



HAL
open science

Caractérisation des styles littéraires par l'extraction automatique des patrons syntaxiques dans des romans français du 19ème siècle

Yuanfeng Lu

► **To cite this version:**

Yuanfeng Lu. Caractérisation des styles littéraires par l'extraction automatique des patrons syntaxiques dans des romans français du 19ème siècle. Linguistique. Université Paris sciences et lettres, 2021. Français. NNT : 2021UPSLE085 . tel-04160741

HAL Id: tel-04160741

<https://theses.hal.science/tel-04160741>

Submitted on 12 Jul 2023

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.



THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à École Normale Supérieure

**Caractérisation des styles littéraires par l'extraction
automatique des patrons syntaxiques dans des romans français
du 19^{ème} siècle**

Soutenue par

Yuanfeng LU

Le 15 novembre 2021

Ecole doctorale n° 540

**École doctorale Lettres, Arts,
Sciences humaines et sociales**

Spécialité

Sciences du langage



Composition du jury :

Dominique LEGALLOIS

Professeur des Universités

Université Sorbonne Nouvelle

Président du jury

Agnès TUTIN

Professeure des universités

Université Grenoble Alpes

Rapporteuse

Iris ESHKOL-TARAVELLA

Professeure des universités

Université Paris Nanterre

Rapporteuse

Michel ESPAGNE

Directeur de recherche au CNRS

École normale supérieure

Examineur

Thierry POIBEAU

Directeur de Recherche au CNRS

École normale supérieure

Directeur de thèse

**Caractérisation des styles littéraires par l'extraction
automatique des patrons syntaxiques dans des romans français
du 19^{ème} siècle**

RÉSUMÉ

Comment l'analyse syntaxique automatique peut-elle être utile pour identifier et caractériser les styles littéraires ? Dans cette thèse, nous proposons de répondre à cette question en considérant les patrons syntaxiques comme des marqueurs stylistiques performants pour identifier, mais aussi caractériser les styles littéraires dans les romans français du 19^{ème} siècle. Un parcours des notions de style, stylistique et stylométrie est présenté au début de la thèse pour mieux situer nos recherches dans la stylistique. Les extractions de patrons syntaxiques (ainsi qu'une méthode originale permettant de mesurer la complexité syntaxique d'un corpus) se fondent sur l'analyse syntaxique en dépendances fournie par le système UDPipe. La méthode d'extraction des patrons caractérisants (à savoir les figures de style d'ordre syntaxique) et la méthode d'extraction des patrons identificatoires (dont la spécificité est de nature purement statistique) sont appliquées à l'œuvre romanesque de Victor Hugo d'une part, et à un corpus plus grand, constitué des romans de Chateaubriand, Flaubert, Hugo et Stendhal d'autre part. Les résultats d'application démontrent la robustesse de notre méthode à caractériser les styles littéraires. Cette méthode, fondée sur des patrons plus précis et plus interprétables que ceux produits par les méthodes précédentes, permet de mener à bien l'analyse stylistique des auteurs du corpus. Notre méthode combinatoire accomplit à la fois la tâche d'identification et de caractérisation de la stylistique, mêlant analyse quantitative et analyse qualitative des patrons identifiés. Elle démontre la capacité satisfaisante des patrons syntaxiques à différencier les styles littéraires, et permet de confirmer (ou au contraire, parfois, de contredire) les observations passées sur les auteurs visés.

MOTS CLÉS

stylistique computationnelle, stylométrie, analyse syntaxique en dépendances, patrons syntaxiques, extraction syntaxique, modélisation statistique

ABSTRACT

How can automatic syntactic analysis be useful to identify and characterize literary styles? In this thesis, we propose to answer this question by considering syntactic patterns as efficient stylistic markers to identify, but also to characterize literary styles in 19th century French novels. A review of the notions of style, stylistics and stylometry is presented at the beginning of the thesis to better situate our research in stylistics. The extraction of syntactic patterns (as well as an original method for measuring the syntactic complexity of a corpus) is based on the syntactic dependency analysis provided by the UDPipe system. The method for extracting characterising patterns (i.e. syntactic figures of speech) and the method for extracting identifying patterns (whose specificity is of a purely statistical nature) are applied to the novels of Victor Hugo on the one hand, and to a larger corpus, consisting of the novels of Chateaubriand, Flaubert, Hugo and Stendhal on the other. The application results demonstrate the robustness of our method in characterising literary styles. This method, based on more precise and interpretable patterns than those produced by the previous methods, makes it possible to carry out a stylistic analysis of the authors in the corpus. Our combinatorial method accomplishes both the task of identification and characterization of stylistics, combining quantitative and qualitative analysis of the identified patterns. It demonstrates the satisfactory capacity of syntactic patterns to differentiate literary styles, and makes it possible to confirm (or, on the contrary, sometimes contradict) past observations on the styles of the authors concerned.

KEYWORDS

computational stylistics, stylometry, dependency parsing, syntactic patterns, syntactic extraction, statistical modeling

*À mes parents,
qui m'ont soutenu tout au long de la thèse*

Remerciements

Je tiens tout d'abord à remercier mon directeur de thèse, M. Thierry Poibeau, qui m'a guidé dès le début pour mes recherches, non seulement par ses conseils et son soutien scientifique, mais aussi par la relecture minutieuse de la thèse.

Je tiens à souligner ma gratitude à M. Dominique Legallois qui, par gentillesse, m'a inspiré à travailler sur le rôle des patrons syntaxiques dans la stylométrie. Les deux rapporteuses, Mme Agnès Tutin et Mme Iris Eshkol-Taravella, et l'examineur de la thèse, M. Michel Espagne, m'ont donné de conseils précieux par leur rapport sur mon travail, mais aussi pendant la soutenance de la thèse.

J'adresse mes sincères remerciements aux différents et nombreux collègues de LATTICE avec qui j'ai collaboré depuis quatre ans. Ils m'ont appris à utiliser les outils informatiques, ainsi que les langages de programmation pour faciliter mes recherches dans le domaine du traitement automatique des langues.

Je remercie aussi M. Zhibin Zhao avec toute mon amitié, qui n'a jamais arrêté de m'encourager pendant les périodes les plus difficiles de la thèse.

Enfin, un grand merci à mes parents, c'est grâce à leur soutien et leur accompagnement à distance que j'ai pu finir la thèse.

Table des matières

<i>Chapitre 1 Introduction</i>	1
1.1 La problématique	2
1.2 Les corpus	4
1.3 Les apports de la thèse	8
1.4 Organisation de la thèse	9
<i>Chapitre 2 Le style, la stylistique et la stylométrie</i>	11
2.1 Introduction	11
2.2 Le style et la stylistique	13
2.2.1 Style : génie ou expressivité ?	14
2.2.2 Le style « parfait » et la singularité stylistique	19
2.2.3 Stylistiques psychologique et structuraliste	24
2.2.4 Le style de texte : le texte et son contexte	29
2.2.5 La question de la littéarité	36
2.2.6 Les deux variétés de stylistiques	41
2.3 La stylométrie et la stylistique de corpus	47
2.4 État de l'art sur la stylométrie	53
2.4.1 Les débuts de la stylométrie : le niveau lexical	53
2.4.2 Vers la combinaison de paramètres stylistiques	56
2.4.3 Niveau syntaxique	63
2.5 Conclusion	77
<i>Chapitre 3 Présentation de la méthodologie</i>	79
3.1 Introduction	79
3.2 Les relations syntaxiques : linéarité ou hiérarchie	81
3.3 Comparaison des performances des différents analyseurs syntaxiques en	

dépendances pour la langue française.....	88
3.4 Les patrons caractérisants et les patrons identificatoires	93
3.5 Une nouvelle méthode pour mesurer la complexité syntaxique.....	99
3.5.1 Les méthodes précédentes pour mesurer la complexité syntaxique	99
3.5.2 Présentation de la méthode	102
3.6 Les extractions des patrons caratérissants.....	105
3.6.1 L'identification et l'extraction des inversions du sujet	106
3.6.2 L'identification et l'extraction des phrases clivées	113
3.6.3 L'identification et l'extraction des répétitions	118
3.7 L'extraction des patrons identificatoires.....	124
3.7.1 La méthode d'extraction linéaire des patrons identificatoires	125
3.7.2 La méthode d'extraction hiérarchique des patrons syntaxiques	127
3.8 Conclusion	147
<i>Chapitre 4 Caractérisation du style romanesque de Victor Hugo par la méthode des patrons syntaxiques.....</i>	<i>149</i>
4.1 Introduction.....	149
4.2 Le style romanesque de Victor Hugo.....	151
4.2.1 Aperçu sur le style de Victor Hugo	151
4.2.2 La matérialisation de l'hétérogénéité dans les œuvres de Victor Hugo	154
4.3 Mesurer la complexité syntaxique dans les romans de Victor Hugo.....	159
4.3.1 Calcul de la complexité syntaxique des romans de Victor Hugo.....	159
4.4 Application de la méthode d'extraction supervisée sur le corpus Victor Hugo	162
4.4.1 L'inversion du sujet.....	162
4.4.2 Le clivage.....	178
4.4.3 La répétition.....	192
4.5 Application de la méthode d'extraction non-supervisée au corpus Victor	

Hugo	201
4.5.1 Analyses quantitatives des patrons syntaxiques.....	202
4.5.2 Analyses qualitatives des patrons identificatoires.....	208
4.6 Conclusion	237
<i>Chapitre 5 Caractérisation du style des quatre auteurs du 19^{ème} siècle de notre corpus par la méthode des patrons syntaxiques.....</i>	239
5.1 Introduction.....	239
5.2 Les styles romanesques de Chateaubriand, Flaubert et Stendhal.....	241
5.2.1 Aperçu sur le style de Chateaubriand.....	241
5.2.2 Aperçu sur le style de Flaubert.....	245
5.2.3 Aperçu sur le style de Stendhal.....	249
5.3 Mesurer la complexité syntaxique dans les romans de Chateaubriand, Flaubert, Hugo et Stendhal	253
5.4 Application de la méthode d'extraction supervisée	256
5.4.1 L'inversion du sujet.....	256
5.4.2 Le clivage.....	275
5.4.3 La répétition.....	293
5.5 Application de la méthode d'extraction non-supervisée à notre corpus ...	306
5.5.1 Analyses quantitatives des patrons syntaxiques.....	307
5.5.2 Analyses qualitatives des patrons syntaxiques.....	311
5.6 Conclusion	335
<i>Chapitre 6 Conclusion</i>	337
6.1 Introduction.....	337
6.2 Les résultats d'analyse.....	337
6.2.1 La complexité syntaxique	337
6.2.2 Les patrons caractérisants.	339
6.2.3 Les patrons identificatoires	341

6.3 Perspectives et suggestions pour de futures recherches	344
6.3.1 Pour les patrons caractérisants	344
6.3.2 Pour les patrons identificatoires.....	345
6.4 Conclusion	346
<i>Annexe 1 Deux échantillons de phrases pour tester la performance des analyseurs syntaxiques</i>	<i>347</i>
<i>Annexe 2 Liste des verbes courants dans la construction d'incise</i>	<i>361</i>
<i>Annexe 3 Les algorithmes d'extraction.....</i>	<i>367</i>
<i>Bibliographie.....</i>	<i>371</i>

Glossaire

ADJ	adjective
ADV	adverbe
AFC	analyse factorielle des correspondances
DT	déterminant
foregrounding	la surreprésentation des traits linguistiques qui « font référence à un stimulus culturellement non attendu dans une situation sociale et capables de provoquer une attention particulière » (Garvin, 1964, p. 8)
littérarité	« ce qui fait d'une œuvre donnée une œuvre littéraire » (Jakobson et al. 1977)
n-gram	une séquence contiguë de n éléments provenant d'un échantillon donné de texte ou de parole
NN	nom
NP	noun phrase (groupe nominal)
patron caractérisant	patron stylistiquement significatif qui définit les qualités d'un style littéraire
patron identificateur	patron syntaxique statistiquement spécifique au sein d'un texte
PRON	pronom
PUNCT	ponctuation
VB	verbe
VP	verb phrase (groupe verbal)

Chapitre 1 Introduction

Le point de départ de la thèse est le rôle indéniable de la syntaxe dans la stylistique. En effet, la longueur des phrases, leur structure et la façon dont elles sont construites révèlent des choix fondamentaux de l'écrivain, ce qui fonde le style. Pourtant, les études stylistiques traditionnelles sont souvent limitées à certains passages ou extraits d'un texte (ceci était vrai avant l'introduction des outils informatiques dans le domaine, mais reste encore largement vrai aujourd'hui). Par conséquent, la généralisation des caractéristiques stylistiques est souvent difficile, voire « impressionniste » (au sens où il s'agit essentiellement de généralisations hypothétiques à partir d'impressions partielles).

Le développement de la stylistique de corpus depuis le 20^{ème} siècle a permis aux chercheurs de dépasser des études partielles et fragmentaires ; de plus, les méthodes statistiques ont apporté la possibilité de quantifier les phénomènes observables au sein des textes. Mais, bien que la stylistique de corpus intègre différentes méthodes statistiques, elle s'est concentrée au début presque exclusivement sur le lexique, au détriment de la syntaxe.

On peut identifier deux raisons à cet état des choses : d'une part, les paramètres syntaxiques sont plus difficiles à repérer dans les textes et demandent des outils informatiques performants, qui ne sont apparus (avec un bon niveau de performance) que ces dernières années ; d'autre part, les paramètres lexicaux présentent déjà des résultats satisfaisants pour différencier les œuvres (voir notre état de l'art dans le Chapitre 2). La syntaxe a d'abord été utilisée selon le même modèle que le lexique, en valorisant la capacité identificatoire des paramètres, c'est-à-dire leur capacité à identifier une œuvre par rapport à une autre et à l'attribuer à un auteur supposé (tâche dite d'« identification d'auteur »). Cependant, la tâche d'identification n'est que la première étape des démarches stylistiques, et l'apport des outils informatiques ne doit

pas se limiter à cela. La stylométrie, d'après nous, doit accomplir à la fois la tâche d'identification (dire ce qui est spécifique au sein d'un corpus) et la tâche de caractérisation (dire en quoi ce qui est repéré permet de définir un style particulier ; comment caractériser ce style ?). Notre projet a ainsi pour objectif de relier la tradition stylistique qui se concentre sur les qualités du style par l'analyse littéraire (en quoi un trait linguistique produit un effet littéraire) et les techniques de la stylométrie contemporaine, qui exploite systématiquement des paramètres quantifiables et identificatoires.

« Le style de Hugo est inspiré, et ses inventions puissantes¹ ». « La marque propre du style de Hugo est le fait d'être constamment plein d'images² » (Puková, 2018). « Le style de Hugo est varié, avec de beaux mouvements oratoires, un vocabulaire riche, mais parfois un peu grandiloquent³ ». Sur quoi se fondent ces affirmations ? Sont-elles vraies à l'échelle d'une œuvre comme celle de Hugo ? Et surtout, comment le déterminer vu la longueur imposante du corpus hugolien ?

On s'attache dans cette thèse à fournir des éléments de réponse à ces questions, en proposant des méthodes automatiques permettant d'explorer le style de Hugo et d'autres auteurs du 19^e siècle, tout en laissant une large place à l'interprétation humaine, indispensable dans ce type d'étude.

1.1 La problématique

La problématique qui motive cette thèse est la suivante : comment l'analyse syntaxique automatique peut-elle être utile pour identifier et caractériser les styles

¹ <https://www.french.hku.hk/victor-hugo>, consulté le 6 septembre, 2021.

² Puková, Lucie. 2018. « Lyrification et dramatisation dans Les Misérables de Victor Hugo ». Masarykova univerzita, Filozofická fakulta. <https://theses.cz/id/ybgbt0/>.

³ <https://www.sculfort.fr/articles/grammaire/avance/laphrasefrancaise.html>, consulté le 6 septembre, 2021.

littéraires ? Afin de répondre à cette question, nous essayons d'évaluer la capacité des patrons syntaxiques à caractériser le style des écrivains français du 19^{ème} siècle à partir de corpus de romans. Deux études comparatives sont mises en place dans la thèse : l'une entre les premiers et les derniers romans de Victor Hugo, l'autre entre l'ensemble des romans de Victor Hugo et ceux de ses contemporains : Chateaubriand, Flaubert et Stendhal.

Le choix de ces quatre auteurs se base sur le fait que les critiques littéraires sont relativement unanimes pour qualifier leur style. Hugo a produit ses œuvres romanesques en deux périodes bien distinctes (avant et après son exil à Jersey et Guernesey, pendant le second Empire) et un net changement de style peut être distingué entre les œuvres d'avant et d'après l'exil. D'après Judith Wulf, le style de Hugo après l'exil trahit une « hétérogénéité⁴ » (Wulf, 2014), alors qu'avant l'exil, Hugo cherchait plutôt à « restaurer », « harmoniser » et « combler ⁵ » (Hugo, 1985 [1827], p. 24). Brombert partage la même idée en disant « qu'aucun des romans de jeunesse de Hugo ne peut prétendre être un grand roman⁶ » (Brombert, 1985, p. 23). Pour Chateaubriand, les stylisticiens tels que Sainte-Beuve pense « qu'il y a innovation et rénovation du langage en même temps que tradition⁷ » (Sainte-Beuve, 1926). Ainsi, Chateaubriand marie tradition littéraire et innovation langagière, ce qui est représentatif du courant romantique. Le style de Flaubert est également reconnaissable par sa fluidité, son usage de la coupe oratoire et par « l'harmonie⁸ » générale de ses phrases (Thibaudet, 1935). Enfin, le style que Stendhal cherche à atteindre est celui qui est « clair » et « naturel⁹ » (Talbot, 1985).

⁴ Wulf, Judith. 2014. *Étude sur la langue romanesque de Victor Hugo: le partage et la composition*. Paris, France: Classiques Garnier.

⁵ Hugo, Victor. 1985 [1827]. « Préface de Cromwell » dans *Œuvres complètes*. Édité par Guy Rosa, Jacques Seebacher, et Groupe Hugo. 16 vol. Paris, France : R. Laffont, p. 24.

⁶ Brombert, Victor, et Héloïse Neefs. 1985. *Victor Hugo et le roman visionnaire*. Paris, France: PUF, p. 23.

⁷ Sainte-Beuve. 1926. *Mes poisons: cahiers intimes inédits*. Édité par Victor Giraud. Paris, France: Plon-Nourrit et Cie, impr.1926.

⁸ Thibaudet, Albert. 1935. *Gustave Flaubert*. 4e éd. 1 vol. Paris: Gallimard. [https://fr.m.wikisource.org/wiki/Gustave_Flaubert_\(Thibaudet\)](https://fr.m.wikisource.org/wiki/Gustave_Flaubert_(Thibaudet)), consulté le 25 juin, 2021.

⁹ Talbot, Emile J. 1985. *Stendhal and Romantic Esthetics*. French Forum Monographs 61. Lexington, Ky: French forum.

La méthode que nous proposons dans la thèse, basée sur les patrons syntaxiques, vise à tester la capacité de ce paramètre à relever les caractéristiques stylistiques qui peuvent confirmer (ou éventuellement infirmer) ce que disent les stylisticiens sur le style de ces quatre auteurs. Au-delà, on espère évidemment que la méthode puisse aussi mettre au jour des faits nouveaux, même si beaucoup a déjà été dit sur le style de ces quatre auteurs (c'est d'ailleurs pour cela que nous les avons choisis, les études antérieures constituant des points de référence par rapport à nos résultats et nos propositions). Notre méthode propose une combinaison d'analyses quantitative et qualitative de ces patrons, afin d'évaluer leur capacité non seulement à identifier, mais aussi à caractériser les styles littéraires.

1.2 Les corpus

Deux corpus sont considérés pour notre projet de recherche. Le premier est composé des neuf romans de Victor Hugo publiés entre 1818 et 1874. La comparaison que nous essayons de mettre en place se base sur le fait qu'après la publication de *Claude Gueux* en 1834, Victor Hugo n'a pas publié un seul roman pendant 28 ans, en raison de son exil en Belgique et en Angleterre. La reprise de sa publication romanesque se situe en 1862 quand il publie son roman le plus volumineux, *Les Misérables*, écrit pendant son exil. Ainsi, les créations romanesques de Victor Hugo se divisent chronologiquement en deux parties, celle qui comprend les premiers romans avant l'exil et celle qui comprend ses derniers romans pendant et après l'exil. Selon les critiques littéraires comme Judith Wulf, les différences stylistiques sont perceptibles entre les deux périodes de publication de Hugo¹⁰ (1818-1834 et 1862-1874) (Wulf, 2014). Ainsi, la construction de notre corpus Victor Hugo vise à vérifier, par la

¹⁰ Wulf, Judith. 2014. *Étude sur la langue romanesque de Victor Hugo: le partage et la composition*. Paris, France: Classiques Garnier.

méthodologie de la thèse, s'il y a bien des caractéristiques stylistiques propres à chaque période de publication de Hugo.

Le tableau suivant présente les neuf romans du corpus Victor Hugo ainsi que leur date de première publication. Les éditions utilisées sont celles du recueil *Œuvres Complètes* de Victor Hugo, publié en 2013 par Arvensa Editions¹¹. Le choix de ce recueil se base sur le fait que cet ouvrage existe en version numérique facilement accessible par les chercheurs, il ne demande ainsi pas d'océrisation pour obtenir les textes bruts. De plus, ce recueil rassemble presque toutes les œuvres de Victor Hugo (122 titres) avec les préfaces et les notes, cela nous permet de construire le corpus Victor Hugo sans éléments manquants dans les textes.

Romans	Date de première publication
<i>Bug-Jargal</i>	1818
<i>Han d'Islande</i>	1823
<i>Le Dernier Jour d'un condamné</i>	1829
<i>Notre-Dame de Paris</i>	1831
<i>Claude Gueux</i>	1834
<i>Les Misérables</i>	1862
<i>Les Travailleurs de la mer</i>	1866
<i>L'Homme qui rit</i>	1869
<i>Quatrevingt-treize</i>	1874

Tableau 1.1 Les romans dans le corpus Victor Hugo.

Le deuxième corpus est construit afin de démontrer la robustesse de notre méthode pour caractériser les styles d'auteur (et non pas seulement les styles de texte). Chateaubriand, Flaubert et Stendhal sont des contemporains de Victor Hugo. Nous

¹¹ Hugo, Victor, et ' Arvensa Editions. 2013. *Victor Hugo: Oeuvres complètes - 122 titres (Annotés et illustrés)* - Arvensa Editions. Arvensa Editions.

sélectionnons uniquement leurs romans dans la construction de ce corpus pour que la comparaison stylistique soit possible, dans la mesure où le genre joue un rôle important (on n'écrit pas du théâtre comme on écrit un roman). Le corpus de Chateaubriand, Flaubert, Hugo et Stendhal se compose de 35 romans au total. Le tableau ci-dessous présente les œuvres romanesques dans ce corpus. Les cinq romans de Chateaubriand proviennent *Des œuvres intégrales de François-René de Chateaubriand* de Ligram éditions¹², les sept romans de Flaubert proviennent de *Flaubert l'intégrale : Œuvres complètes, 69 titres* d'Ink Book édition¹³, les 14 romans de Stendhal proviennent de *Stendhal Œuvres complètes* d'Arvensa éditions¹⁴. Le choix des éditions se base également sur le fait que ce sont les éditions qui couvrent le plus de titres pour chaque auteur, avec les préfaces et les notes pour chaque texte. Par ailleurs, ces éditions sont toutes en version numériques, ce qui facilite la préparation des textes bruts pour chaque auteur.

¹² Chateaubriand, François-René de. 2013. *Les œuvres intégrales - François-René de Chateaubriand* - Ligram éditions - ebook (ePub) - Le Hall du Livre NANCY. Ligram éditions.

¹³ Flaubert, Gustave. 2021. *Flaubert l'intégrale : Œuvres complètes, 69 titres*. 1er édition. Ink book.

¹⁴ Stendhal, et ' Arvensa Editions. 2014. *Stendhal : Œuvres complètes*. Arvensa Editions.

Chateaubriand	Flaubert	Victor Hugo	Stendhal
<i>Atala</i>	<i>Bouvard et Pécuchet</i>	<i>Bug-Jargal</i>	<i>Armance</i>
<i>René</i>	<i>La tentation de Saint Antoine</i>	<i>Han d'Islande</i>	<i>Féder</i>
<i>Les Martyrs</i>	<i>L'Éducation sentimentale</i>	<i>Le Dernier Jour d'un condamné</i>	<i>La Chartreuse de Parme</i>
<i>Les Natchez</i>	<i>Madame Bovary</i>	<i>Notre-Dame de Paris</i>	<i>Lamiel</i>
<i>Les Aventures du dernier Abercerage</i>	<i>Mémoires d'un fou</i>	<i>Claude Gueux</i>	<i>Le Chevalier de Saint-Ismier</i>
	<i>Salammbô</i>	<i>Les Misérables</i>	<i>Le Coffre et le revenant</i>
	<i>Trois contes</i>	<i>Les Travailleurs de la mer</i>	<i>Le Juif</i>
		<i>L'Homme qui rit</i>	<i>Le Philtre</i>
		<i>Quatrevingt-treize</i>	<i>Le Rose et le Vert</i>
			<i>Le Rouge et le Noir</i>
			<i>Mina de Vanghel</i>
			<i>Lucien Leuwen</i>
			<i>Philibert Lescale</i>
			<i>Souvenirs d'un gentilhomme italien</i>

Tableau 1.2 Les romans dans le corpus Chateaubriand, Flaubert, Victor Hugo et Stendhal.

1.3 Les apports de la thèse

L'exploitation des paramètres d'ordre syntaxique dans la stylométrie outillée est possible grâce aux développements récents d'outils de traitement automatique des langues performants. Les analyseurs syntaxiques nous permettent non seulement d'annoter d'une manière automatique les composants phrastiques (en associant aux mots des étiquettes morphosyntaxiques), mais aussi d'identifier les relations syntaxiques entre ces composants phrastiques. La méthode que nous avons élaborée se fonde sur l'analyse syntaxique en dépendances des textes bruts. L'annotation, l'analyse syntaxique ainsi que l'extraction des patrons syntaxiques sont automatiques, mais la vérification de ces annotations demande un travail partiellement manuel.

Les apports de la thèse sont les suivants :

1. Extraction automatique de trois figures de style repérable par analyse syntaxique (l'inversion du sujet, le clivage et la répétition).

2. Extraction, d'une manière non-supervisée, des patrons caractéristiques, sur une base syntaxique, à partir des relations de dépendance. L'intégration des relations syntaxiques augmente considérablement le pourcentage de patrons interprétables par rapport aux méthodes traditionnelles généralement fondées sur une simple analyse de séquences linéaires d'étiquettes. Cette méthode améliore la précision d'extraction et, par conséquent, facilite l'analyse qualitative des patrons syntaxiques.

3. Élaboration d'une nouvelle méthode d'analyse permettant de caractériser le style sur la base de la complexité syntaxique. Cette méthode se fonde sur l'analyse des relations syntaxiques de dépendance au sein des phrases et mesure la complexité de ces relations (en particulier les niveaux de coordination et de subordination au sein des phrases).

4. Proposition d'un point de vue combinatoire pour la stylistique. Les deux tâches de la stylistique, à savoir celle d'identification (différencier les œuvres) et celle de caractérisation (définir les qualités du style) doivent prendre en compte à la fois les

patrons identificatoires (tels que les patrons syntaxiques statistiquement spécifiques) et les patrons caractérisants (tels que les figures de style dont les fonctions stylistiques sont bien connues). La thèse insiste sur le fait que les deux catégories de patrons sont indispensables pour les études stylistiques.

En se basant sur les phénomènes d'ordre syntaxique dans les textes, la thèse apporte une nouvelle méthode qui exploite les patrons syntaxiques comme marqueurs stylistiques. Elle démontre, par son application sur des textes français du 19^{ème} siècle, que les unités syntaxiques peuvent être utilisées pour identifier, mais aussi pour caractériser les styles littéraires. L'application de notre méthode sur nos deux corpus a permis de mettre au jour de nouvelles caractéristiques stylistiques qui pourront intéresser les futures études, tout en relevant avec succès les phénomènes stylistiques propres à chaque auteur déjà relevés dans différentes études littéraires.

1.4 Organisation de la thèse

Les grandes questions et débats autour des notions de style, de stylistique et de stylométrie sont discutés dans le Chapitre 2. Nous présentons des discussions historiques sur ces notions, afin de mieux situer notre propre démarche concernant la stylistique. Ce chapitre présente également notre point de vue combinatoire non seulement en ce qui concerne les tâches d'identification et de caractérisation au sein de la démarche stylistique, mais aussi en matière d'analyses quantitative et qualitative qui doivent être appliquées sur ces patrons. Le Chapitre 3 est la présentation des méthodes d'analyse développées dans le cadre de la thèse, notamment la méthode d'extraction supervisée des patrons caractérisants (les figures de style) et la méthode d'extraction non-supervisée de patrons syntaxiques. Ces différents points sont discutés en détail dans ce chapitre et servent de base aux études des chapitres suivants. Dans le Chapitre 4, nous présentons l'application de la méthodologie au corpus Victor Hugo : il est ainsi

possible de montrer la robustesse de la méthode et sa capacité à identifier et caractériser le style de Hugo, en particulier les différences de style entre les romans d'avant et d'après l'exil de l'auteur. Le Chapitre 5 enfin est consacré à l'application de la méthode aux corpus de Chateaubriand, Flaubert, Hugo et Stendhal. Cette étude a pour objectif principal de montrer la robustesse de la méthode et son applicabilité à différents auteurs. Elle permet en particulier de différencier les auteurs entre eux, et de mettre en avant les particularités de chacun par le truchement des patrons syntaxiques identificatoires et caractérisants.

Chapitre 2 Le style, la stylistique et la stylométrie

2.1 Introduction

Pour mieux situer notre point de départ théorique sur la stylistique, nous présentons, dans ce chapitre, une étude des théories du style, de la stylistique et de la stylométrie. La thèse étant avant tout une enquête sur les capacités de la syntaxe à caractériser les styles littéraires, en particulier dans les romans français du 19^e siècle, nous essayons d'explicitier le point de départ de nos recherches, ainsi que leurs bases théoriques. La stylométrie outillée est d'après nous une branche de la stylistique de corpus : d'une part, la stylométrie emploie des outils informatiques qui permettent de relever systématiquement les phénomènes linguistiques (pour pouvoir démontrer leurs significations quantitatives), d'autre part, elle recourt aux théories littéraires (pour comprendre l'apport de ces phénomènes formels à l'analyse qualitative). Nous présentons dans ce chapitre un parcours général de la stylistique et de la stylométrie, en mettant en relief notre point de vue par rapport aux différents courants qui ont marqué l'histoire de la stylistique.

Au lieu de faire un résumé de l'histoire de la stylistique, nous présentons dans la section 2.2 six contradictions théoriques qui ont suscité de longs débats dans cette discipline. Ces contradictions portent sur : la nature du style, la notion de singularité stylistique, le problème de la subjectivité dans la stylistique, l'objet d'étude de la stylistique (le texte et le contexte), la recherche des significations stylistiques pour les phénomènes textuels saillants et les deux variétés de stylistiques selon Stanley Fish (la stylistique identificatoire et caractérisante). Nous avons fait le choix de présenter des points de vue contradictoires, afin de mieux mettre en avant nos propres points de vue sur le style et la stylistique. Nous n'essayons pas de redéfinir ces notions, mais nous

proposons un parcours des grandes idées du domaine, pour trouver les bases théoriques sur lesquelles devraient selon nous se situer la stylistique de corpus, ainsi que la stylométrie outillée de nos jours.

Dans la section 2.3, une étude spécifique sur la stylométrie est présentée. Cette excroissance de la stylistique de corpus à l'ère informatique s'est initialement intéressée aux aspects quantitatifs des traits formels dans les textes, mais elle vise également l'interprétation subjective de ces traits linguistiques. Nous pensons que ce tournant vers le corpus a confirmé la nécessité d'étudier les phénomènes saillants dans les textes, mais la possibilité de traiter une grande quantité de données ne doit pas faire oublier le but de la stylistique, à savoir interpréter ces phénomènes saillants dans leurs contextes respectifs.

Nous présentons l'état de l'art sur la stylométrie dans la section 2.4. Les premiers essais stylométriques ont été réalisés au niveau du lexique : les chercheurs ont d'abord cherché à confirmer la possibilité de résoudre des problèmes d'attribution d'auteur à travers l'analyse de la distribution des mots fonctionnels. Ensuite, la même volonté de combiner analyses statistiques et qualitatives a été constatée quand les recherches ont commencé à intégrer le champ sémantique. Nous présentons enfin quelques études importantes sur la syntaxe, qui ont inspiré et influencé notre propre recherche sur ce sujet.

Ce parcours théorique nous fait comprendre que la stylométrie outillée, comme la stylistique elle-même, fait se croiser analyses quantitatives et qualitatives. Bien que la plupart des études que nous présentons dans cette partie s'intéressent uniquement aux possibilités apportées par les outils informatiques dans les recherches stylistiques, l'interprétation qualitative n'est jamais absente. Au lieu de proposer une seule méthode qui fusionne les deux variétés de stylistiques (c'est-à-dire l'approche qui se concentre sur les traits identificatoires et celle qui se concentre sur les traits caractérisants), nous proposons de considérer les deux variétés comme complémentaires en donnant une vue élargie des caractéristiques du texte.

2.2 Le style et la stylistique

Les définitions de la notion de style sont aussi diverses que les sujets que recouvre ce terme. Nous adoptons dans la thèse une définition linguistique : « le style se définit d'une part, comme l'écart de la norme linguistique par rapport aux formes standards, et d'autre part, comme la récurrence de l'apparition de ces écarts¹⁵ » (Molinié, 1994). Nous adoptons ensuite la définition de Raymond Chapman pour la stylistique : la stylistique est « l'étude linguistique des différents styles¹⁶ » (Chapman, 1973), elle consiste à nous aider à comprendre comment les êtres humains verbalisent leurs conceptions littéraires par les codes langagiers.

Il ne nous semble pas nécessaire de retracer toute l'histoire de la stylistique dans cette partie, en revanche, nous présentons six grands problèmes qui ont fait l'objet de longs débats dans les milieux s'intéressant au style et à la stylistique. Nous commençons par nous interroger sur la nature du style (2.2.1) : qu'est-ce que le style ? Est-il saisissable par les outils informatiques ? Ensuite, nous discutons de la notion de singularité stylistique (2.2.2) : existe-t-il un « style parfait » sans auteur ? La section 2.2.3 parle du problème de la subjectivité dans la stylistique. Dans 2.2.4, nous présentons l'objet d'étude de la stylistique : le texte. Nous argumentons également sur l'importance du contexte dans les démarches stylistiques. Dans 2.2.5, nous explorons comment les traits linguistiques saillants peuvent être associés à des significations stylistiques selon la théorie sémiostylistique de Georges Molinié. Enfin, dans 2.2.6, nous présentons les deux variétés de stylistiques (identificatoire et caractérisante) selon

¹⁵ Molino, Jean. 1994. « Pour une théorie sémiologique du style », dans Molinié, Georges Organisateur de réunion, Pierre, Cahné Organisateur de réunion, Colloque international Qu'est-ce que le style ?, et École doctorale Sciences du langage. 1994. *Qu'est-ce que le style ? : actes du colloque international*. Paris, France: Presses universitaires de France.

¹⁶ Chapman, Raymond. 1973. *Linguistics and literature: an introduction to literary stylistics*. London, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord: Edward Arnold, 1973.

Stanley Fish : nous défendons la position que la combinaison de ces deux variétés est idéale pour chaque recherche stylistique.

2.2.1 Style : génie ou expressivité ?

Nous abordons d'abord la contradiction qui porte sur la nature du style. Le problème de la nature du style est crucial parce que la définition adoptée nous aidera à répondre à la question suivante : comment étudier le style ? Les deux parties opposées dans cette contradiction répondent à cette question de deux manières différentes : pour les uns, le style est le génie de l'auteur, qui n'est pas directement saisissable par les phénomènes formels. Aussi le style est-il inséparable du langage lui-même : par conséquent, étudier le style doit être une démarche qui cherche à déchiffrer ce génie de l'auteur ; pour les autres, le style est une expressivité, c'est-à-dire, une fonction expressive qui est ajoutée volontairement par l'auteur sur les informations linguistiques. Aussi le style serait-il tout à fait séparable du langage. Pour mieux étudier cette fonction expressive, la linguistique a proposé la méthode structurale, tandis que Léo Spitzer propose une méthode qui consiste à dévoiler certains aspects psychologiques de l'auteur par le truchement des phénomènes expressifs dans les textes.

Avant d'entrer dans les détails de cette contradiction qui porte sur la nature du style, il est impératif d'évoquer un grand tournant dans la stylistique moderne : à savoir l'individuation des artistes. Selon François Rastier, cette individuation des artistes, ou plus précisément, l'intention des artistes de se démarquer les uns des autres, fait naître une définition plus étroite de la stylistique par rapport à la rhétorique antique : avec l'individuation progressive des artistes, la stylistique moderne se concentre de plus en plus sur la notion de style d'auteur, alors que la rhétorique, dans sa définition antique, servait principalement à étudier et à distinguer les différents genres de discours¹⁷ (Rastier, 2001). Pour mieux comprendre la relation entre la stylistique moderne et la

¹⁷ Rastier, F. 2001. *Arts et sciences du texte*. Paris cedex 14, France : Presses Universitaires de France.

rhétorique antique, disons que la stylistique moderne a hérité uniquement de l'une des cinq techniques de la rhétorique : l'*élocutio*, c'est-à-dire la sélection du style pour un effet approprié¹⁸ (Carter, 2008, p. 292). Avec cette individuation que constatent Carter et Rastier dans la stylistique moderne, nous commençons à considérer le style comme une forme particulière à chaque auteur, qui porte l'empreinte de sa personnalité. C'est dans ce contexte historique et par ce tournant stylistique que se situent les deux parties de cette contradiction sur la nature du style.

Commençons par présenter la première partie de cette contradiction : la manifestation la plus connue de ce courant est sans doute la phrase de Buffon dans son *Traité du Style* : « Le style, c'est l'homme même¹⁹ » (Buffon, 1886). Buffon attribue le style non pas aux genres littéraires, mais aux auteurs. Plusieurs théoriciens à son époque prennent son relais, dont Schopenhauer, qui définit le style comme « la physionomie de l'esprit²⁰ » (Schopenhauer, 2020 [1851], p. 44). Le style est ainsi inséparable de l'auteur lui-même. Victor Hugo, dont les œuvres font l'objet d'analyses dans notre corpus, résume cette théorie du « génie » :

« Le style a quelque chose de préexistant. Il reste toujours de son espèce. Il jaillit de tout l'écrivain, de la racine de ses cheveux aussi bien que des profondeurs de son intelligence. Tout le génie, son côté terrestre comme son côté cosmique, son humanité comme sa divinité, le poète comme le prophète, sont dans le style. Le style est âme et sang ; il provient de ce lieu profond de l'homme où l'organisme aime ; le style est entrailles²¹. » (Hugo, 1985 [1901], p. 622)

¹⁸ Carter, Ronald, et Peter Stockwell, éd. 2008. *The Language and Literature Reader*. 1 vol. London : Routledge.

¹⁹ Buffon, Georges-Louis Leclerc. 1886. *Discours sur le style*. Paris, France : C. Delagrave.

²⁰ Schopenhauer, Arthur. 2020 [1851]. *Essays of Schopenhauer*. Otbebookpublishing. P. 44.

²¹ Hugo, Victor. 1985 [1901]. « Les traducteurs » dans *Œuvres complètes*. Édité par Guy Rosa, Jacques Seebacher, et Groupe Hugo. 16 vol. Paris, France : R. Laffont, p. 622.

Le point de vue de Victor Hugo sur le style marque une rupture avec la rhétorique de l'Antiquité : si la rhétorique implique de produire un discours conforme aux lois langagières, si les lieux rhétoriques évoquent l'idée d'obédience et de normes pour écrire correctement et esthétiquement, le style, selon Hugo, « est libre. Il traverse en pleine liberté tous les alambics de la grammaire²² » (Hugo, 1985 [1901]). Hugo va ensuite plus loin dans sa théorie du style, en exprimant son intention d'« oser, hasarder, créer, inventer son style²³ » (Hugo, 1985 [1827], p. 30). Cette théorie donne lieu à une réflexion sur l'articulation entre la langue et la pensée²⁴ (Wulf, 2015, p. 24) : pour Hugo, le fond n'est pas différent de la forme. Le langage doit prendre en compte ces deux aspects, tandis que le style est une manifestation du génie de la langue. Le style est ainsi inséparable du texte lui-même.

Les auteurs que nous venons de citer, le courant du « génie » (pour emprunter ce même terme de Victor Hugo), viennent se heurter à une autre conception du style. Nous présentons maintenant l'autre parti, qui entre en débat avec la conception de Hugo.

Cette autre conception met l'accent sur la fonction expressive du style : les théoriciens de ce courant pensent que le style est un moyen à rendre la langue plus expressive. Nous pouvons citer un auteur de notre corpus — Stendhal — à l'égard de ce problème. Stendhal pense que l'essence du style est d'« ajouter à une pensée donnée toutes les circonstances propres à produire tout l'effet que doit produire cette pensée²⁵ » (Ullmann, 1957, p. 2). Le mot-clé de cette théorie est la notion de « choix » : tout repose sur le fait que l'auteur peut choisir entre différentes formes d'expressions. En effet, l'étude stylistique n'est possible, d'après ce courant, que par le truchement des analyses comparatives concernant les choix de l'auteur pour façonner son texte. La mise en relief de la notion d'expressivité a fait naître deux écoles stylistiques qui, selon nous,

²² *Idem.*

²³ Hugo, Victor. 1985 [1827]. « Préface de Cromwell » dans *Œuvres complètes*. Édité par Guy Rosa, Jacques Seebacher, et Groupe Hugo. 16 vol. Paris, France : R. Laffont, p. 622, p. 30.

²⁴ Wulf, Judith. 2014. *Étude sur la langue romanesque de Victor Hugo : le partage et la composition*. Paris, France: Classiques Garnier, p. 24.

²⁵ Cité par Ullmann, Stephen. 1957. *Style in the French novel*. Cambridge Eng., Royaume-Uni de Grande-Bretagne et d'Irlande du Nord : University Press, p. 2.

représentent chacune à sa manière un courant de pensée s'opposant à la notion de « génie » langagier : l'école structuraliste qui part des particularités structurales et grammaticales des textes, et l'école psychologique, qui essaie de saisir les traits linguistiques pouvant être les indices de la psychologie de l'auteur. Les deux écoles ne sont pas apparues dans la même époque, ni ne sont-elles issues d'une même tradition littéraire, mais elles continuent à enrichir respectivement les débats sur la notion de style de nos jours. Ainsi, le parcours de ces deux écoles nous semble éclairant pour comprendre les différents avis modernes sur les études stylistiques.

Héritier de Ferdinand de Saussure, Charles Bally est sans doute le linguiste le plus représentatif de l'école structuraliste de la stylistique. La stylistique structuraliste « recenserait les divers moyens dont disposent les locuteurs pour exprimer une idée ou une vision²⁶ » (Philippe, 2013, p. 147). Selon Fowler, le structuralisme en linguistique fournit trois perspectives importantes pour les études du texte :

« Le texte peut être vu comme une suite de phrases, chacune pourrait être analysée linguistiquement ; ou comme une construction unifiée unique avec sa propre structure particulière en plus des schémas de phrases qu'elle tire des règles de la langue ; troisièmement, un texte littéraire peut être vu comme une unité au sein d'un système littéraire, dans le contexte d'un ensemble d'autres œuvres pertinentes liées à la structure sémiotique de toute la culture²⁷. » (Fowler, 1981, p. 14)

Selon Bally, la stylistique doit étudier la langue en général et ses ressources proprement stylistiques en particulier. Au lieu d'analyser le « génie » du langage et le

²⁶ Philippe, Gilles. 2013. « Les deux corps du style ». *Les Temps Modernes*, 5(5), P. 147.

²⁷ « The text may be seen as a sequence of sentences each to be analyzed linguistically; or as a single unified construction with its own particular internal structure in addition to the sentence-patterns it draws from the rules of the language; third, a literary text may be seen as a unit within a literary system, within the context of a set of relevant other words related to the semiotic structure of the whole culture ». Fowler, Roger. 1981. *Literature as Social Discourse: The Practice of Linguistic Criticism*. Bloomington: Indiana university press, p. 14.

« génie » de l'auteur, le structuralisme propose d'étudier toutes les ressources stylistiques et structurales qui contribuent à l'expressivité du texte. Bally nomme ces ressources stylistiques les « soulignements expressifs²⁸ », c'est-à-dire des sortes de mises en relief qui possèdent des fonctions expressives propres. Ces « soulignements » sont détectables par les lecteurs et ainsi saisissables dans tous les textes, qu'ils soient littéraires ou non.

L'école psychologique, en revanche, propose une autre approche pour étudier les textes. L'un des pionniers les plus connus de cette école est sans doute Léo Spitzer. La méthode que propose ce linguiste s'appelle le « cercle philologique²⁹ » (Ullmann, 1966, p. 26). Elle met davantage l'accent sur les connexions entre l'aspect psychologique de l'auteur et son langage. Nous résumons brièvement cette démarche : lire un texte jusqu'à ce que le lecteur rencontre un phénomène étonnant dans le texte, que ce soit au niveau lexical, syntaxique ou sémantique ; ensuite, essayer de faire une analyse quasi-psychanalytique de l'auteur pour expliquer ce phénomène étonnant ; trouver enfin d'autres indices dans le texte pour confirmer cette analyse psychologique. La méthode de Spitzer étant fortement subjective, elle se distingue du courant du « génie » des deux manières suivantes : premièrement, elle part du principe que les traits linguistiques saillants dans le texte sont détectables par le lecteur ; deuxièmement, elle se base sur le fait que ces traits linguistiques sont liés à la psychologie de l'auteur. Ainsi, elle met en relation les traits linguistiques et l'intention de l'auteur, la dernière étant impossible à saisir pour le courant du « génie ».

Nous avons présenté brièvement deux courants stylistiques contradictoires dans cette partie : le courant du « génie » met en relief le lien intrinsèque entre le style et le langage, tandis que le courant de l'expressivité s'y oppose de deux manières : le structuralisme propose une nouvelle méthode qui se concentre sur les ressources

²⁸ *Idem.*

²⁹ Ullmann, Stephen. 1966. *Language and Style: Collected Papers*. Language and Style Series 1. Oxford: B. Blackwell, p. 26.

formelles et structurales d'un texte ; le courant psychologique propose de réconcilier le « génie » du langage avec les traits linguistiques saillants détectés par le lecteur.

Nous pensons quant à nous que le « courant de l'expressivité » est le plus riche et le plus pertinent pour le cadre qui est le nôtre. La stylométrie, une branche outillée de la stylistique, peut difficilement adopter la théorie de « génie » du langage, parce que cette théorie ne fournit pas de méthode opérationnelle. Le structuralisme et la stylistique psychologique, bien que les deux aient fait l'objet de critiques au cours de l'histoire, nous sont donc précieux. Certes, nous n'allons pas adopter les mêmes méthodes que Spitzer et Bally pour analyser notre corpus, mais les idées mises en avant par ces deux courants nous serviront de base théorique, parce que notre point de départ vise à pouvoir détecter les traits linguistiques saillants dans les textes, alors que le corps de notre thèse est de produire des analyses qualitatives à partir de ces traits caractéristiques.

2.2.2 Le style « parfait » et la singularité stylistique

En abandonnant la notion du « génie », nous abordons maintenant une autre question importante pour la stylistique : dans quelle mesure un style particulier se distingue-t-il des autres ? Pour bien élucider et répondre à cette question, il nous semble impératif d'évoquer la notion de « singularité stylistique ». Les deux courants évoqués ici s'opposent en effet sur cette notion de singularité : certains croient à l'idée d'un style parfait, un style universellement bon que les auteurs pourraient atteindre après de laborieux travaux (Flaubert et Maupassant incarneraient cette recherche du style parfait dans la langue française), tandis que d'autres valorisent la singularité de chaque style personnel, en affirmant qu'un style ne peut s'apprécier que par ses qualités et ses particularités, forcément différentes de celles d'un autre auteur.

C'est en France, au 19^e siècle qu'a eu lieu le grand débat du « style parfait » selon Gilles Philippe³⁰ (Philippe, 2013). Le débat se lance autour du style de Flaubert, qui, de l'avis général, « écrivait bien »³¹ (Philippe, 2013). En effet, il est connu que Flaubert consacrait beaucoup d'efforts à corriger ses manuscrits, afin d'éviter notamment les mauvaises répétitions qui alourdissent le style. Émile Zola dans son *Gustave Flaubert écrivain* a remarqué ce soin obsessionnel de Flaubert dans sa prose :

« Il poursuit les mots répétés jusqu'à trente ou quarante lignes de distance. Il se donne un mal infini pour éviter les consonances fâcheuses, les redoublements de syllabe offrant quelque dureté. Surtout, il proscriit les rimes, les retours de fin de phrase apportant le même son ; rien ne lui semble gêner autant un morceau de style³². » (Zola, 1895, p. 506)

Contrairement à ce que croyait Victor Hugo, le style pour Flaubert n'est pas « dans la racine des cheveux »³³ (Hugo, 1985 [1901], p. 622) de l'auteur : le style, au contraire, est une chose que l'auteur doit perfectionner par son travail. La conception du style, chez Flaubert, manifeste une certaine modernité parce qu'il ne considère pas le style comme un objet abstrait, insaisissable, mais plutôt comme un outil que chaque auteur doit travailler pour polir et perfectionner son texte. Il l'exprime ainsi : « J'ai en moi, et très net, il me semble, un idéal (pardon du mot), un idéal de style, dont la poursuite me fait haleter sans trêve³⁴ » (Flaubert, 1927 [1857]). Si nous admettons avec Flaubert que le style est une technique, il est logique de considérer que cette technique peut être travaillée. C'est exactement l'opinion de Guy de Maupassant, disciple de Flaubert, quand il écrit la préface de *Pierre et Jean* :

³⁰ Philippe, Gilles. 2013. *Le rêve du style parfait*. 1 vol. Paris : Presses universitaires de France.

³¹ *Idem*.

³² Zola, Émile. 1895. *Les romanciers naturalistes*. Paris, France : Bibliothèque-Charpentier, p. 506.

³³ Hugo, Victor. 1985 [1901]. *Op.cit.*, p. 622.

³⁴ Flaubert, Gustave. 1927 [1857]. « à Ernest Feydeau, Croisset, 06 août 1857 » dans *Correspondance*. Paris, France : L. Conard.

« Quelle que soit la chose qu'on veut dire, il n'y a qu'un mot pour l'exprimer, qu'un verbe pour l'animer et qu'un adjectif pour la qualifier. Il faut donc chercher, jusqu'à ce qu'on les ait découverts, ce mot, ce verbe et cet adjectif, et ne jamais se contenter de l'à peu près, ne jamais avoir recours à des supercheries, même heureuses, à des clowneries de langage pour éviter la difficulté³⁵. » (Maupassant, 2019 [1888])

Le fait que l'auteur ait la possibilité de travailler le style pour efficacement exprimer ses idées sous-entend la confirmation que la singularité d'un style est un produit ultime de perfection. Ainsi est née la notion de style parfait, le style référentiel, sans empreinte personnelle de l'auteur : c'est ce que constate Gilles Philippe sur le style de Maupassant³⁶. Ce dernier est, dès le 19^e siècle, considéré comme un héritier de Flaubert à ce sujet : « La place de Guy de Maupassant dans la mythographie littéraire nationale s'est donc fixée très tôt : c'est celle d'un écrivain "sans style", disais-je, mais c'est aussi celui de l'héritier de Flaubert³⁷ » (Philippe, 2013). Philippe va plus loin : il résume et formalise les idées sur le style que partagent Flaubert et son disciple :

« Il y a une spécificité de la prose comme il y a une spécificité du vers ; l'art est impersonnel, et le style ne doit pas laisser entendre une voix ; on ne saurait séparer l'idée et son expression, mais c'est la seconde qui importe d'abord à l'écrivain, et la qualité de l'idée se mesure à celle de la forme³⁸. » (Philippe, 2013, p. 32)

³⁵ Maupassant, Guy de, et Noëlle Benhamou. 2019 [1888]. « Préface » dans *Pierre et Jean*. Paris : Flammarion.

³⁶ Malgré la volonté, qui s'exprime parfois chez Maupassant, de valoriser la diversité dans le style : « Un critique intelligent devrait, au contraire, rechercher tout ce qui ressemble le moins aux romans déjà faits, et pousser autant que possible les jeunes gens à tenter des voies nouvelles³⁶ ». *Idem*.

³⁷ Philippe, Gilles. 2013. *Op. cit.*, p. 32.

³⁸ *Idem*.

Nous revenons inévitablement à la question du fond et de la forme en stylistique : si Hugo, comme nous l'avons évoqué dans la sous-partie précédente, pense que les deux aspects n'en font qu'un. Flaubert met davantage l'accent sur la forme. Les idées de Flaubert sur le style sont anti-aristotéliennes, parce que si Aristote pense que nous ne pouvons pas « parler pour le plaisir de parler³⁹ » (Aristote, 2008 [IV^e siècle av. J.-C.]) et que nous ne pouvons pas non plus parler du non-être comme le « bouc-cerf⁴⁰ », pour Flaubert, l'autorité du fond est récusée et l'importance de la forme est reconnue.

Un auteur « sans style » mériterait-il la couronne du grand écrivain ? La réponse à cette question est paradoxale parce que la poursuite du style parfait fait souvent naître des styles particuliers. Roland Barthes a justement constaté ce paradoxe en affirmant que « la phrase de Balzac n'est pas cet objet incroyablement reconnaissable qu'est la phrase de Flaubert⁴¹ » (Cahen, 1966, p. 436). En effet, malgré la recherche par Flaubert d'un « style parfait », qui serait donc dès lors « impersonnel », le style de Flaubert est reconnaissable, et donc singulier, au sein de la littérature française.

La contradiction entre la recherche d'un style parfait, et ainsi universellement bon de Flaubert et la singularité du style flaubertien que constataient les stylisticiens tels que Roland Barthes nous amène à réexaminer la notion de singularité dans la stylistique. Jean Starobinski revient sur cette contradiction entre la singularité et l'universalité du style en tissant un lien entre ces deux notions contradictoires :

³⁹ Aristote, et Myriam Hecquet-Devienne. 2008 [IV^e siècle av. J.-C.]. *Métaphysique gamma : édition, traduction, études*. Édité par Annick Stevens. Louvain-la-Neuve, Belgique, France, États-Unis d'Amérique.

⁴⁰ Pour Aristote, c'est le fond qui définit la forme : la signification est unidirectionnelle, du fond à la forme, si le « bouc-cerf » n'existe pas, nous ne pouvons pas le dire. Ainsi, le travail sur la forme du texte est moins important que celui sur le fond. Aristote oppose son idée aux démarches des sophistes, qui d'après lui, cherchent à tout prix à convaincre par les lieux rhétoriques dans leurs discours sans se soucier du fond ni de la vérité. *Idem*.

⁴¹ Cahen, D. 1966. *Entretien avec Roland Barthes*. S.l. : Aletheia, p. 436.

« Le style n'est donc ni le particulier pur, ni l'universel, mais un particulier en instance d'universalisation, et un universel qui se dérobe pour renvoyer à une liberté singulière⁴² » (Starobinski, 1970, p. 56).

Cette citation de Starobinski met en avant deux points cruciaux pour comprendre les notions de singularité et universalité : 1. Le style singulier d'un auteur pourrait être universalisé. Nous constatons que certaines techniques stylistiques de Flaubert sont encore de rigueur de nos jours, telles que la suppression des répétitions ou la coupe oratoire (discutée en détail au Chapitre 5). 2. Malgré cette universalisation de certaines techniques stylistiques, les styles singuliers ont toujours le moyen de se distinguer dans la littérature, Gilles Philippe constate ainsi que « depuis le XIXe siècle, le mot style désigne prioritairement la variante personnelle de la langue commune⁴³. » (Philippe, 2013, p. 145)

Nous avons présenté dans cette partie la contradiction entre la poursuite du style parfait de Flaubert et la singularité de son style dans la littérature française. Selon nous, la vraie contradiction réside entre le style individuel et la norme linguistique historique de l'époque, ou entre différentes œuvres ou différents auteurs. Nous pouvons citer ici un passage de Halliday qui parle du « style reconnaissable » : « S'il existe un style reconnaissable, qu'il s'agisse d'une œuvre, d'un auteur, d'une période entière ou d'une tradition littéraire, sa qualité distinctive peut en dernière analyse être exprimée en matière de fréquences relatives⁴⁴ » (Halliday, 2020 [1969], p. 343). Halliday ne précise pas de quelle manière et par rapport à quelle référence les fréquences pourraient être

⁴² Spitzer, Leo, Jean Starobinski, Éliane Kaufholz-Messmer, Alain Coulon, et Michel Foucault. 1970. *Études de style /Précédé de Leo Spitzer et la lecture stylistique*. Paris: Gallimard, p. 56.

⁴³ Philippe, Gilles. 2013. *Op. cit.*, p. 145.

⁴⁴ Halliday, M. 2020 [1969]. « Linguistic function and literary style: an inquiry into the language of William Golding's *The Inheritors* ». In : Chatman, Seymour Benjamin, éd. 1971. *Literary style: a symposium*. London, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, Etats-Unis d'Amérique, p. 343.

relatives, mais l'idée du relativisme, c'est-à-dire le rejet d'une référence absolue, est bien exprimée.

Les comparaisons stylistiques doivent, selon nous, se baser sur la singularité du style d'un texte par rapport aux autres textes ou sur la singularité du style d'un auteur par rapport à d'autres auteurs. Par nature, le travail d'analyse stylistique commence par la forme du texte et ensuite creuse le fond : en conséquence, nos propres démarches stylistiques se fondent sur des corpus où les analyses stylistiques sont faites soit en comparant des œuvres, soit en comparant des auteurs. Nous essayons de montrer qu'en étudiant des traits formels, de nature syntaxique (corrigibles et perfectibles selon Flaubert), il est possible de révéler des différences stylistiques, par une approche comparative entre œuvres et entre auteurs.

2.2.3 Stylistiques psychologique et structuraliste

Dans la première sous-partie, nous avons opposé les théories stylistiques de Bally et de Spitzer à l'idée d'une stylistique fondée sur la notion du « génie ». Nous allons explorer dans cette sous-partie la dispute qui opposa Spitzer et Bally eux-mêmes, ou plus généralement, la contradiction entre la stylistique psychologique et la stylistique structuraliste. Nous présentons ensuite un autre théoricien, Michael Riffaterre, qui a, à son tour, adopté un point de vue faisant une synthèse entre stylistique structuraliste et psychologique. La notion à la base de la confrontation qui nous occupe est en effet la subjectivité : tandis que la stylistique structuraliste essaie d'éliminer la subjectivité, le courant psychologique se base essentiellement sur l'interprétation subjective du texte. L'étude comparée de ces deux courants nous permet de mettre au jour la façon d'aborder et d'intégrer la notion de subjectivité en stylistique.

Nous avons mentionné dans la première sous-partie que la stylistique de Léo Spitzer est une approche psychologique, parce que son objectif est de révéler l'état

psychologique de l'écrivain par le truchement des singularités langagières de son texte. Sans doute influencé par Croce et Vossler⁴⁵ (Ullmann, 1966, p. 26), Spitzer pense que le style littéraire (contrairement au style de la langue parlée) est le paradigme parfait de la créativité et de la fonction d'auto-expression artistique du langage : le langage littéraire est auto-expressif parce qu'il représente l'esprit et le tempérament de l'auteur.

La stylistique psychologique a reçu de nombreuses objections. Les démarches d'analyse stylistique précisées dans *Linguistics and Literary History* (1948) de Léo Spitzer ont rapidement semblé contestables aux stylisticiens. D'abord, rappelons-nous que la première étape de la méthode de Spitzer est de relever les traits singuliers en lisant attentivement le texte (*close reading*), mais il n'existe aucun critère formel pour la détection de ces traits singuliers. Le processus dépend uniquement du lecteur : aucune justification n'est demandée ; ensuite, l'analyse de ces phénomènes singuliers vise essentiellement une psychanalyse de l'auteur. Spitzer a lui-même appliqué cette méthode au style de Diderot : en repérant le rythme passionnel des phrases dans *Le Neveu de Rameau* de Diderot, Spitzer pense que le neveu de Rameau représente, d'une manière caricaturale, le tempérament de Diderot lui-même⁴⁶ (Spitzer, 1948). Nous n'allons pas argumenter sur la pertinence de cette analyse psychanalytique, mais il est impossible de nier le caractère subjectif de cette méthode, que Spitzer appelle le « cercle philologique ».

De l'autre côté, nous trouvons la stylistique structuraliste de Charles Bally. La stylistique structuraliste « recenserait les divers moyens dont disposent les locuteurs pour exprimer une idée ou une vision⁴⁷ » (Philippe, 2013, p. 147). Nous citons la définition de la stylistique de Bailly dans son *Précis de stylistique* en 1905 :

⁴⁵ Ullmann, Stephen. 1966. *Language and Style: Collected Papers*. Language and Style Series 1. Oxford: B. Blackwell, p. 26.

⁴⁶ Spitzer, Leo. 1948. *Linguistics and literary history: essays in stylistics*. Princeton, États-Unis d'Amérique: Princeton University Press, Ch. IV.

⁴⁷ Philippe, Gilles. 2013. « Les deux corps du style ». *Les Temps Modernes*, 5(5), P. 147.

« Elle (la stylistique) observe les rapports qui existent dans une langue donnée entre les choses à exprimer et leur expression : elle cherche à déterminer les lois et les tendances que suit cette langue pour arriver à l'expression de la pensée sous toutes ses formes. Elle recherche enfin une méthode propre à faire découvrir ces moyens d'expression, à les définir, à les classer, et à en montrer l'emploi⁴⁸. » (Bally, 1905, p. 7)

Contrairement à Spitzer, Bally pense que la stylistique ne peut pas se concentrer uniquement sur la langue littéraire, elle gagnerait à étudier la langue en général, y compris la langue parlée. En repérant les traits linguistiques saillants (appelés les « soulignements expressifs⁴⁹ ») qui pourraient porter des fonctions stylistiques, Bally propose d'attribuer à ces traits formels les fonctions stylistiques correspondantes, empruntant la notion de « fonction formelle » au courant formaliste russe et saussurien⁵⁰ (Dmitriev, 2002) qui considère le langage comme un système de formes plutôt qu'une matière propre. Son approche met l'accent sur l'aspect linguistique (c'est-à-dire sur des marques explicites dans le texte) : elle s'oppose à la stylistique d'auteur (c'est-à-dire la stylistique qui remonte à l'origine de la production du texte, avec une dimension subjective assumée) représentée par Léo Spitzer.

Plus récemment, mais dans le sillage de cette polémique, le rôle de Michel Riffaterre dans l'histoire de la stylistique nous semble particulièrement intéressant, car il essaie de réconcilier, d'une certaine manière, les stylistiques structuraliste et psychologique. Dans son *Essais de stylistique structurale*, il donne sa propre définition du style :

⁴⁸ Bally, Charles. 1905. *Précis de stylistique : esquisse d'une méthode fondée sur l'étude du français moderne*. Genève, Suisse: A. Eggimann, p. 7.

⁴⁹ *Idem*.

⁵⁰ Dmitriev, Aleksander. 2002. « Le contexte européen (français et allemand) du formalisme russe ». *Cahiers du monde russe*. Russie - Empire russe - Union soviétique et États indépendants 43 (43/2-3): 423-40.

« Le style est compris comme un soulignement (*emphasis*) ajouté à l'information transmise, sans altération de sens. [...]. Le style est la mise en relief qui impose certains éléments de la séquence verbale à l'attention du lecteur, de telle manière que celui-ci ne peut les omettre sans mutiler le texte et ne peut les déchiffrer sans les trouver significatifs et caractéristiques (ce qu'il rationalise en y trouvant une forme d'art, une personnalité, une intention, etc.). Ce qui revient à dire que le langage exprime et que le style met en valeur⁵¹. » (Riffaterre, 1971, pp. 30-31)

Nous considérons que la stylistique de Riffaterre opère une soudure entre Bally et Spitzer pour les deux raisons suivantes.

Premièrement, en prenant en compte les traits structuraux du langage, Riffaterre renonce à une explication psychologique de ces traits, ou plus précisément, il refuse la stylistique d'auteur en mettant l'accent sur le texte lui-même ainsi que sur la réception du texte par les lecteurs. Ceci est important parce que si Bally et Spitzer cherchent les singularités dans les textes littéraires pour soit révéler les fonctions linguistiques et expressives de telle ou telle expression, soit faire une analyse psychanalytique de l'auteur, Riffaterre pense que les significations et les caractéristiques des « emphasis » doivent être perçues et interprétées par les lecteurs. Il inaugure une « stylistique de réception⁵² » (Philippe, 2005) et se débarrasse ainsi de la « stylistique de production⁵³ » (Philippe, 2005).

Deuxièmement, nous constatons que la stylistique de Riffaterre est essentiellement structuraliste, comme celle de Bally, dans le sens où Riffaterre propose également une théorie permettant de formaliser les phénomènes linguistiques. Sa démarche consistant à séparer le fond de la forme (afin d'étudier les effets donnés par la forme sur le texte)

⁵¹ Riffaterre, Michael, et Daniel Delas. 1971. *Essais de stylistique structurale*. Paris, France : Flammarion, pp. 30-31.

⁵² Philippe, Gilles. 2005. « Traitement stylistique et traitement idiolectal des singularités langagières ». *Cahiers de praxématique*, n° 44 (janvier).

⁵³ *Idem*.

est la même que celle de Bally. Cependant, nous constatons que cette théorie se base aussi sur la détection subjective de traits singuliers par le lecteur : Riffaterre ne propose donc pas une méthode radicalement différente de celle de Spitzer en ce qu'elle conserve la détection subjective de traits singuliers. La stylistique de Riffaterre est formaliste, mais elle partage au moins une chose avec la stylistique de Spitzer : celle de privilégier les phénomènes linguistiques saillants qui suscitent l'attention du lecteur. Ainsi, Riffaterre trouve selon nous un bon équilibre entre objectivité et subjectivité dans ses démarches stylistiques.

Comme précédemment, nous avons présenté ici la contradiction entre la stylistique structuraliste et la stylistique psychologique, afin de pouvoir mieux situer notre propre démarche. Une limite de la stylistique structurale est d'être trop formaliste : à chaque phénomène linguistique saillant dans les textes, il faudrait pouvoir associer une fonction formelle invariable, autrement dit un effet stylistique particulier. Une limite de la stylistique psychologique est au contraire sa tendance excessivement subjective : les résultats dépendent en effet uniquement de l'observation et de l'interprétation d'un lecteur particulier ; ce résultat peut toujours être remis en cause. Ce qui semble singulier à un lecteur peut être tout à fait normal à un autre, alors que toutes les interprétations psychologiques sont permises selon ce qui a suscité l'attention d'un stylisticien particulier. La stylistique de Riffaterre, comme nous l'avons montré, peut de notre point de vue être considérée comme une réconciliation entre ces deux pôles. Riffaterre combine en effet les deux méthodes, en valorisant la surreprésentation de certains traits linguistiques et en cherchant la signification psychologique de ces traits. Nous situons en effet notre propre méthode dans la lignée de celle de Riffaterre, sans adopter aveuglément l'ensemble de sa théorie stylistique. Nous proposons une méthode combinant l'extraction supervisée et non-supervisée de traits formels, afin de trouver la bonne dose de subjectivité dans notre démarche stylistique.

2.2.4 Le style de texte : le texte et son contexte

Nous abordons ici une autre question cruciale pour notre objet d'étude : le texte ou, plus précisément, la question suivante : faut-il limiter les analyses stylistiques au texte lui-même, ou faut-il prendre en considération le contexte social et historique ? Il nous semble d'autant plus important de poser cette question que la stylométrie, bien que portant par essence sur le texte lui-même, fait preuve d'une grande hésitation entre les deux approches. Tandis que le New Criticism et les premiers essais de stylométrie cherchent à isoler le texte de ses contextes afin de se concentrer sur le texte lui-même, des linguistes comme François Rastier proposent une théorie systématique permettant de relier l'étude du texte avec le contexte, via la stylométrie. Par le truchement de la théorie linguistique de Rastier, nous évoquerons également les deux fonctions de la stylistique : la fonction d'identification (l'identification des styles littéraires) et la fonction de caractérisation (la définition des qualités du style), qui seront reprises dans notre propre démarche.

Si on parle d'exclure le contexte et en particulier la notion d'auteur de l'étude stylistique, il est inévitable d'évoquer le New Criticism, un courant à la source de la stylistique moderne, selon Fowler⁵⁴ (Fowler, 1981, p. 11) (les autres influences sont les linguistiques modernes américaine et européenne, et le structuralisme français, toujours selon cet auteur). Le New Criticism est qualifié de « nouveau » et s'oppose à la critique littéraire du début du 20^e siècle en Angleterre, qui, par une analyse pratique et philologique des œuvres littéraires, cherche à pratiquer une sorte de Close Reading sur les traits saillants des textes par rapport aux normes grammaticales standards⁵⁵ (Carter, 2008, p. 292). Selon Carter, ce courant critique anglais s'est ensuite développé aux États-Unis avec quelques modifications. « Là où le premier [le courant critique

⁵⁴ Fowler, Roger. 1981. *Literature as Social Discourse: The Practice of Linguistic Criticism*. Bloomington: Indiana university press, p. 11.

⁵⁵ Carter, Ronald, et Peter Stockwell. 2008. éd. *Op. cit.*, p. 292.

anglais du début du 20^e siècle] plaçait l'interprétation par le lecteur en premier avec la lecture attentive, le New Criticism se concentrait sur les mots eux-mêmes⁵⁶ » (Carter, 2008, p. 292). Ce courant stylistique se concentre sur la saillance des traits linguistiques, sans se soucier de l'intention de l'auteur. Naturellement, cette démarche exclut complètement les analyses psychologiques que nous avons évoquées dans les sections 2.2.1 et 2.2.3 :

« Pour ces auteurs [les stylisticiens du New Criticism], le texte littéraire (dont le court poème lyrique semble être le paradigme) est un artefact verbal autonome, une structure unique du langage. Son mode d'existence est linguistique et non historique ; il doit être étudié comme un ensemble de patrons verbaux, non comme le produit des forces sociales ou de la psychologie d'un auteur⁵⁷. » (Fowler, 1981, p. 12)

Ainsi, le New Criticism isole-t-il le texte de son contexte historique et social. L'auteur est comme absent, il n'est pas pris en compte par ce courant critique. Nous constatons le côté idéaliste de cette démarche. D'après notre propre définition (rappelée au début de ce chapitre), le texte est une manifestation de l'auteur, qui cherche à verbaliser sa conception littéraire. Pourtant, le New Criticism n'a pas été abandonné, la présence de cette méthode se manifeste en particulier dans la stylométrie outillée. Comme nous le verrons, la plupart des essais stylométriques pourraient être classés dans cette catégorie : ils se concentrent sur les traits linguistiques saillants dans un texte, sans fournir d'explication ni sur l'intention de l'auteur, ni sur les effets créés chez les lecteurs. L'unique et principale mission de ces recherches stylométriques est, selon nous, de

⁵⁶ *Idem.*

⁵⁷ « For these writers, the literary text (of which the short lyric poem seems to be the paradigm) is a self-contained verbal artefact, a unique structure of language. Its mode of existence is linguistic, not historical; it is to be studied as a complex of integrated verbal patterns, not as the product of social forces or of the psychology of an author ». Fowler, Roger. 1981. *Op. cit.*, p. 12.

repérer statistiquement les traits saillants d'un texte, il s'agit sans doute d'une version automatisée de la méthode du New Criticism par les outils informatiques et statistiques.

Si le New Criticism ne prend en compte que le texte, avec Riffaterre, nous découvrons une théorie qui relie les notions d'auteur, de texte et de lecteur. Cet auteur propose une théorie englobante qui met en évidence les enjeux stylistiques :

« *Le style est l'homme même*. Buffon a tout dit et son axiome est une vérité d'évidence. Mais si tout est dit, tout reste à expliquer, car la transmutation d'un être pensant en écriture ne s'accomplit qu'au prix de deux opérations paradoxales, l'une réconciliant l'énonciation et l'énoncé, l'autre développant un idiolecte textuel sans aller jusqu'à rendre impossible son déchiffrement à partir du sociolecte⁵⁸. » (Riffaterre, 1994, p. 283)

Nous avons évoqué dans la sous-partie précédente le fait que Riffaterre a pour ambition de formaliser les études linguistiques et stylistiques. Explorons maintenant plus en détail cette citation. D'abord, la phrase « le style est l'homme même » est peut-être la phrase la plus citée de Buffon (au moins dans le domaine de la stylistique), mais Riffaterre réfute cette idée trop subjective du style d'auteur. Il argumente que :

« Le texte fonctionne comme le programme d'un ordinateur pour nous faire faire l'expérience de l'unique. Unique auquel on donne le nom de style, et qu'on a longtemps confondu avec l'individu hypothétique appelé auteur : en fait, *le style, c'est le texte même*⁵⁹. » (Riffaterre, 1979, p. 8)

⁵⁸ Riffaterre, Michael, « L'inscription du sujet », in Molinié, Georges Organisateur de réunion, Pierre Organisateur de réunion Cahné, Colloque international Qu'est-ce que le style ?, et École doctorale Sciences du langage. 1994. *Qu'est-ce que le style ? : actes du colloque international*. Paris, France : Presses universitaires de France, p. 283.

⁵⁹ Riffaterre, Michael. 1979. *La production du texte*. Paris, France : Éditions du Seuil, p. 8.

Cette citation est trompeuse, en ce qu'elle pourrait suggérer que Riffaterre refuse d'intégrer la notion d'auteur à l'étude du style. Mais ce n'est pas le cas : Riffaterre s'oppose aux démarches qui partent du texte pour remonter à l'auteur en proposant ainsi une analyse trop subjective et psychologique. Cette deuxième citation marque plutôt la rupture de Riffaterre avec Spitzer (comme nous l'avons présentée dans la section précédente). Ces deux citations nous semblent utiles pour donner une vue plus claire de sa théorie.

En mettant l'accent sur le texte (et non sur l'auteur), Riffaterre propose de considérer le texte comme une « réconciliation » de « l'énonciation ⁶⁰ » et de « l'énoncé⁶¹ ». Il soutient aussi, toujours dans la première citation, que le style doit être considéré comme un déchiffrement des idiolectes « à partir du sociolecte⁶² » (Rastier, 2001, p. 178-179)). Riffaterre affirme que le texte se place dans un contexte social et extérieur qui se manifeste par les idiolectes d'auteur. Nous pouvons maintenant développer cette théorie en reformulant cette idée : le texte intègre les idiolectes de l'auteur, à savoir les traits statistiquement suremployés ou sous-employés, qui peuvent donner lieu à une interprétation stylistique (pour Riffaterre, cette tâche d'interprétation se réalise par les lecteurs et ne devrait pas renvoyer à la psychologie de l'auteur), et l'interprétation de ces idiolectes doit se réaliser par la prise en charge des sociolectes. Nous voyons que même si Riffaterre met l'accent sur le texte comme objet d'étude de la stylistique, il combine en fait la stylistique de production (par l'affirmation des idiolectes d'auteur) et la stylistique de réception (par la tâche qu'il attribue aux lecteurs de déchiffrer ces idiolectes d'auteur).

⁶⁰ « Acte de production individuelle du discours dans des circonstances données ». Le Robert, <https://dictionnaire.lerobert.com/definition/enonciation>, consulté le 5 août, 2021.

⁶¹ « Segment de la chaîne parlée produit par un seul locuteur et situé entre deux silences ». <https://www.cnrtl.fr/definition/enonce>, consulté le 5 août, 2021.

⁶² Riffaterre n'a jamais précisé les définitions d'idiolecte et de sociolecte, nous empruntons ici la définition donnée par François Rastier, qui nous semble pertinente dans ce contexte sémantique : considérons « les normes sociales à l'œuvre dans tout texte. On peut nommer *sociolectes* les types de discours instaurés par ces normes. [...] On peut appeler *idiolecte* l'ensemble des régularités personnelles ou "normes individuelles" dont elles témoignent ». Rastier, François. 2001. *Op. cit.*, p. 178-179.

Nous présentons maintenant une étude plus systématique de la théorie de François Rastier qui prolonge, nous semble-t-il, le New Criticism et la théorie énonciative de Riffaterre. Rastier dans *Arts et sciences du texte* (2001) présente son point de vue sur ce que la stylistique devrait être. Il présente d'abord la stylistique comme héritage de la rhétorique. Il explore ensuite la notion d'idiolecte, en distinguant trois systèmes linguistiques qui touchent directement les analyses stylistiques :

« (i) Le plus rigoureux (système linguistique) reste celui du système fonctionnel de la langue, qui impose, croit-on, ses règles à tout usage ; sans préjuger de l'homogénéité de ce système, on peut le nommer *dialecte*. (ii) Viennent ensuite les normes sociales à l'œuvre dans tout texte. On peut nommer *sociolectes* les types de discours instaurés par ces normes. Un sociolecte correspond à une pratique sociale (ex. judiciaire, politique, religieuse). Chacun a son lexique propre structuré en domaines sémantiques, et s'articule en divers genres textuels (ex. la plaidoirie, la question publique, l'homélie). (iii) Enfin, chaque usage de la langue est immanquablement marqué par les dispositions particulières du prétendu "émetteur" : sans présumer qu'elles fassent système, on peut appeler *idiolecte* l'ensemble des régularités personnelles ou "normes individuelles" dont elles témoignent⁶³. »
(Rastier, 2001, p. 178-179)

La stylistique, selon Rastier, « gagnerait à se fonder sur une théorie générale des idiolectes, qu'ils soient littéraires ou non⁶⁴ » (Rastier, 2001, p. 180). La théorie de Rastier pourrait donc être considérée comme une réconciliation entre le New Criticism et Riffaterre.

⁶³ *Idem.*

⁶⁴ Rastier, François. 2001. *Op. cit.*, p. 180.

En effet, Rastier refuse vivement d'isoler le texte de son contexte. « Admettons qu'un texte se rattache à une langue par son dialecte, à un sociolecte par son genre et son discours, à un idiolecte par son style⁶⁵ » (Rastier, 2001, p. 179). Il est ainsi impossible d'analyser le contenu d'un texte sans prendre en compte les facteurs extérieurs. Mais, dans le même temps, son intention de ramener la stylistique à « une théorie générale des idiolectes » se rapproche de ce que cherche à faire le New Criticism : rappelons-nous que l'idiolecte est également « le style sans la signification⁶⁶ » (Philippe, 2005, p. 82). Ainsi, le but de Rastier est de proposer une stylistique linguistique qui se concentre sur les traits idiolectaux, sans signification stylistique hors de leur contexte, mais avec une signification propre quand ils sont réinsérés dans les trois systèmes linguistiques que nous venons de citer.

Rastier renonce aussi à l'idée de combiner la stylistique de production et la stylistique de réception. « Si l'on identifie ce caractère du style, une perspective subjectivante peut le rapporter à l'auteur et l'expliquer par sa biographie psychologique, alors qu'une perspective objectivante le rapporte à des formes textuelles particulières. Nous choisissons la seconde, car nous avons à expliquer les œuvres en termes d'œuvres⁶⁷ » (Rastier, 2001, p. 184). Le point commun entre Riffaterre et Rastier est que tous les deux refusent la stylistique de production, c'est-à-dire la démarche qui cherche à expliquer l'œuvre d'après la psychologie de l'auteur. Mais il y a toutefois une différence radicale entre les deux auteurs : Riffaterre ne pense qu'à la fonction énonciative des idiolectes dans le texte. Pour lui, tous les idiolectes doivent être interprétés par les lecteurs, alors que Rastier distingue deux fonctions de ces idiolectes : la fonction d'identification (c'est-à-dire la fonction qui sert à identifier et à différencier une œuvre des autres) et la fonction de caractérisation (la fonction qui permet d'apprécier la qualité du style). La première a fait naître « maints travaux de linguistique quantitative »⁶⁸ (Rastier, 2001, p. 186) et la deuxième se dirige vers « une

⁶⁵ Rastier, François. 2001. *Op. cit.*, p. 179.

⁶⁶ Philippe, Gilles. 2005. *Op. cit.*, p. 82.

⁶⁷ Rastier, François. 2001. *Op. cit.*, p.184

⁶⁸ *Ibid.*, p.186.

élaboration stylistique »⁶⁹ (Rastier, 2001, p. 186), c'est-à-dire, la caractérisation des styles à travers les qualités stylistiques des idiolectes dans leur contexte. Ces deux fonctions sont en fait interchangeables : « par exemple, le jeu dramatique de Perec avec la fréquence des lettres, dans *La disparition* en premier lieu. La fréquence anormale d'une lettre, en occurrence zéro, est alors promue au rang de trait spitzérien, entendu comme principe organisateur⁷⁰ » (Rastier, 2001, p. 186). « Le rang de trait spitzérien » désigne la fonction de caractérisation que possèdent ces traits : Rastier oppose la stylistique de Spitzer à la stylistique purement identificatoire par les statistiques. Les deux fonctions dont parle Rastier rendent compte du fait qu'un idiolecte n'a pas de signification en dehors de son contexte. Nous pouvons ainsi conclure sur les théories des deux auteurs : Riffaterre cherche dans les textes le style lui-même, à savoir les idiolectes qui peuvent faire l'objet d'une interprétation qualitative par les lecteurs, tandis que Rastier cherche dans les textes les mêmes idiolectes, mais essaie de leur attribuer une interprétation à la fois quantitative et qualitative. Ainsi, Rastier, par sa théorie de la stylistique linguistique, construit un pont entre le New Criticism et la démarche énonciative de Riffaterre.

Nous avons présenté dans cette section la confrontation de deux courants stylistiques accordant un statut particulier à la notion de texte même : le New Criticism ainsi que quelques premiers essais de la stylométrie qui essaient d'exclure tout ce qui est en dehors du texte, par opposition à Riffaterre et Rastier qui abordent le problème du style en intégrant le contexte historique et social. Nous en concluons que, pour nous : (i) il est impossible d'isoler le texte de son contenu social et historique, les études stylistiques doivent tenir compte du contexte ; (ii) il faut renoncer à la stylistique de production qui a tendance à tomber dans une subjectivité dangereuse pour la stylométrie ; (iii) les traits saillants dans les textes sont utiles non seulement pour les analyses qualitatives, mais aussi pour les analyses quantitatives qui cherchent à

⁶⁹ *Idem.*

⁷⁰ *Idem.*

identifier et à montrer les différences entre les œuvres ; (iv) les deux fonctions des traits saillants, identificatoires et caractéristiques, sont interchangeableables, selon l'usage de ces traits par les stylisticiens.

2.2.5 La question de la littérarité⁷¹

Après avoir consacré la sous-partie précédente à mettre en évidence le rôle incontournable du contexte dans les démarches stylistiques, une autre question se pose naturellement : qu'est-ce que la stylistique cherche à repérer dans le texte ? Existe-t-il un certain genre de texte idéal pour les démarches stylistiques ? C'est la notion de genre littéraire qui est ici en jeu : tandis que Charles Bally propose d'appliquer la méthode structuraliste en stylistique à tous les textes sans discrimination, les formalistes russes pensent que seuls les textes littéraires méritent de recevoir une analyse stylistique. Nous présenterons ensuite la théorie linguistique de Georges Molinié, qui étudie systématiquement la question de la littérarité.

Commençons par Charles Bally. Nous avons évoqué dans les sections précédentes que pour Bally, la stylistique doit se définir comme une étude de l'expressivité caractéristique de la langue en général :

« La stylistique étudie les moyens d'expression dont dispose une langue, les procédés généraux employés par elle pour rendre par la parole les phénomènes du monde extérieur aussi bien que les idées, les sentiments et en général tous les mouvements de notre vie intérieure⁷². » (Bally, 1905, p. 7)

⁷¹ Nous adoptons la définition de Roman Jakobson pour cette notion : « ce qui fait d'une œuvre donnée une œuvre littéraire ». Jakobson, Roman, et Tzvetan Todorov. 1977. *Huit questions de poétique*. Paris, France: Éditions du Seuil.

⁷² Bally, Charles. 1905. *op.cit.*, Genève, Suisse: A. Eggimann, p. 7.

D'après Colas-Blaise et al. (2016), Bally représente un courant stylistique qui « refuse l'étude des textes littéraires » :

« La stylistique s'est d'emblée définie ou manifestée comme une praxis : ce fut d'abord une étude des faits expressifs caractéristiques d'une langue et non d'un discours ou d'un texte littéraire ; c'était la position de Bally, qui refusait l'étude des textes littéraires du fait qu'ils présentent un état de langue artistique, c'est-à-dire pratiquement artificiel⁷³. » (Colas-Blaise et al., 2016)

Si Bally pense que les caractéristiques inhérentes aux textes littéraires sont artificielles et ne méritent pas d'études stylistiques, le Formalisme russe a une position opposée. Viktor Shklozsky est l'un des pionniers de ce courant linguistique, et il parle explicitement du rôle du langage littéraire par rapport au langage ordinaire. Selon ce courant stylistique, toute la littérature a un statut particulier (au sens où les textes y sont travaillés et ont une finalité esthétique), elle est ainsi distincte de la communication ordinaire. À partir de l'effort des formalistes russes pour faire de la critique littéraire une discipline scientifiquement fondée, Roman Jakobson a formulé le concept de « littérarité » : « ce qui fait d'une œuvre donnée une œuvre littéraire⁷⁴ » (Jakobson, 1977). Proposant une opposition fondamentale entre les usages littéraires et non-littéraires du langage, Jakobson pense que la littérarité était une caractéristique qui offre au lecteur un mode d'expérience spécial mettant en relief les caractéristiques formelles du texte. Cette école structuraliste se lance alors « à la recherche d'une certaine littérarité qui est supposée caractériser cette entité littéraire et qui est censée trouver son expression dans une langue littéraire distinctive, opposée à la langue ordinaire⁷⁵ » (Fowler, 1981, p. 20). Les unités formelles stylistiquement significatives sont, selon

⁷³ Colas-Blaise, Marion, et Claire Stolz. 2016. « Comment faire dialoguer stylistique et sémiotique ? Éléments pour une pensée de la frontière disciplinaire ». In *Stylistiques ?*, édité par Laurence Bougault et Judith Wulf, 99-110. Interférences. Rennes: Presses universitaires de Rennes.

⁷⁴ Jakobson, Roman, et Tzvetan Todorov. 1977. *Op.cit.*.

⁷⁵ Fowler, Roger. 1981. *Op. cit.*, p. 20.

eux, plus présentes dans la langue littéraire et plus susceptible de susciter l'attention du lecteur. La valorisation des textes littéraires comme objet d'étude de la stylistique est souvent contestée depuis cette époque, parce qu'elle met en place une discrimination en fonction des genres littéraires.

La notion de littérarité de Jakobson est souvent contestée. Comme le soulignent Marion Colas-Blaise et Claire Stolz, un texte, considéré comme non-littéraire quelques siècles avant, pourrait bien posséder un style particulier de cette époque, qui nous apparaît quand nous le lisons aujourd'hui⁷⁶ (Colas-Blaise et al., 2016). Par conséquent, la stylistique, ainsi que la critique littéraire en général, doivent souvent adopter un point de vue évolutif sur les textes.

Nous évoquons ici la notion de littérarité, mais il ne s'agit pas d'opposer textes littéraires et non-littéraires, et nous n'avons pas l'intention de réduire la stylistique à cette dichotomie. Néanmoins, la littérarité, comme les mises en reliefs caractéristiques du langage, est avant tout marquée par la production contrôlée d'effets stylistiques par l'exploitation subtile des phénomènes saillants du texte. Georges Molinié propose d'élargir cette définition. Dans *Approches de la réception* (1993), il propose une étude systématique de la littérarité. Selon lui, il existe trois types de littérarités : la littérarité générale, générique et singulière⁷⁷ (Molinié et al., 1993, p. 13). La première est celle dont parlait Jakobson, la littérarité qui distingue un texte littéraire des autres. La deuxième est celle des genres et la troisième est celle de chaque production littéraire individuelle. Ainsi, il existe trois niveaux de littérarité, permettant de différencier les types de discours, les genres littéraires et les textes individuels. Molinié met ensuite l'accent sur le rôle du récepteur pour déchiffrer ces trois littérarités quand il lit les textes :

⁷⁶ Colas-Blaise, Marion, et Claire Stolz. 2016. *Op. cit.*

⁷⁷ Molinié, Georges, et Alain Viala. 1993. *Approches de la réception*. Presses Universitaires de France, p. 13.

« On n'a sans doute pas besoin de souligner à quel point toutes ces considérations n'ont de sens que pour une stylistique de la réception : c'est le récepteur, on ne le répétera jamais assez, qui mesure, et fait exister, ce ressenti, variable, de l'identité et de la variation⁷⁸. » (Molinié et al., 1993, p. 33)

L'outil dont dispose le lecteur pour accéder aux littérarités d'un texte est ce qu'il nomme le « stylème » : « Le stylème, qui est l'objet apparent de la recherche en sémiostylistique, est une réalité langagière à valeur littéraire (un caractérisème de littérarité, avons-nous redit)⁷⁹ » (Molinié et al., 1993, p. 34). Le stylème, d'après cette définition est une fonction caractérisante des traits linguistiques, qui révèle les littérarités d'un texte. Dans ce sens-là, nous pouvons conclure, avec Molinié, que chacune des trois littérarités est l'ensemble de traits saillants d'un certain texte, qui le rendent particulier. Le stylème n'émerge pas d'une manière arbitraire, Molinié rappelle le rôle important des traits répétitifs et saillants dans cette démarche stylistique :

« La répétition est le plus puissant outil dont nous disposons ; c'est même le seul vraiment efficace. [...] Il peut s'agir, dans un ensemble *a priori* délimité, de l'itération d'un phonème (ou d'un graphème), d'une syllabe, d'une lexie, d'un syntagme, d'un ensemble phrastique entier ; dans chaque cas, on a itération essentielle du signifiant ; à partir du niveau de la lexie, cette itération, selon le système figuré mis en jeu, peut correspondre à l'indication d'un unique et même signifié, ou à la production d'un effet de sens irréductible à chaque occurrence concrète du même signifiant⁸⁰. » (Molinié et al., 1993, p. 33)

⁷⁸ Molinié, Georges, et Alain Viala. 1993. *Op. cit.*, p. 33.

⁷⁹ *Ibid.*, p. 34.

⁸⁰ *Ibid.*, p. 33.

Ainsi, les stylèmes sont nés avant tout des structures linguistiques répétitives dans le texte et servent ensuite au lecteur à déchiffrer les littéarités du texte. Le problème entre le fond et la forme, entre les analyses quantitatives et qualitatives est théoriquement résolu par l'introduction de la notion de stylème, reste à savoir comment le lecteur doit interpréter et trouver les stylèmes à partir des répétitions statistiques. À partir de là, la théorie de Molinié semble être moins pragmatique : il met l'accent sur les fonctions stylistiques que nous devrions trouver aux différents plans linguistiques. En plus des effets textuels perçus par le lecteur grâce à la littéarité singulière, il faudrait aussi étudier le texte dans son contexte social et historique, et tenir compte du genre. Pourtant, Molinié ne détaille pas la notion de stylème au niveau de la « littéarité singulière » : au lieu de proposer une méthode pour montrer, par exemple, comme un patron syntaxique devient un stylème singulier et significatif, il insiste sur le fait qu'il existe d'autres littéarités, de plus haut niveau qui prennent en compte le contexte social et générique.

Molinié, comme les autres linguistes et stylisticiens (que ce soit dans le cadre de la sémiotique, de la grammaire transformationnelle ou du structuralisme) propose une théorie linguistique qui essaie de relier les phénomènes linguistiques aux significations littéraires. Cette théorie est donc intéressante, mais reste dans les faits peu opérationnelle.

Bien que ce parcours théorique sur la notion de la littéarité ne nous ait pas apporté une méthode opérationnelle, il faut noter que la notion de littéarité est cruciale parce qu'elle montre comment des phénomènes linguistiques saillants peuvent produire des effets stylistiques. Le but de cette thèse n'est pas de proposer une méthode d'analyse à partir des résultats statistiques permettant de cerner la littéarité d'un texte, c'est-à-dire, essentiellement l'analyse de l'ensemble des traits particuliers et significatifs d'un texte. Influencée par Bally, la stylistique a commencé par ignorer le côté littéraire en se concentrant sur les structures générales de la langue ; guidée par le courant du Formalisme Russe, elle s'est recentrée sur les textes littéraires, mais exprime toujours l'intention de formaliser la langue. Pourtant, la linguistique contemporaine n'a jamais

arrêté de lier la stylistique aux différentes théories linguistiques : la sémantique interprétative de Rastier ou la sémiostylistique de Molinié, ont proposé toutes deux un système linguistique dans lequel les idiolectes deviennent des caractéristiques stylistiques. Cette thèse se fonde aussi sur le côté formel de la linguistique pour extraire les patrons syntaxiques, elle envisage ensuite de lancer les analyses de ces patrons en tenant en compte leurs contextes littéraire, générique et historique.

2.2.6 Les deux variétés de stylistiques

Dans cette dernière partie consacrée au style et à la stylistique, nous explorons les deux variétés de stylistiques mises en avant par Stanley Fish. Ces deux variétés se fondent sur deux types de traits linguistiques que nous avons déjà évoqués : les traits identificatoires, qui servent à identifier les œuvres en soulignant les phénomènes singuliers qui s’y trouvent, et les traits caractérisants, qui sont censément plus aptes à déchiffrer les qualités du style. L’opposition de points de vue que nous abordons dans cette partie met en évidence deux approches stylistiques, avec deux objectifs différents, à partir de ces deux types de traits linguistiques. Les critiques qui mettent en avant les traits identificatoires pensent que la stylistique doit, avant tout, être descriptive, c’est-à-dire qu’elle doit se concentrer sur la mise en avant des traits qui permettent d’identifier et de différencier les œuvres ; ceux qui valorisent les traits caractérisants cherchent uniquement les traits linguistiques qui possèdent déjà des fonctions stylistiques (les figures de style, en particulier), sans se soucier de leur prééminence dans le texte. Selon Fish, les deux variétés de stylistique, inspirées respectivement par ces deux types de traits, présentent chacune des difficultés : les traits identificatoires parviennent rarement à donner des analyses qualitatives sur le style, parce que ces traits n’ont pas de fonction stylistique prédéfinie ; à l’opposé, les traits caractérisants ont une

interprétation souvent contrainte, et les stylisticiens ont donc peu de liberté par rapport à cette norme. Nous défendons ici le point de vue que demander à un seul type de traits d'accomplir à la fois la tâche d'identification et la tâche de caractérisation n'est pas une solution idéale, ni même très réaliste. Il est en fait possible de combiner les deux variétés de stylistiques pour avoir une vue élargie sur le style.

La dichotomie entre traits identificatoires et traits caractérisants est déjà présente dans les théories sémantiques de Rastier et dans la sémiostylistique de Molinié : le premier aborde cette question par la formalisation des traits identificatoires (descriptifs) et des traits de caractérisation (interprétatifs). Les traits identificatoires sont souvent des traits de « bas niveaux⁸¹ » (Rastier, 2001) qui servent à différencier les œuvres, alors que les traits caractérisants possèdent des significations stylistiques. La même dichotomie a été reprise dans la sémiostylistique de Molinié, les répétitions statistiques des traits saillants (descriptifs) dans les textes peuvent devenir les littérarités de trois plans linguistiques (interprétatifs), sauf que le moyen de cette transformation ne se trouve pas au niveau sémantique (comme dans la théorie de Rastier), mais plutôt dans les fonctions sémiotiques que Molinié nomme le stylème. La stylistique exprime ainsi sa volonté de viser à la fois la description et l'interprétation des styles. Nous allons maintenant montrer que les deux fonctions de la stylistique (l'identification et la caractérisation) ont chacune des limites. Ainsi, une recherche stylistique qui ne se baserait que sur une des fonctions ne serait jamais complète.

Les traits caractérisants possèdent généralement déjà des fonctions stylistiques prédéfinies assez claires, la démarche est systématique (les fonctions d'une figure de style sont souvent peu variées). Le contexte n'influence que très peu l'attribution des fonctions stylistiques pour ce type de traits. De plus, examiner uniquement les traits caractérisants, indépendamment du texte, ne permet pas obligatoirement de repérer des faits significatifs sur le plan statistique.

⁸¹ Rastier, François. 2001. *Op. cit.*

Parallèlement, nous affirmons que l'interprétation des traits identificatoires est impossible quand ces traits n'ont pas de valeur stylistique identifiable (nous parlons ici des traits linguistiques extraits de manière non-supervisée dans les textes). Les patrons syntaxiques, qui n'ont pas de signification littéraire ou sémantique marquée (mais qui sont bel et bien surreprésentés statistiquement dans les textes) sont souvent impossibles à interpréter. Le fait qu'ils soient grammaticalement corrects, mais qu'ils ne soient pas liés de manière fixe et systématique à un effet stylistique clair et prédéfini, permet aux stylisticiens de leur attribuer librement les interprétations voulues. Ce processus est donc aussi discutable, parce que l'idée de rendre les traits identificatoires également caractérisants demande un travail subjectif de la part du stylisticien.

Stanley Fish a remarqué ce problème dans son article *What is Stylistics and Why Are They Saying Such Terrible Things about It? — Part II*. Il montre que le stylisticien, avant d'étudier un phénomène linguistique repéré dans un texte, a déjà dans sa tête une idée de la manière dont il l'interprétera. Ainsi, la démarche visant à donner des significations aux traits linguistiques, a priori neutre, est en réalité une pétition de principe : le lecteur a une certaine idée sur le style de l'auteur ; il repère les phénomènes saillants, il essaie de les interpréter d'une manière subjective pour confirmer son idée ; il continue à chercher d'autres phénomènes qui renforcent cette idée⁸² (Fish, 1979).

Fish a également exploré le problème des traits caractérisants, ce que nous avons tout juste évoqué au sujet des traits linguistiques extraits de manière supervisée. Ensuite, il résume les deux tendances interprétatives fondées sur deux catégories de traits linguistiques différentes. Il pense qu'il s'agit de « deux variétés de stylistiques » :

« Il existe deux variétés de stylistique et aucune d'entre elles ne soutiendra les revendications les plus fortes des stylisticiens. L'une s'effondre au milieu parce qu'il n'y a aucune façon légitime (et toutes les façons illégitimes) de

⁸² Fish, Stanley E. 1979. « What is Stylistics and Why Are They Saying Such Terrible Things about It? - Part II ». *boundary 2 boundary 2* 8 (1): 129.

relier ses composantes formelles et sémantiques ; dans l'autre, les composantes formelles et sémantiques sont si parfaitement liées que la distinction entre elles est perdue : ses étapes sont interprétatives de la première à la dernière. Les stylisticiens effectuent souvent ces deux types d'analyse sans voir la différence entre eux, car ils restent attachés à une distinction forme/sens même lorsqu'ils l'ont abandonnée dans leur pratique⁸³. » (Fish, 1979, p. 143)

Ensuite, il propose de changer de point de vue sur les fonctions de la stylistique :

« Ma thèse est ici que les patrons formels sont eux-mêmes les produits de l'interprétation et qu'il n'y a donc pas de patron formel, du moins dans le sens nécessaire à la pratique de la stylistique : c'est-à-dire qu'aucun patron que l'on observe avant l'interprétation est hasardeux et qui puisse donc être utilisé pour préférer une interprétation à une autre. La conclusion, cependant, n'est pas qu'il n'y a pas de patrons formels, mais qu'il y a toujours des patrons formels ; c'est juste que les patrons formels qui existent toujours seront toujours le produit d'un acte interprétatif antérieur, et ne pourront donc être discernés que tant que cet acte sera en vigueur⁸⁴. » (Fish, 1979, p. 143)

⁸³ « There are two varieties of stylistics and neither of them will support the stylisticians' strongest claims. The one falls apart in the middle because there is no legitimate way (and every illegitimate way) to relate its formal and semantic components; in the other the formal and semantic components are so perfectly related that the distinction between them is lost: its stages are interpretive from the first to the last. The stylisticians often perform these two kinds of analysis without seeing the difference between them, because they remain committed to a form/meaning distinction even when they have in their practice abandoned it ». *Ibid.*, p. 143.

⁸⁴ « Here my thesis is that formal patterns are themselves the products of interpretation and that therefore there is no such thing as a formal pattern, at least in the sense necessary for the practice of stylistics: that is, no pattern that one can observe before interpretation is hazarded, and which therefore can be used to prefer one interpretation to another. The conclusion, however, is not that there are no formal patterns, but that there are always formal patterns; it is just that the formal patterns there always are will always be the product of a prior interpretive act, and therefore will be available for discerning only so long as that act is in force ». *Idem.*

Stanley Fish sans doute ici un ton pessimiste sur la stylistique, mais il nous semble pertinent de terminer ce parcours sur le style et la stylistique en donnant notre point de vue général sur cette discipline.

Le grand problème de la stylistique, selon nous, est de ne pas avoir remarqué que ses deux variétés (celle qui part des traits caractérisants et celle qui part des traits identificatoires) ont deux points de départ différents. Nous avons montré que la recherche des traits stylistiquement signifiants demande une méthode de repérage supervisée (que ce soit par l'homme ou par l'ordinateur). Les traits caractérisants rencontrent souvent les limites de généralisation faute de saillance statistique : ils ne sont pas généralisables et ne peuvent pas être utilisés à grande échelle pour distinguer les œuvres ou les auteurs (à l'inverse des traits identificatoires).

De même, les traits identificatoires doivent rester identificatoires. Ces traits sont extraits sur une base purement statistique qui, par nature, échappe à la subjectivité, mais l'interprétation de ces traits, elle, bien que possible, reste difficile. Nous avons, d'une part, des patrons extraits qui sont statistiquement significatifs, mais ne sont pas interprétables d'un point de vue stylistique, en ce qu'il ne semble en rien spécifique d'un style ou d'un auteur malgré leur suremploi statistique. De l'autre, les patrons interprétables souffrent aussi d'une certaine subjectivité dans l'interprétation. Ainsi, bien que nous puissions leur attribuer des significations stylistiques, nous n'arriverons probablement jamais à produire des analyses systématiquement reproductibles et excluant complètement la subjectivité par ce biais.

Nous présentons un résumé des particularités des deux variétés de stylistiques dans le tableau suivant :

	Types d'extraction dans les textes	Significations statistiques	Fonctions stylistiques	L'utilisation stylistique
Les traits caractérisants	Extraction supervisée	Non-prouvées	Fixées, interprétations plus ou moins figées	Montrer partiellement des caractéristiques stylistiques d'un texte
Les traits identificatoires	Extraction non-supervisée	Prouvée	Arbitraires, interprétations libres et subjectives par les lecteurs	Identifier et différencier les textes ou les auteurs

Tableau 2.1 Les traits identificatoires et les traits caractérisants.

Répetons qu'il ne s'agit pas ici de nier l'importance de ces deux variétés d'analyse pour la stylistique. Nous affirmons au contraire que, de notre point de vue, il s'agit de deux faces complémentaires et nécessaires de l'analyse stylistique. Il ne serait pas pertinent de prédéfinir l'interprétation d'éléments repérés sur une base purement statistique, ou d'interpréter des éléments sans savoir s'ils sont statistiquement significatifs. Selon nous, la stylistique devrait adopter une perspective combinatoire, c'est-à-dire combiner les deux types d'indicateurs.

C'est ce qui a inspiré notre recherche, qui combine méthode supervisée et méthode non-supervisée, autrement dit traits caractérisants et traits identificatoires.

2.3 La stylométrie et la stylistique de corpus

Nous utilisons le terme « stylistique de corpus » pour désigner spécifiquement l'étude stylistique des textes littéraires. La stylistique de corpus applique les méthodes de la linguistique de corpus à la pratique de la stylistique. Elle implique un certain nombre de méthodologies différentes, facilitées par les outils informatiques ainsi que par les analyses quantitatives. Ces méthodologies sont diverses tant dans leur pratique, que dans leurs objectifs et leurs résultats. La stylométrie est une de ces méthodologies, nées à l'ère informatique. Nous parcourons dans cette partie les caractéristiques de cette nouvelle méthodologie, ses limites ainsi que ses objectifs visant à produire des résultats à la fois qualitatifs et quantitatifs à l'aide d'outils informatiques.

Nous avons déjà montré, dans la partie précédente, que l'objet d'étude de la stylistique est le texte. Grâce à cette délimitation, la stylométrie s'intéresse principalement aux unités linguistiques quantifiables. Leech et Short expriment une certaine appréciation de ce qu'ils appellent le « tournant vers le corpus de la stylistique⁸⁵ » (Leech et Short, 2013, p. 286) et rejettent les craintes que l'analyse informatique prenne le dessus sur la stylistique. Le développement des outils informatiques renforce la capacité à rechercher rapidement et avec précision des éléments tels que le vocabulaire, les collocations et certaines catégories syntaxiques simples. Pourtant, Leech et Short ne rejettent pas l'idée qu'une partie des recherches reste toujours manuelle. Il faut dire que les recherches informatisées suivent toujours une intuition critique, elles visent rarement à trouver quelque chose d'inattendu dans un texte. Par conséquent, ils concluent que, quelle que soit l'utilité des méthodes de corpus, un esprit humain est toujours nécessaire pour l'interprétation du texte. Nous

⁸⁵ Leech, Geoffrey Neil, et Michael Henry Short. 2013. *Style in fiction: a linguistic introduction to English fictional prose*. Harlow, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord : Longman, p. 286.

constatons que les opinions de Leech et Short valorisent le côté quantitatif de la stylistique outillée sans compromettre ni l'interprétation humaine ni la critique littéraire.

Le tournant vers la stylistique du texte naît de la réussite des premiers essais statistiques dans le domaine. Les premières études sur la fréquence des mots (que nous allons présenter dans la partie suivante) confirment la nécessité d'étudier les traits linguistiques et statistiques des textes. L'une des analyses stylistiques les plus citées est l'article de Michael Halliday sur le langage de *The Inheritors* de William Golding. À travers cette analyse, l'auteur veut montrer la « fonction textuelle » de sa théorie générale des fonctions du langage⁸⁶ :

« C'est par cette fonction [textuelle] que le langage fait des liens avec lui-même et avec la situation ; et le discours devient possible, parce que l'orateur ou l'écrivain peut produire un texte et que l'auditeur ou le lecteur peut en reconnaître un⁸⁷. » (Halliday, 2020 [1969], p. 334)

Halliday parle ensuite de la notion de « proéminence » dans les textes littéraires. Il préfère éviter de définir la « proéminence » comme un écart par rapport à une norme en raison des difficultés à choisir une norme appropriée, et parce que différents lecteurs sont susceptibles d'avoir des normes différentes en tête. Cela ne l'empêche pourtant pas d'affirmer qu'il sera possible d'identifier la « proéminence » statistique. Cela le conduit à rejeter l'idée que le style est insaisissable et trop idiosyncrasique pour pouvoir être analysé quantitativement, comme le disent parfois certains critiques littéraires. Nous voyons dans son étude une des premières confirmations de l'utilité et de la

⁸⁶ « It is through this function that language makes links with itself and with the situation ; and discourse becomes possible, because the speaker or writer can produce a text and the listener or reader can recognize one ». Halliday, M. 2020 [1969]. « Linguistic function and literary style: an inquiry into the language of William Golding's *The Inheritors* ». In : Chatman, Seymour Benjamin, éd. 1971. *Literary style: a symposium*. London, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, Etats-Unis d'Amérique, p. 334.

⁸⁷ *Idem*.

nécessité de recourir aux traits statistiques singuliers et récurrents pour analyser le style d'un auteur :

« S'il existe un style reconnaissable, qu'il s'agisse d'une œuvre, d'un auteur, d'une période entière ou d'une tradition littéraire, sa qualité distinctive peut en dernière analyse être exprimée en matière de fréquences relatives, bien que les caractéristiques linguistiques qui montrent la variation peuvent être simples et évidentes ou extrêmement subtiles et complexes⁸⁸. » (Halliday, 2020 [1969], p. 343)

Concernant les phénomènes quantifiables dans les textes, nous avons rencontré la notion de « proéminence » chez Halliday et de « emphasis » chez Riffaterre. Nous pensons que ces deux notions réfèrent essentiellement au même phénomène dans les textes, à savoir, le « foregrounding » défini par l'École de Prague dans les années 30. Le « foregrounding » désigne essentiellement la surreprésentation de certains faits linguistiques et nous pensons pouvoir extraire et quantifier ceux-ci par les outils informatiques. La notion du « foregrounding » a été traduite pour la première fois en anglais par P. L. Garvin, il explique que ce terme « fait référence à un stimulus culturellement non attendu dans une situation sociale et capable de provoquer une attention particulière⁸⁹ » (Garvin, 1964, p. 8).

Cette tendance à quantifier les traits proéminents (et ce qui nous semble être un avantage indéniable de l'aspect systématique de la stylométrie) provoque aussi des oppositions. Selon Gilles Philippe, les traits linguistiques, quelque proéminents qu'ils soient, ne font pas « correspondre ces variations à un contenu sémantique spécifique,

⁸⁸ « If there is such a thing as a recognizable style, whether of a work, an author, or an entire period or literary tradition, its distinctive quality can in the last analysis be stated in terms of relative frequencies, although the linguistic features that show significant variation may be simple and obvious or extremely subtle and complex ». Halliday, M. 2020 [1969]. *op. cit.*, p. 343.

⁸⁹ GARVIN, Paul Lucian. 1964. *A Prague School Reader on Esthetics, Literary Structure, and Style*. Selected and Translated by Paul L. Garvin. (Third Printing.). Georgetown University Press: Washington, p. 8.

ni à un souci expressif local⁹⁰ » (Philippe, 2005). Martin Wynne, malgré sa confiance dans le développement de la stylométrie outillée, a remarqué cette résistance contre l'approche automatique du domaine :

« Bien que l'on puisse soutenir que l'utilisation d'ordinateurs pour analyser des versions électroniques de textes et pour établir des preuves de normes linguistiques dans l'utilisation de la langue n'est qu'un moyen de vérifier et d'affiner les déclarations et les résultats empiriques, certains voient le danger que la recherche se préoccupe des procédures informatiques, de l'encodage et de l'annotation des textes électroniques, ce qui conduit à un manque regrettable d'attention à la textualité et au sens⁹¹. » (Wynne, 2006, p. 225)

En effet, la naissance de la stylométrie confirme l'utilité de l'analyse systématique des traits linguistiques pour la stylistique. Cependant, le développement très fort de ce domaine ne doit pas faire oublier le fait que l'interprétation des phénomènes statistiques garde toujours un côté subjectif indispensable. En 1980, Stanley E. Fish constate et résume les deux aspects de la stylistique :

« La stylistique est née d'une réaction à la subjectivité et à l'imprécision des études littéraires. Aux ravissements appréciatifs de la critique impressionniste, les stylisticiens prétendent substituer des descriptions linguistiques précises et rigoureuses et procéder à partir de ces descriptions à des interprétations pour lesquelles ils peuvent revendiquer une certaine

⁹⁰ Philippe, Gilles. 2005. *Op. cit.*

⁹¹ « Although it can be argued that the use of computers for analyzing electronic versions of texts, and for establishing evidence of linguistic norms in language use, is merely a means of verifying and refining empirical statements and findings, some see the danger of research becoming preoccupied with computational procedures, and the encoding and annotation of electronic texts, leading to a regrettable lack of attention to textuality and the meaning ». Wynne, Martin. 2006. « Stylistics : Corpus Approaches ». Elsevier. In: K. Brown, ed. *Encyclopedia of Language and Linguistics*. 2nd ed. Oxford: Elsevier, p. 225.

objectivité. La stylistique, en somme, est une tentative de donner à la critique une base scientifique⁹². » (Fish, 1980, p. 69-70)

La stylistique de corpus, selon Fish, est devenue une science qui consiste à réconcilier le côté littéraire/subjectif et le côté linguistique/objectif. Nous constatons que le problème de la subjectivité et de l'objectivité est récurrent et, d'une certaine manière, assez représentatif. Ronald Carter évoque aussi ce problème dans sa définition de la stylistique de corpus :

« La stylistique de corpus est une procédure essentiellement quantitative et implique une évaluation de l'importance tirée statistiquement d'un dénombrement basé sur le corpus. L'application effective de la stylistique de corpus aux textes implique nécessairement des décisions qualitatives et des actes d'interprétation faits par l'analyste à la lumière et dans une certaine mesure en amont des résultats de la banque de données assemblée. L'analyse stylistique du corpus est une procédure méthodologique relativement objective qui, au mieux, est guidée par un processus d'interprétation relativement subjectif⁹³. » (Carter, 2010, p. 66-7)

Cette définition est aussi celle que nous adoptons pour nos propres recherches. Nous admettons d'une part que les outils informatiques rendent les études à grande échelle

⁹² « Stylistics was born of a reaction to the subjectivity and imprecision of literary studies. For the appreciative raptures of the impressionistic critic, stylisticians purport to substitute precise and rigorous linguistic descriptions and to proceed from these descriptions to interpretations for which they can claim a measure of objectivity. Stylistics, in short, is an attempt to put criticism on a scientific basis ». Stanley Eugene. 1980. « What is Stylistics and Why Are They Saying Such Terrible Things About It? ». In: Fish, Stanley Eugene. 1980. *Is There a Text in This Class?: The Authority of Interpretive Communities*. Cambridge, Mass.: Harvard University Press, p. 69-70.

⁹³ « Corpus stylistic analysis is an essentially quantitative procedure and involves an assessment of significance drawn statistically from a corpus- informed count. The actual application of corpus stylistics to texts necessarily involves ... qualitative decisions and interpretive acts made by the analyst in the light of and to some degree in advance of the results from the assembled data-bank. Corpus stylistic analysis is a relatively objective methodological procedure that at its best is guided by a relatively subjective process of interpretation ». Carter, Ronald. 2010. « Methodologies for Stylistic Analysis: Practices and Pedagogies », janvier, p. 66-67.

possibles et nous affirmons, d'autre part, l'importance de la partie subjective et manuelle, qui consiste essentiellement à interpréter et à analyser les phénomènes récurrents et singuliers de nos résultats.

2.4 État de l'art sur la stylométrie

2.4.1 Les débuts de la stylométrie : le niveau lexical

Nous présentons dans cette partie les études stylométriques qui se concentrent sur les paramètres lexicaux. Les débuts de la stylométrie ont été marqués par une grande quantité d'études portant sur le lexique. Carter (2010) argumente que cette tendance est une approche générale inhérente de cette discipline :

« Dans un sens fondamental, la description de la langue par la linguistique de corpus donne la priorité au lexique. Alors que la stylistique accorde plus d'attention aux écarts par rapport aux normes linguistiques qui conduisent à la création d'effets artistiques, la linguistique de corpus se concentre sur ce qui peut être identifié informatiquement - ce qui tend à être sur les patrons lexicaux, en particulier les patrons fréquemment répétés⁹⁴ ». (Carter, 2010, p. 65)

La première recherche systématique de type stylométrique date de 1887. Thomas Mendenhall (1887) a étudié la distribution de la fréquence de milliers de mots de longueurs différentes dans les œuvres de Shakespeare, Bacon et Marlowe, pour montrer que la longueur de mots pourrait être considérée comme un marqueur stylistique satisfaisant dans l'attribution d'une œuvre à un certain auteur⁹⁵. Les démarches méthodologiques de Mendenhall ont inspiré les statisticiens et par conséquent, ont inauguré les recherches stylométriques qui se concentrent sur les unités mesurables. La longueur de mots et leur fréquence sont reprises dans plusieurs études stylométriques. L'application la plus réussie de cette méthode est sans doute la recherche qui porte sur

⁹⁴ « In a basic sense, corpus linguistic description of language prioritizes lexis. Whereas stylistics pays more attention to deviations from linguistic norms that lead to the creation of artistic effects, corpus linguistics focuses on what can be identified computationally – which tends to be on lexical patterns, especially patterns that are frequently repeated ». *Ibid.*, p. 65.

⁹⁵ Mendenhall, T. C. 1887. « THE CHARACTERISTIC CURVES OF COMPOSITION ». *Science ns-9* (214S): 237-46.

la dispute des *Federalist Papers*. Mosteller et Wallace (1984) ont utilisé les mots fonctionnels tels que les prépositions, les conjonctions et les articles comme paramètres et ont réussi à attribuer 12 articles, publiés dans *Federalist Papers*, à deux auteurs différents, Alexander Hamilton and James Madison⁹⁶.

L'étude de Mosteller et al. (1984) eut un tel succès qu'elle est considérée comme une étape clé dans l'histoire de la stylométrie. Sous cette influence, plusieurs travaux ont cherché à démontrer la bonne performance des mots fonctionnels en stylométrie, et en particulier pour l'attribution d'auteur. Argamon et Levitan (2005) ont aussi réussi à montrer que les mots les plus fréquents (c'est-à-dire essentiellement les mots fonctionnels) sont les mieux à même de distinguer les œuvres dans un corpus de vingt romans anglais contemporains, en suivant une méthode d'apprentissage fondée sur un modèle de discrimination linéaire⁹⁷. Ils arrivent enfin à la conclusion que « les mots fonctionnels s'avèrent encore étonnamment utiles pour l'attribution d'auteur, même après de nombreuses décennies de recherche sur les fonctionnalités et les algorithmes d'analyse stylométrique⁹⁸ ».

La méthode des mots fonctionnels est sans doute pertinente pour la tâche d'attribution d'auteur, mais ces mots fonctionnels ne permettent pas de caractériser les styles d'écriture. Un des exemples les plus importants concernant l'exploitation de la fréquence de mots avec une visée explicitement stylistique est celui de Burrows (1986), portant sur la caractérisation du style de Jane Austen⁹⁹. En sélectionnant, mais aussi en analysant le lexique des romans de Jane Austen selon les types de mots et leurs fréquences, Burrows se concentre particulièrement sur les verbes modaux (« could »,

⁹⁶ Mosteller, Frederick, et David Wallace. 1984. *Applied Bayesian and classical inference: the case of The Federalist papers*. New York, États-Unis d'Amérique, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord.

⁹⁷ Argamon, S., & Levitan, S. 2005. « Measuring the usefulness of function words for authorship attribution ». In *Proceedings of the Joint Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing*.

⁹⁸ *Idem*.

⁹⁹ BURROWS, J. F. 1986. « Modal Verbs and Moral Principles: An Aspect of Jane Austen's Style ». *Literary and Linguistic Computing* 1 (1): 9-23.

« may », etc.) et sur les pronoms personnels, possessifs et réflexifs (« myself », « himself », etc.). Il propose d'abord une méthode statistique visant à analyser la fréquence lexicale dans les textes de son corpus, et il s'intéresse ensuite aux mots sémantiquement pleins (les adjectifs, adverbes, verbes, noms). Grâce à une sélection manuelle et à son interprétation littéraire des mots pleins, il suggère que les personnages fictifs peuvent être comparés d'après les différences et les similarités parmi les mots pleins qui les qualifient. Dans son article, il confirme qu'Elizabeth Bennet dans *Pride and Prejudice* et Emma Woodhouse dans *Emma* se ressemblent beaucoup. Burrows insiste particulièrement sur l'importance de lier les résultats statistiques aux préoccupations propres à la critique littéraire. Suite à cette analyse stylistique du lexique des romans de Jane Austen, il constate les nouvelles possibilités que les outils informatiques apportent à la stylistique moderne. Les études de Burrows pourraient être considérées comme constituant un des premiers modèles prenant en compte à la fois l'aspect quantitatif et qualitatif : elles sont d'autant plus inspirantes pour notre thèse qu'elles proposent une méthode impliquant un double processus d'extraction, à la fois supervisé et non-supervisé.

Cependant, bien que la plupart des études prennent en compte les nouvelles possibilités apportées par les outils informatiques, la subjectivité, que ce soit au niveau de la sélection des données d'entrée ou au niveau de l'interprétation des résultats finaux, reste une question sensible. La polémique sur l'intérêt d'intégrer ou non les mots pleins dans les démarches stylométriques a connu une évolution importante avec l'arrivée des algorithmes d'apprentissage artificiel (machine learning) dans ce domaine. Diederich et al. (2003) appliquent par exemple l'algorithme SVM (Support Vector Machine) sur les articles publiés dans le journal allemand, *Berliner Zeitung*, de 1998 à 1999, sans annotations spécifiques. Ils obtiennent de bons résultats (60-80 %) ¹⁰⁰ pour la tâche d'attribution d'auteur (retrouver automatiquement les auteurs des articles). Dans la

¹⁰⁰ Diederich, Joachim, Jörg Kindermann, Edda Leopold, et Gerhard Paass. 2003. « Authorship Attribution with Support Vector Machines ». *Applied Intelligence* 19 (1): 109-23.

conclusion, ils indiquent que « les résultats expérimentaux montrent que les SVM obtiennent systématiquement de bonnes performances pour les tâches d'identification. Il n'est pas nécessaire de sélectionner des caractéristiques spécifiques, nous pouvons simplement prendre toutes les fréquences de tous les mots ». Les auteurs évitent ainsi d'avoir à choisir a priori les traits pertinents pour leur étude, mais leur unique objectif est d'attribuer les articles à un auteur, il ne s'agit pas d'étudier le style en lui-même.

Comme nous l'avons constaté, de nombreux chercheurs se concentrent sur la recherche du paramètre le plus performant pour distinguer des œuvres ou procéder à une tâche d'attribution d'auteur, sans chercher à interpréter les paramètres eux-mêmes. C'est pour cette raison que nous jugeons l'étude de Burrows (1986) inspirante à notre recherche : il ne réduit pas la stylométrie à une étude purement statistique des textes.

Les premières recherches stylistiques ont donc eu tendance à ignorer l'aspect qualitatif de la stylistique, d'une part parce ce sont souvent les mots fonctionnels qui sont les plus utiles pour différencier les œuvres, et d'autre part parce que ces mots fonctionnels (comme les prépositions ou les articles) ne possèdent pas vraiment de qualités stylistiques à analyser. C'est pour cette raison que notre thèse n'envisage pas de se recourir à ces paramètres pour caractériser le style. Notre méthode essaie d'extraire les unités linguistiques qui possèdent à la fois la capacité à distinguer les œuvres, mais aussi la possibilité d'être analysées stylistiquement par les lecteurs.

2.4.2 Vers la combinaison de paramètres stylistiques.

Nous présentons dans cette partie les études qui combinent les paramètres de différents « niveaux » linguistiques en stylométrie. Certaines études présentées dans la partie précédente ont commencé à intégrer des paramètres linguistiques à la méthode

lexicale. La période récente a, elle, été marquée par l'intégration de critères syntaxico-sémantiques aux études purement statistiques sur le lexique. Ces études sont possibles grâce aux avancées informatiques. La combinaison de plusieurs paramètres linguistiques est une tendance dans la stylométrie, avec des auteurs tels que Gwénolé Lecorvé (2020) qui relie les manifestations formelles de l'énoncé (le lexique, la morphologie, la syntaxe, la sémantique, la prosodie, la phonologie, etc.) avec la stylistique¹⁰¹.

Les études stylométriques purement sémantiques sont très rares. Pourtant, nous avons déjà constaté la tendance à combiner les paramètres syntaxico-sémantiques avec les paramètres lexicaux dans l'étude de Burrows (1986), présentée dans la section précédente. Culpeper (2009) emprunte partiellement la méthode de Burrows pour étudier les personnages dans la pièce de théâtre *Romeo and Juliet* de Shakespeare¹⁰². Il annote son corpus avec des étiquettes sémantiques (« love », « hate », etc.) et extrait ensuite une liste de mots qu'il nomme « keywords » (« les mots [pleins] qui sont statistiquement caractéristiques d'un texte ou des textes¹⁰³ ») pour chaque personnage afin de pouvoir les caractériser. Selon cette étude, les listes de mots-clés générées pour chacun des six personnages principaux reflètent divers aspects de leurs préoccupations ou de leur caractère. L'étude de Culpeper (2009) illustre les objectifs de la stylistique de corpus : extraire des données qui pourraient difficilement l'être par une simple lecture, et utiliser ces données pour explorer le style et l'effet d'un texte littéraire. La méthode de Culpeper (2009) est essentiellement une amélioration de celle de Burrows.

¹⁰¹ Lecorvé, Gwénolé. 2020. « Traitement Automatique Du Style Dans Le Langage Naturel : Quelques Contributions et Perspectives », thèse d'habilitation.

¹⁰² Culpeper, Jonathan. 2009. « Keyness: Words, Parts-of-Speech and Semantic Categories in the Character-Talk of Shakespeare's *Romeo and Juliet* ». *International Journal of Corpus Linguistics* 14 (1): 29-59.

¹⁰³ *Idem.*

D'autres études plus récentes proposent d'utiliser plus de deux paramètres linguistiques. Gamon (2004) présente une étude d'identification stylistique¹⁰⁴ qui mérite une explication particulière. Il utilise différents paramètres linguistiques pour essayer de différencier les cinq romans des sœurs Brontë. Des paramètres linguistiquement superficiels tels que la fréquence des mots fonctionnels et la longueur de phrases, ainsi que des paramètres linguistiquement plus profonds (les dépendances sémantiques) sont appliqués sur le corpus. L'étude teste la performance de chaque paramètre ainsi que la performance globale de la combinaison de ces paramètres. De son article, nous nous intéressons en premier lieu à la méthode définie pour tenir compte de la sémantique.

« Nous extrayons deux types d'informations à partir des graphes de dépendances sémantiques produits par le système NLPWin¹⁰⁵ : les caractéristiques sémantiques et les relations de modification sémantique. Des exemples de caractéristiques sémantiques sont le nombre et la personne [c'est-à-dire plus ou moins humain] pour les noms et les pronoms, le temps et l'aspect pour les verbes, et la sous-catégorisation (indiquant une sous-catégorisation réalisée par opposition à une sous-catégorisation potentielle) pour les verbes. Il y a un total de 80 traits sémantiques de ce type¹⁰⁶. »

Les « relations de modification sémantique » ressemblent en fait beaucoup aux relations syntaxiques de dépendance¹⁰⁷. Voici un des exemples mis en avant par l'auteur dans son article : « Verb Tsub Pron Tobj Noun (un nœud verbal avec un sujet profond

¹⁰⁴ Gamon, Michael. 2004. « Linguistic correlates of style: authorship classification with deep linguistic analysis features ». In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, 611-17. Geneva, Switzerland: COLING.

¹⁰⁵ NLPWin est un analyseur multifonction lexical, syntaxique et sémantique développé par Microsoft. <https://www.microsoft.com/en-us/research/project/nlpwin/>, consulté le 28 juillet, 2021.

¹⁰⁶ « We extract two kinds of information from the semantic dependency graphs produced by the NLPWin system: binary semantic features and semantic modification relations. Examples of semantic features are number and person features on nouns and pronouns, tense and aspectual features on verbs, and subcategorization features (indicating realized as opposed to potential subcategorization) on verbs. There is a total of 80 such semantic features ». Gamon, Michael. 2004. *Op.cit.*.

¹⁰⁷ « Dependency Grammar », https://en.wikipedia.org/wiki/Dependency_grammar, consulté le 17 avril, 2021.

pronominal et un objet profond nominal) ». Les fréquences de ces patrons sémantiques sont calculées et utilisées comme un paramètre linguistique profond.

L'article de Gamon (2004) nous intéresse parce qu'en faisant une comparaison entre différents paramètres, il arrive à la conclusion que plus un paramètre est linguistiquement profond, plus il produit de bons résultats :

« Nous avons montré que l'utilisation des paramètres linguistiques profonds pour l'attribution d'auteur peut entraîner une réduction significative du taux d'erreur par rapport à l'utilisation des paramètres linguistiques superficiels telles que la fréquence des mots fonctionnels et de trigrammes d'étiquettes morphosyntaxiques¹⁰⁸. »

Cela justifie nos propres démarches visant à utiliser des paramètres profonds au niveau linguistique. Ensuite, les « relations de modification sémantique » dont parle l'auteur sont en fait des relations syntaxico-sémantiques au sein de la phrase ou de la proposition. L'auteur utilise l'outil NLPWin (un analyseur disponible en 2004), mais la plupart des parseurs syntaxiques actuels peuvent associer à chaque mot des étiquettes de nature syntaxique ou sémantique, et ensuite mettre les mots en relation en suivant une grammaire de dépendance¹⁰⁹ (ceci est valable pour le français et l'anglais, ainsi que pour une vingtaine d'autres langues ; au-delà, les analyseurs restent hélas encore soit inexistants, soit de trop mauvaise qualité pour ce type d'analyse). Les patrons sémantiques présentés dans cet article sont un des premiers essais pour exploiter des patrons non-linéaires, fondés sur des relations linguistiques typées entre les mots. Ceci a bien évidemment influencé notre approche.

¹⁰⁸ « We have shown that the use of deep linguistic analysis features in authorship attribution can yield a significant reduction in error rate over the use of shallow linguistic features such as function word frequencies and part of speech trigrams ». Gamon, Michael. 2004. *Op. cit.*.

¹⁰⁹ « Dependency Grammar », https://en.wikipedia.org/wiki/Dependency_grammar, consulté le 17 avril, 2021.

Nous présentons enfin une autre étude qui exploite une combinaison des paramètres linguistiques pour différencier les œuvres. McCarthy, Lewis, Dufty, and McNamara (2006) se sont servis de l'outil Coh-Metrix pour calculer différentes mesures et relations sémantiques¹¹⁰. Coh-MétriX est un logiciel qui propose l'extraction d'un grand éventail de marqueurs stylistiques, dont les informations sur les synonymes et les hyperonymes. Les auteurs de cet article ont également appliqué l'analyse sémantique latente¹¹¹ de Deerwester et al.¹¹² (1990) aux traits lexicaux, afin de détecter automatiquement les mots sémantiquement similaires. Les résultats obtenus semblent satisfaisants, mais l'évaluation de ce modèle reste inachevée. Les auteurs mentionnent aussi une tentative pour baliser les textes selon la *Systemic Functional Grammar* (une théorie issue des travaux de Halliday¹¹³, 2004).

En plus de Coh-MétriX et du NLPWin, plusieurs outils informatiques ont vu le jour depuis les années 2000. Wmatrix est un outil logiciel d'analyse et de comparaison de corpus développé par Paul Rayson¹¹⁴. Il fournit une interface Web de visualisation d'annotation de corpus, ainsi que fonctionnalités standard telles que l'accès à des listes de fréquence ou à des concordances. Les paramètres superficiels tels que la fréquence de mots et les étiquettes morphosyntaxiques sont présentés en même temps que les paramètres plus complexes tels que les n-grams et les mots-clés, qui servent à délimiter les champs sémantiques évoqués dans les textes. Grâce à cette grande diversité de

¹¹⁰ McCarthy, Philip, Gwyneth Lewis, David Dufty, et Danielle McNamara. 2006. *Analyzing Writing Styles with Coh-Metrix*. FLAIRS 2006 - Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference. Vol. 2006.

¹¹¹ Landauer, Thomas K., Peter W. Foltz, et Darrell Laham. 1998. « An introduction to latent semantic analysis ». *Discourse Processes* 25 (2-3): 259-84.

¹¹² Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, et Richard Harshman. 1990. « Indexing by Latent Semantic Analysis ». *Journal of the American Society for Information Science* 41 (6): 391-407.

¹¹³ Halliday, Michael Alexander Kirkwood. 2004. *An Introduction to Functional Grammar*. Édité par Christian Matthias Ingemar Martin Matthiessen. 3rd ed. 1 vol. London New York: Arnold Oxford university press.

¹¹⁴ Rayson, Paul. 2009. « Wmatrix: a web-based corpus processing environment », Computing Department, Lancaster University. <http://ucrel.lancs.ac.uk/wmatrix/>.

paramètres disponibles, le logiciel a été appliqué à différentes tâches stylistiques, y compris la stylistique de corpus.

Un autre package qui regroupe plusieurs méthodes statistiques pour la stylométrie est *stylo*¹¹⁵ (Eder et al., 2016). Le package fournit les implémentations de diverses analyses établies dans le domaine de la stylistique computationnelle, y compris l'attribution d'auteur, la reconnaissance de genre, le développement du style (« stylochronométrie »), etc. Le package est essentiellement un recueil d'outils statistiques (analyse multidimensionnelle, analyse en composantes principales, analyse de cluster, arbres de consensus bootstrap), qui se base sur le lexique ou sur les n-grams. Si *Wmatrix* fournit davantage de paramètres stylistiques, l'accent du package *stylo* est mis sur le côté statistique, il ne se concentre pas sur l'exploitation de différents paramètres de la stylistique, mais se concentre plutôt sur différents calculs et différentes manipulations statistiques possibles et paramétrables. De plus, un certain nombre de méthodes d'apprentissage automatique supervisé sont aussi disponibles dans ce package : machines à vecteur support (SVM), bayésien naïfs (naive Bayes), k-plus proches voisins (k-means), etc. Grâce à cette grande diversité de traitements statistiques sur les données, *stylo* permet aux utilisateurs de partitionner leurs données (« clustering¹¹⁶ ») et aussi de procéder à l'analyse contrastive de deux ensembles de textes (l'outil peut par exemple générer des listes de mots statistiquement sur ou sous-employé dans un corpus donné par rapport à un autre).

Les recherches ainsi que les outils que nous venons de présenter dans cette partie ont tous pour objectif de combiner plusieurs méthodes pour identifier ou caractériser les styles littéraires. Pourtant, leurs réalisations restent très diverses : si les uns préfèrent

¹¹⁵ Eder, Maciej, Jan Rybicki, et Mike Kestemont. 2016. « Stylometry with R: A Package for Computational Text Analysis ». *The R Journal* 8 (1): 107-21.

¹¹⁶ Le partitionnement de données vise à diviser un ensemble de données en différents « paquets » homogènes, en ce sens que les données de chaque sous-ensemble partagent des caractéristiques communes, qui correspondent le plus souvent à des critères de proximité (similarité informatique) que l'on définit en introduisant des mesures et classes de distance entre objets. https://fr.wikipedia.org/wiki/Partitionnement_de_données, consulté le 6 août, 2021.

appliquer le plus de paramètres possible pour arriver à des résultats d'identification meilleurs, les autres préfèrent explorer un seul paramètre en le mettant à l'épreuve de différentes manipulations statistiques. Bien que ces études soient très différentes, sur le plan méthodologique notamment, elles témoignent, selon nous, des attentes de la stylistique outillée moderne. Le style, comme nous l'avons évoqué précédemment, est une notion holiste et chaque méthode ne peut dévoiler qu'un nombre limité d'éléments pertinents sur le plan stylistique. L'idée de combiner plusieurs méthodes permet de répondre à cette remarque.

Certes, la plupart des études présentées dans cette partie ont pour but de différencier les œuvres, et non de produire des analyses qualitatives. La tâche d'attribution d'auteur est omniprésente dans ces articles, ainsi que dans l'histoire de la stylométrie. Les études purement statistiques sur les traits proéminents du texte permettent bien de distinguer les œuvres, mais elles ignorent largement l'interprétation littéraire, qui doit suivre les résultats statistiques. Elles intègrent des traits syntaxiques ou sémantiques, soit pour compléter d'autres paramètres discriminatifs jugés trop peu discriminants, soit pour prouver que des paramètres linguistiquement plus profonds permettent de mieux différencier les œuvres. Enfin, presque tous les articles cités reposent sur des méthodes d'extraction non-supervisées : ils proposent rarement d'utiliser les outils informatiques pour extraire des phénomènes formels évoqués dans les critiques littéraires. Nous revenons ici à la question de la subjectivité, dont nous essayons de démontrer l'importance dans notre thèse.

C'est pour cette raison que nous avons décidé de ne pas nous servir des outils présentés ci-dessus, d'une part parce que notre but n'est pas une tâche d'identification d'auteur ou de caractérisation de textes, d'autre part parce que les outils, malgré la grande diversité de paramètres qu'ils proposent, ne fournissent pas de paramètres qui permettraient une analyse littéraire profonde. Néanmoins, ces études nous sont précieuses parce qu'elles montrent que la combinaison de méthodes différentes donne

souvent de meilleurs résultats, ce qui justifie aussi notre démarche fondée sur des méthodes supervisées et non-supervisées.

2.4.3 Niveau syntaxique

À partir des années 1990, la stylométrie commence à explorer la syntaxe grâce au développement d'analyseurs syntaxiques. La syntaxe est définie comme « la partie de la grammaire qui se préoccupe de la structure des phrases¹¹⁷ » (Aarts, 2008, p. 4). La relation entre la syntaxe et la stylistique est un sujet important dans la stylistique de corpus¹¹⁸. L'exploitation de la syntaxe dans la stylométrie est arrivée très tardivement, parce qu'il manquait jusque-là d'outils informatiques performants pour analyser d'une manière correcte la syntaxe. Sans outil d'analyse automatique, impossible de tenir compte de la syntaxe. Or, la syntaxe est indispensable pour pouvoir tenir compte des relations entre les mots. La syntaxe est en effet tout à fait conforme à ce que devrait être un paramètre stylistique selon R. W. Bailey (1979) : un paramètre « doit être saillant, structurel, fréquent, facilement quantifiable, et relativement à l'abri d'un contrôle conscient¹¹⁹ ». Dans cette dernière section de l'état de l'art, nous commençons par présenter les études stylométriques explorant des paramètres syntaxiques superficiels (longueur de phrases, etc.), avant d'évoquer des études impliquant des paramètres syntaxiques plus complexes grâce à l'arrivée des analyseurs syntaxiques ; nous présentons enfin deux études fondées sur la notion de n-grams, qui ont fortement influencé notre propre méthodologie.

¹¹⁷ Aarts, Bas. 2008. *English syntax and argumentation*. Basingstoke, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, Pays multiples.

¹¹⁸ Novakova, Iva, et Dirk Siepmann. 2020. *Phraseology and Style in Subgenres of the Novel: A Synthesis of Corpus and Literary Perspectives*. Palgrave Macmillan (Springer).

¹¹⁹ Bailey, R. W. 1979. « Authorship attribution in a forensic setting ». In: *Advances in Computer-aided Literary and Linguistic Research*, D. E. Ager, F. E. Knowles, and J. Smith (Eds.) AMLC, Birmingham.

Une des premières recherches stylométriques fondées sur des informations syntaxiques a été réalisée par le statisticien G. Udny Yule en 1939. Celui-ci propose de mesurer la longueur des phrases pour différencier les œuvres de différents auteurs¹²⁰. D'autres tentatives sur la longueur des phrases sont faites par Williams¹²¹ (1940) et de Wake¹²² (1957).

Une avancée dans le domaine est sans aucun doute le recours aux étiquettes morphosyntaxiques. Plusieurs recherches se fondent sur ce type d'étiquettes, dont celle de Baayen et al. (1996) : ces chercheurs partent d'un corpus anglais annoté pour attribuer les œuvres à différents auteurs¹²³. La préparation de leur corpus consiste à annoter morpho-syntaxiquement chaque mot dans le texte : aux mots sont associées des étiquettes morphosyntaxiques telles que le nombre (singulier, pluriel) pour les noms et les pronoms, la voix pour les verbes (active, passive), etc. L'annotation ne prend pas en compte les relations entre les mots. Ils réussissent ainsi à obtenir de meilleurs résultats que ceux basés uniquement sur la fréquence de mots.

Toujours avec la même idée de fonder l'analyse sur des étiquettes morphosyntaxiques et syntaxiques, Stamatatos, et al. (2000, 2001) se servent d'un outil de traitement automatique des langues (SCBD) pour accomplir avec succès la tâche d'attribution d'auteur. Le SCBD découpe les phrases en syntagmes, puis les annote au niveau syntagmatique (par exemple, en attribuant les étiquettes telles que NP pour les groupes nominaux, VP pour les groupes verbaux, etc.), avant de lancer le processus de différenciation, non pas sur les mots, mais sur les syntagmes¹²⁴¹²⁵. Les auteurs

¹²⁰ Yule, G. Udny. 1939. « On Sentence-Length as a Statistical Characteristic of Style in Prose: With Application to Two Cases of Disputed Authorship ». *Biometrika* 30 (3/4): 363 90.

¹²¹ Williams, C. B. 1940. « A Note on the Statistical Analysis of Sentence-Length as a Criterion of Literary Style ». *Biometrika* 31 (3/4): 356 61.

¹²² Wake, William C. 1957. « Sentence-Length Distributions of Greek Authors ». *Journal of the Royal Statistical Society. Series A (General)* 120 (3): 331 46.

¹²³ Baayen, H, H van Halteren, et F Tweedie. 1996. « Outside the cave of shadows: using syntactic annotation to enhance authorship attribution ». *Literary and Linguistic Computing* 11 (3): 121 32.

¹²⁴ Stamatatos, Efstathios, Nikos Fakotakis, et Kokkinakis George. 2000. « Automatic Text Categorization In Terms Of Genre, Author. » *Computational Linguistics* 26 (décembre): 471 95.

¹²⁵ Stamatatos, Efstathios, Nikos Fakotakis, et Kokkinakis George. 2001. « Computer-Based Authorship Attribution Without Lexical Measures ». *Computers and the Humanities* 35 (mai): 193 214.

confirment que cette méthode donne de meilleurs résultats que la méthode fondée simplement sur les mots fonctionnels.

Les méthodes basées sur les étiquettes morphosyntaxiques ont été rapidement complétées par des méthodes syntaxiques, grâce aux outils informatiques qui permettent aux chercheurs d'analyser la syntaxe d'une manière plus profonde. L'objet d'étude a été par conséquent élargi des simples mots avec leur étiquette morphosyntaxique aux séquences de mots avec leur étiquette syntaxique correspondante, à savoir, les n-grams linéaires. Argamon, Koppel et Avneri (1998) proposent d'utiliser des séquences de trois mots (des 3-grams, qui sont chacun une séquence d'étiquettes morphosyntaxiques) pour distinguer des articles de magazines par opposition à des articles de journaux. D'après eux, les 3-grams sont « suffisamment grands pour coder des informations syntaxiques utiles, et suffisamment petits pour rester gérables au niveau du calcul¹²⁶ ». Kukushkina, Polikarpov et Khmelev (2001) ont une démarche très similaire, en tenant compte des séquences de lettres et des séquences de classes grammaticales des mots dans leur étude. Ils concluent que « les fréquences d'occurrences des séquences de lettres et des séquences de classes grammaticales dans un texte russe sont des caractéristiques plutôt stables au niveau d'un auteur donné (que les mots seuls) et, apparemment, elles pourraient être utilisées dans l'attribution d'auteur contestée¹²⁷ ». Zhao et Zobel (2007) comparent les performances des méthodes basées respectivement sur les mots fonctionnels, les étiquettes morphosyntaxiques et les séquences d'étiquettes morphosyntaxiques sur un grand corpus de 634 œuvres de 55 auteurs différents. Ils montrent que les mots fonctionnels produisent toujours les meilleurs résultats, mais argumentent aussi que la performance inférieure des étiquettes morphosyntaxiques pourrait être due aux performances

¹²⁶ Argamon-Engelson, Shlomo, Moshe Koppel, et Galit Avneri. 1998. « Style-based text categorization: What newspaper am I reading ? », In *Proceedings of AAAI Workshop on Learning for Text Categorization*, 1-4.

¹²⁷ Kukushkina, O. V., A. A. Polikarpov, et D. V. Khmelev. 2001. « Using Literal and Grammatical Statistics for Authorship Attribution ». *Problems of Information Transmission* 37 (2): 172-84.

insatisfaisantes des tagueurs¹²⁸ utilisés. Karlgren et Eriksson (2007), quant à eux, ont recours à des séquences des mots qu'ils nomment « patrons » : ils étudient non seulement la typologie des patrons, mais aussi leur position dans la phrase afin de pouvoir distinguer les auteurs et les genres littéraires. Ce modèle utilise la complexité des phrases ainsi que la fréquence des patrons syntaxiques répétitifs. L'étude donne des résultats satisfaisants pour les tâches d'attribution d'auteur et de genres littéraires¹²⁹. Charles Dale Hollingsworth (2012) propose quant à lui de nouvelles règles de réécriture syntaxique (qu'il nomme « DepWords ») pour les phrases en prenant en compte les relations de dépendance syntaxique entre les mots : son objectif est de donner à chaque mot une étiquette originale qui représente à la fois sa fonction morphosyntaxique et sa relation de dépendance syntaxique¹³⁰, par exemple, « #c » veut dire un « usubj » qui dépend directement du « root » de la phrase. Ensuite, il teste la performance de ces étiquettes originales ainsi que les séquences linéaires de ces étiquettes sur son corpus et trouve que cette méthode de réécriture est au moins aussi performante que celle des mots fonctionnels. Sa méthode consistant à intégrer les relations de dépendance est une progression dans la stylométrie. Cependant, ces étiquettes originales sont syntaxiquement significatives, mais sémantiquement non-interprétables.

Les études de Ganascia et al. (2001, 2017) s'intéressent également aux patrons syntaxiques. Dans son étude publiée en 2001, l'auteur explore les patrons avec leur relation syntaxique. L'idée de représenter les relations syntaxiques d'une phrase se matérialise par l'Arbre Stratifié Ordonné (« un arbre ordonné est un arbre dans lequel les relations de gauche à droite, entre fils d'un même nœud, sont porteuses

¹²⁸ Zhao, Ying, et Justin Zobel. 2007. « Searching with style: authorship attribution in classic literature ». In *Proceedings of the thirtieth Australasian conference on Computer science - Volume 62*, 59-68. ACSC '07. AUS: Australian Computer Society, Inc.

¹²⁹ Karlgren, Jussi, et Gunnar Eriksson. 2007. « Authors, genre, and linguistic convention ». In *Proceedings of the SIGIR Workshop on Plagiarism Analysis, Authorship Attribution, and Near-Duplicate Detection*, 23-28.

¹³⁰ Hollingsworth, Charles. 2012. « Syntactic Stylometry: Using Sentence Structure for Authorship Attribution ». 2012.

d'information ¹³¹ »). Contrairement aux arbres syntaxiques de dépendance et de constituants, les Arbres Stratifiés Ordonnés possèdent souvent une profondeur unique (la profondeur 1), parce qu'il s'agit essentiellement de représenter les composants (les mots et les ponctuations) dans une phrase d'une manière linéaire. Les relations syntaxiques entre les composants ne sont pas prises en compte dans cette étude. Ces arbres syntaxiques sont ensuite utilisés pour mettre en évidence la similitude des patrons dans un texte par la distance d'édition (« une édition est une transformation élémentaire qui remplace un caractère dans une chaîne ou plus généralement un nœud dans un arbre, par un autre, éventuellement vide ¹³² »). La distance d'édition est essentiellement une méthode statistique de similarité, elle permet à l'auteur de repérer les patrons similaires ou répétitifs dans le texte. Une brève analyse qualitative dans laquelle l'auteur applique cette méthode sur 100 nouvelles et romans tirés de la littérature classique du 18^{ème} et 19^{ème} siècle est présentée dans cette étude. L'auteur conclut que la méthode est efficace et féconde, mais l'interprétation de ces patrons reste limitée.

Dans Frontini et al. (2017), les auteurs étudient les patrons syntaxiques séquentiels de taille 3 à 5 dans quatre romans français de Hugo, Balzac, Flaubert et Zola. Les patrons sont extraits par l'outil EREMOS sur les textes annotés préalablement par les étiquettes morphosyntaxiques. L'Analyse Factorielle des Correspondances est appliquée sur les résultats pour démontrer la capacité de ces patrons à identifier les œuvres, tandis que l'analyse littéraire des patrons se fonde sur la sélection manuelle des patrons interprétables. Les auteurs concluent que « les résultats suggèrent que cette méthodologie est très prometteuse en tant qu'instrument herméneutique offert aux

¹³¹ Ganascia, Jean-Gabriel. 2001. « Extraction automatique de motifs syntaxiques ». In *TALN 2001*. Tours, France.

¹³² *Idem*.

experts du domaine littéraire pour étudier le style dans les textes et extraire des caractéristiques stylistiques intéressantes¹³³ ».

Nous présentons maintenant plus en détail deux séries d'études qui ont fortement influencé cette thèse. Il s'agit d'une part des recherches sur les motifs à partir des travaux de Legallois et Charnois, et d'autre part des recherches de Sidorov mettant en jeu la syntaxe.

En 2016, Dominique Legallois, Thierry Charnois et Thierry Poibeau (2016) ont extrait et utilisé des « motifs », c'est-à-dire des séquences linéaires répétitives (à partir d'éléments qui peuvent être soit la forme originale du mot, soit son lemme, soit son étiquette morphosyntaxique), dans 150 romans contemporains : 50 romans policiers 50 romans dits « sérieux » et 50 romans sentimentaux¹³⁴. Les textes sont étiquetés au niveau morphosyntaxique et, ensuite, une fenêtre glissante permet d'extraire les n-grams. Les fréquences de ces n-grams sont calculées et utilisées pour les analyses quantitatives, et le « retour au texte » permet de découvrir la séquence originale derrière chaque « motif » dit « caractéristique ». Les auteurs ont listé les « motifs » les plus représentatifs pour chaque genre de romans dans cette étude : l'objectif est de pouvoir caractériser les genres littéraires à partir de ces « motifs » syntaxiques. Ils en concluent que :

« La méthode que nous avons présentée ici n'est pas seulement un moyen efficace pour identifier des patrons flexibles ; permettant d'extraire des unités qui ne sont pas décelables (ou facilement décelables) par les méthodes traditionnelles, elle met en évidence un ensemble élargi de formes — ce qui

¹³³ Frontini, Francesca, Mohamed Amine Boukhaled, et Jean-Gabriel Ganascia. 2017. « Mining for characterising patterns in literature using correspondence analysis: an experiment on French novels ». *Digital Humanities Quarterly*, Göttingen Dialog in Digital Humanities 2015, 11 (2).

¹³⁴ Legallois, Dominique, Thierry Charnois, et Thierry Poibeau. 2016. « Repérer les clichés dans les romans sentimentaux grâce à la méthode des « motifs » ». *Lidil. Revue de linguistique et de didactique des langues*, no 53 (mai): 95-117.

constitue la condition nécessaire pour l'analyse du préconstruit dans les genres¹³⁵. » (Legallois et al., 2016)

Cette étude inspire notre recherche de plusieurs manières :

Premièrement, elle confirme la nécessité de recourir aux unités séquentielles composées de plusieurs mots pour distinguer les œuvres. Si la plupart des articles que nous avons cités dans cette partie ont montré que les unités syntaxiques, d'une taille plus grande que les mots, sont au moins (sur le plan statistique) aussi performantes que les méthodes de référence pour la stylométrie (la fréquence de mots fonctionnels, notamment), nous constatons grâce à cette étude que certains genres littéraires se distinguent effectivement par des structures syntaxiques caractéristiques et particulières. De plus, leur capacité à différencier les genres littéraires a été confirmée au moins statistiquement par les résultats.

Deuxièmement, cette étude prend en compte l'aspect interprétatif et significatif des « motifs ». À part les fréquences absolues qui servent à différencier les genres littéraires, le retour au texte à partir des « motifs » est évidemment une étape importante qui ouvre la possibilité d'interprétation. Contrairement à d'autres articles cités dans cette partie, qui se concentrent sur la capacité des n-grams à différencier les œuvres, l'étude de Legallois et al. (2016) met l'accent sur les n-grams ainsi que sur leurs effets littéraires. Malheureusement, les auteurs de cette étude ne présentent que les « motifs » (en texte original) les plus représentatifs d'un genre littéraire sans entrer dans les détails, mais cette démarche est innovante en ce qu'elle intègre l'aspect subjectif dans l'analyse stylistique des genres littéraires.

Troisièmement, les « motifs » ont été extraits par Legallois et al. (2016) d'une manière non-supervisée, ce qui permet aux chercheurs d'éviter le problème opposant les « unités fonctionnelles » et les « unités de contenu » (les mots pleins). Nous avons

¹³⁵ *Idem.*

montré que, dans les articles de Burrows (1986) et de Culpeper (2009), les auteurs préfèrent utiliser les mots fonctionnels en raison de leur performance prouvée en stylistique ; ensuite, pour intégrer la sémantique afin d'interpréter les résultats sur le plan stylistique, les auteurs utilisent une liste des « mots-clés », qui sont essentiellement les mots pleins surreprésentés statistiquement dans les textes. Dans l'étude de Legallois et al. (2016), aucun filtre n'est appliqué pour sélectionner les « motifs ». Cette méthode présente toutefois des désavantages que nous présenterons également dans cette partie, mais la démarche est intéressante car elle se fonde à la fois sur les mots fonctionnels et sur les mots pleins ; elle prend également en compte, à la fois l'aspect quantitatif/objectif, et l'aspect qualitatif/subjectif de la stylistique.

Cette méthode de « motifs » a été ensuite développée par D. Legallois, T. Charnois et M. Latjavaara dans une étude publiée en 2018¹³⁶. Dans cet article, la technique des « motifs » a été appliquée à des romans français contemporains afin de détecter les caractéristiques stylistiques de chaque auteur du corpus. Legallois et al. (2018) proposent d'appliquer différentes étiquettes sémantiques aux noms, adjectifs et adverbes. Selon l'article, cette démarche rend les « motifs » plus précis et plus interprétables, et, dans le même temps, elle réduit considérablement le nombre de « motifs » extraits par texte. Nous voyons que cette tentative est différente des autres recherches fondées sur la syntaxe, parce qu'elle ne cherche pas à trouver « le meilleur » paramètre pour différencier les œuvres, mais se concentre sur l'amélioration de l'interprétabilité des « motifs », le côté qualitatif, qui, selon les auteurs, est essentiel pour les recherches stylométriques.

Nous avons présenté les avantages de la méthode des « motifs », mais celle-ci reste cependant à améliorer. Plusieurs questions se posent : quelles sont les tailles de n-grams les plus efficaces pour les « motifs » ? Les fréquences absolues sont-elles les meilleurs objets de calcul ? Comment améliorer l'efficacité d'extraction en éliminant dès le début

¹³⁶ Legallois, Dominique, Thierry Charnois, et Meri Latjavaara, éd. 2018. *The grammar of genres and styles : from discrete to non-discrete units*. Berlin, Allemagne : De Gruyter Mouton.

tous les « motifs » qui ne correspondent pas à des syntagmes complets ? Et enfin, comment interpréter ces « motifs » sur le plan stylistique ? Nous proposons, dans cette thèse, plusieurs améliorations permettant de répondre à ces questions.

La deuxième étude sur la syntaxe que nous allons présenter dans cette partie est celle menée par G. Sidorov et al. (2014)¹³⁷. Cette étude partage en fait la même approche que la nôtre, mais son point de départ, sa réalisation et son application restent très différents du nôtre.

Nous pouvons également considérer la démarche de Sidorov et al. (2014) comme une amélioration de la méthode de Legallois et al. (2016) présentée ci-dessus. Dans leur article, Sidorov et al. (2014) proposent d'intégrer les relations syntaxiques de dépendance dans les n-grams séquentiels afin d'attribuer 39 œuvres aux trois auteurs suivants : Booth Tarkington, George Veizey et Louis Tracy. Les auteurs utilisent les algorithmes de SVM¹³⁸, NB et J48¹³⁹ dans leurs expériences. Ils montrent qu'au lieu d'extraire directement des « motifs », ou des n-grams linéaires séquentiels dans les textes, l'ajout de relations syntaxiques dans les n-grams produit une amélioration considérable. « Les résultats montrent que la technique sn-gram surpasse la technique de base [les n-grams linéaires de mots, les étiquettes morphosyntaxiques et les caractères]¹⁴⁰ » (G. Sidorov et al., 2014, p. 860).

Pour extraire les n-grams avec relations syntaxiques, appelées « sn-grams » (« syntactic n-grams ») dans l'article, les auteurs lancent une analyse syntaxique complète avec l'analyseur StanfordNLP, afin d'obtenir les relations de dépendance syntaxique. Les phrases peuvent ainsi être visualisées comme des arbres syntaxiques

¹³⁷ Sidorov, Grigori, Francisco Velasquez, Efstathios Stamatatos, Alexander Gelbukh, et Liliana Chanona-Hernández. 2014. « Syntactic N-Grams as Machine Learning Features for Natural Language Processing ». *Expert Systems with Applications, Methods and Applications of Artificial and Computational Intelligence*, 41 (3): 853-60.

¹³⁸ « Support Vector Machine », <https://dataanalyticspost.com/Lexique/svm/>, consulté le 7 août, 2021.

¹³⁹ Aljawarneh, Shadi, Muneer Bani Yassein, et Mohammed Aljundi. 2019. « An enhanced J48 classification algorithm for the anomaly intrusion detection systems ». *Cluster Computing* 22 (septembre): 1-17.

¹⁴⁰ « The results show that sn-gram technique outperforms the baseline technique ». *Idem*.

dont les mots sont les feuilles et dont les branches sont les relations de dépendance. Par exemple, à partir de la phrase ci-dessous :

1. *Il suivait les laboureurs, et chassait, à coups de motte de terre, les corbeaux qui s'envolaient.* (Madame Bovary, Flaubert)

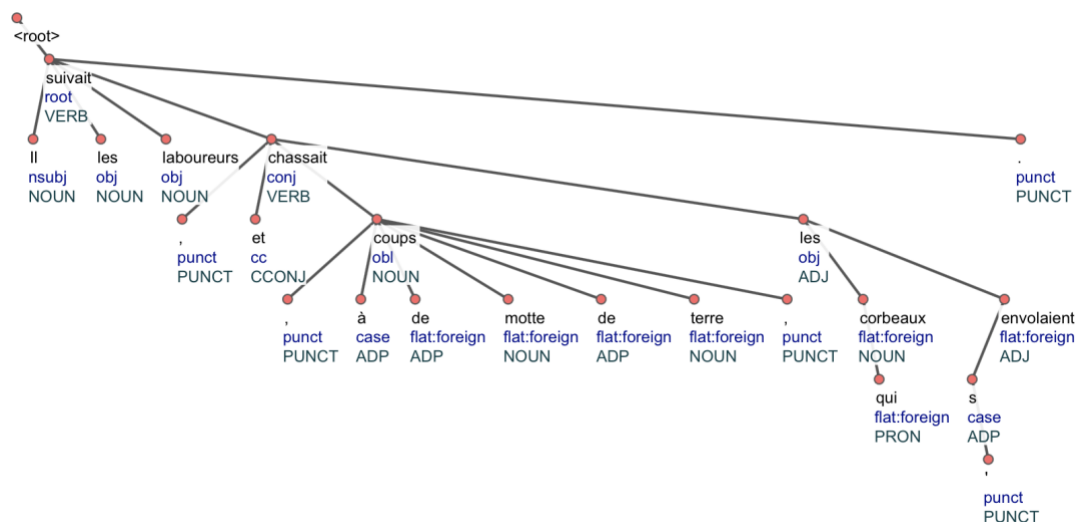


Figure 1.1 Exemple d'un arbre syntaxique de dépendance.

Les « sn-grams » sont présentés linéairement dans les résultats, mais tous les mots dans un certain sn-gram sont liés par des relations de dépendance syntaxique arborescentes. Ensuite, un algorithme extrait tous les « sn-grams » selon une taille linéaire prédéfinie de nombre de composants pour les « sn-grams » (voir les paragraphes suivants pour l'explication de l'algorithme et les exemples d'extraction). Sidorov et al. (2014) confirment que ces sn-grams contiennent plus d'informations syntaxiques que les n-grams linéaires, ainsi, ils sont plus performants pour différencier les œuvres.

Nous avons évoqué le fait que notre propre stratégie d'extraction non-supervisée partage beaucoup de choses avec celle mise au point par Sidorov et al. (2014). Notre méthode comporte toutefois des différences majeures :

1. La méthode de Sidorov et al. (2014) consiste principalement à trouver un paramètre syntaxique plus performant que les n-grams linéaires pour réaliser la tâche d'attribution d'auteur. Ainsi, la méthode ne s'intéresse-t-elle qu'à montrer les différences entre auteurs au niveau des fréquences de « sn-grams », tandis que nous voulons extraire les n-grams syntaxiques à des fins d'interprétation. Sidorov et al. (2014) se concentrent sur l'aspect statistique des structures syntaxiques, alors que nous nous situons, de ce point de vue, davantage dans la lignée de Legallois et al. (2018). Notre point de départ diffère donc de celui de Sidorov et al. (2014), mais nous partageons aussi de nombreux points communs avec ce travail antérieur.

2. Une autre différence fondamentale se situe dans la stratégie d'extraction de n-gram à base syntaxique : les n-grams extraits par notre méthode sont différents de ceux de Sidorov et al. (2014). Dans l'article publié en 2017 de Posadas Durán et al.¹⁴¹ (2017, p. 80), les auteurs expliquent l'algorithme pour extraire les sn-grams.

« L'algorithme extrait les n-grammes syntaxiques d'un arbre de dépendances en exécutant une procédure en deux étapes. Lors de la première étape, l'algorithme effectue une première recherche en largeur et identifie tous les sous-arbres de hauteur égale à 1. Dans la deuxième étape, l'algorithme parcourt l'arbre en remplaçant l'occurrence de nœud dans un sous-arbre par les sous-arbres de niveaux supérieurs où le nœud est la racine. Les sous-arbres extraits correspondent aux n-grams syntaxiques de l'arbre¹⁴². »
(Posadas Durán et al. 2017, p. 80)

¹⁴¹ Posadas Durán, Juan, Grigori Sidorov, Helena Gomez Adorno, Ildar Batyrshin, Elibeth Mirasol-Méendez, Gabriela Posadas-Durán, et Liliana Chanona-Hernández. 2017. « Algorithm for Extraction of Subtrees of a Sentence Dependency Parse Tree ». *Acta Polytechnica Hungarica* 14 (novembre): 79.

¹⁴² « The algorithm extracts the syntactic n-grams from a dependency tree by performing a two-stage procedure. The first stage the algorithm conducts a breadth-first search of the tree and finds all the subtrees of height equal to 1. In the second stage, the algorithm traverses the tree in postorder replacing the node occurrence in a sub-tree with the subtrees from higher levels where the node is the root. The extracted subtrees correspond to syntactic n-grams of the tree ». *Idem*.

Cette citation résume l'algorithme récursif pour la phase d'extraction : le seul paramètre demandé avant l'extraction est la taille finale de nombre de composants dans les « sn-grams ». L'algorithme peut alors parcourir toutes les branches de l'arbre pour découvrir tous les « sn-grams » conformes à la taille demandée. Pour mieux expliquer les résultats de cet algorithme, prenons la phrase prise comme exemple dans l'article des auteurs :

2. *Victor sat at the counter on a plush red stool.* (L'algorithme ne tient pas compte de la ponctuation).

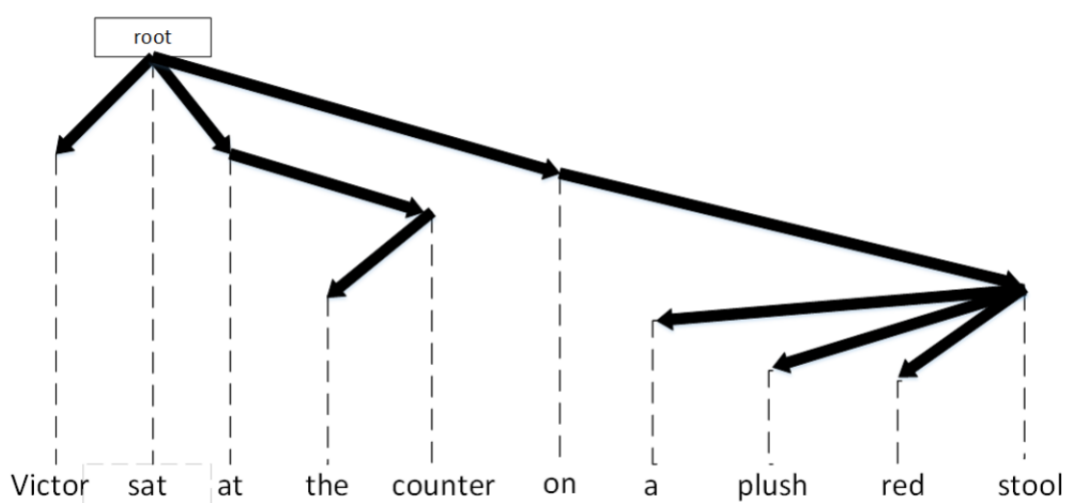


Figure 1.2 Exemple dans l'article de Sidorov et al. (2017).

Les sn-grams de taille 3 extraits par l'algorithme sont : *sat [Victor, at]*, *sat [Victor, on]*, *sat [at, on]*, *sat [on [stool]]*, *sat [at [counter]]*, *on [stool [plush]]*, *on [stool[a]]*, *on [stool [red]]*, *stool [a, plush]*, *stool [a, red]*, *stool [plush, red]*, *sat [at [counter]]*, *at [counter [the]]*.

Nous constatons que l'algorithme cherche d'abord la plus haute feuille de l'arbre (c'est-à-dire la feuille la plus proche de la racine), « sat », puis cherche les branches qui relient ce nœud à d'autres feuilles d'un niveau inférieur (nous en avons trois dans cet exemple : « Victor », « at » et « on »). L'utilisateur ayant prédéfini la taille des « sn-grams » finaux à 3, l'algorithme produit trois sn-grams différents (*sat [Victor, at]*, *sat [Victor, on]*, *sat [at, on]*) pour représenter tous les mots qui dépendent de « sat ». Nous

pouvons constater que les « sn-grams » extraits par l'algorithme n'ont pas de signification sémantique, bien qu'ils soient grammaticalement significatifs. Prédéfinir la taille des sn-grams est similaire à la recherche de la taille idéale pour les n-grams traditionnels. Par ailleurs, la plupart des n-gram obtenus ne sont pas complets sur le plan syntaxico-sémantique, ce qui en rend l'interprétation impossible. Certes, le but des auteurs de l'article n'est pas de produire les « sn-grams » interprétables, ils cherchent simplement à évaluer l'apport de ce nouveau paramètre pour une tâche d'attribution d'auteur ; mais, à l'opposé, les données statistiques ne sont qu'un des moyens d'analyse que nous souhaitons mettre en œuvre de cette thèse. Par conséquent, nous ne pouvons pas appliquer la méthode de Sidorov et al. (2014) telle quelle sur notre corpus. Notre méthode devra prendre en compte à la fois les relations syntaxiques de dépendance et l'interprétabilité des n-grams obtenus. Elle sera présentée dans le Chapitre 3.

3. Le but de notre étude est aussi différent de celle de Sidorov et al. (2014). Nous avons décidé d'élaborer une méthode d'extraction de patrons syntaxiques avant tout pour extraire des patrons caractéristiques pour alimenter les analyses stylistiques. Ainsi, cette méthode d'extraction non-supervisée peut-elle utilement compléter notre méthode d'extraction supervisée, qui est essentiellement tournée vers l'analyse des figures de style. À cause des limites techniques des outils informatiques, un grand nombre de figures de style ne peuvent pas être extraites automatiquement. C'est pourquoi l'introduction des n-grams syntaxiques, extraits d'une manière non-supervisée, permet de donner une vue élargie et plus générale du style, à travers la caractérisation de la syntaxe d'un texte ou d'un auteur. La méthode de « sn-grams » de l'article de Sidorov et al. (2014) est uniquement appliquée pour distinguer les œuvres et les genres littéraires. Leur méthode ne révèle que très peu de caractéristiques stylistiques parce que les « sn-grams », comme nous avons déjà mentionné, sont souvent incomplets sémantiquement et donc difficiles à interpréter.

Une des difficultés dans les études citées dans cette section est le problème de l'interprétabilité. Les patrons extraits d'une manière supervisée ou non-supervisée sont

souvent peu informatifs hors du contexte, ainsi, la sélection des patrons pour l'analyse littéraire et stylistique est souvent difficile. Dans une étude de Longrée et al. (2013), les auteurs proposent de prendre en considération la fonction structurante (« On exemplifiera la fonction structurante par des motifs que l'on pourrait regrouper sous le nom de « clichés de liaison »¹⁴³ ») et la fonction caractérisante (à savoir capacité à distinguer un genre discursif des autres). Cependant, les deux fonctions sont, d'après nous, insuffisantes pour caractériser un style littéraire, l'étude stylistique devrait se concentrer sur tous les phénomènes stylistiques liés aux patrons syntaxiques.

Nous présenterons dans le Chapitre 3 une nouvelle méthode d'extraction non-supervisée qui se base également sur les n-grams syntaxiques, mais qui améliore la méthode de Sidorov et al. (2014). Elle permet de produire les mêmes analyses statistiques tout en gardant la possibilité d'interpréter les n-grams et, ainsi, d'analyser le fond du texte.

¹⁴³ Longrée, Dominique, et Sylvie Mellet. 2013. « Le motif : une unité phraséologique englobante ? Étendre le champ de la phraséologie de la langue au discours* ». *Langages*, Vers une extension du domaine de la phraséologie, 1 (189): 65-79.

2.5 Conclusion

Dans ce chapitre, nous avons parcouru certaines théories importantes sur le style et la stylistique, et nous avons proposé un bref parcours des avancées récentes en matière de stylométrie outillée. Nous avons présenté dans chaque partie nos points de vue sur ces sujets au cœur de notre propre thèse.

Nous avons choisi de présenter des controverses importantes dans les études du style et de la stylistique. Même si notre objectif n'est pas de redéfinir le style ou la stylistique (nous n'avons pas l'intention d'insérer nos recherches stylistiques au sein d'une théorie littéraire ou linguistique particulière), ces idées contradictoires qui ont fait l'objet de longs débats, nous aident à élucider des notions importantes pour cette discipline. Nous adoptons ici un point de vue plutôt linguistique et formel, en partant du principe que le style est saisissable et quantifiable par les outils linguistiques. L'objet d'étude de la stylistique doit être le texte lui-même, et nous proposons ici d'abandonner la stylistique de production, qui se concentre sur l'aspect psychologique de l'auteur du texte. Nous croyons de plus que l'on peut associer aux traits linguistiques saillants des interprétations stylistiques liées à la réception par les lecteurs. Ces interprétations devraient non seulement être liées au contexte individuel du texte, mais aussi à son contexte historique et social. L'idée d'identifier un objet stylistique universel, qui fusionnerait fonction identificatoire et fonction caractérisante, est selon nous une chimère. Néanmoins, la possibilité de combiner ces deux approches peut aboutir à une vue élargie des caractéristiques stylistiques d'un texte, mêlant à la fois les analyses quantitatives et qualitatives.

En examinant les travaux antérieurs en stylométrie outillée, nous constatons que la base de cette discipline est statistique. Les outils informatiques ont permis aux chercheurs d'explorer les traits saillants dans de grandes masses de données textuelles. Cette nouvelle perspective ne doit toutefois pas ignorer l'interprétation des données :

que ce soit au niveau des choix des traits à extraire, ou au niveau de l'interprétation, l'analyse humaine reste indispensable.

L'état de l'art sur la stylométrie confirme cet objectif de fusionner analyses quantitatives et qualitatives. Même si les premiers essais en matière de stylométrie ont été purement statistiques, depuis plusieurs décennies des structures linguistiques plus profondes, telles que les dépendances syntaxiques et le champ sémantique, ont été progressivement intégrées aux modèles. Ainsi, les études les plus récentes proposent souvent un grand éventail de traits, de différents niveaux linguistiques, pour mieux identifier et caractériser le style. La stylistique est interdisciplinaire par nature, ce qui est également vrai de la stylométrie outillée, qui a littéralement absorbé différentes théories et méthodes récentes, issues tant de la littérature que de la linguistique. Quelques études importantes sur le rôle de la syntaxe dans la stylométrie ont été présentées dans cette partie, dont certaines études récentes très similaires aux nôtres. Nous allons présenter, dans le chapitre suivant, une nouvelle méthode qui s'inspire largement de ces travaux récents, mais qui apporte des améliorations quant à l'exploitation des structures syntaxiques pour la caractérisation des styles littéraires.

Chapitre 3 Présentation de la méthodologie

3.1 Introduction

Nous présentons dans ce chapitre la méthode d'extraction supervisée pour les patrons syntaxiques caractérisants et ensuite la méthode d'extraction non-supervisée pour les patrons syntaxiques identificatoires. Nous avons montré dans l'état de l'art que si l'introduction d'aspects syntaxiques est une progression remarquable dans la stylométrie outillée, cette stratégie dépend aussi largement des outils informatiques capables de produire ces annotations syntaxiques.

Notre démarche consiste à explorer les paramètres syntaxiques afin de faire des analyses à la fois qualitatives et quantitatives quant aux caractéristiques syntaxiques d'un texte. Pour accomplir la fonction double de la stylistique, à savoir la fonction identificatoire qui cherche à différencier les œuvres et la fonction caractérisante qui porte sur les spécificités propres d'un texte, nous proposons une méthodologie qui se fonde à la fois sur les traits syntaxiques identificatoires et les traits syntaxiques caractérisants. Les premiers ne sont pas forcément interprétables du point de vue stylistique parce que c'est leur spécificité statistique qui suffit à les mettre en évidence ; les seconds sont de nature stylistique, mais la généralisation des qualités stylistiques dévoilées par ces traits est difficile. La combinaison de ces deux types de traits nous permet de viser les deux types d'analyse évoqués dans la section 2.2.6 du Chapitre 2.

Nous commençons par présenter deux façons de représenter les structures syntaxiques (donc hiérarchiques) des phrases (3.2) : la grammaire de dépendance et la grammaire de constituants. Nous justifions notre choix de représentation sous forme de grammaire de dépendance parce que ce formalisme s'adapte mieux à notre objectif. Dans la section 3.3, les performances de plusieurs analyseurs syntaxiques sont

comparées. Nous choisissons finalement UDPipe en nous fondant sur les performances publiées par les développeurs de cet outil, et aussi sur les résultats obtenus sur deux ensembles de phrases annotées manuellement, représentatives de notre corpus. Dans la section 3.4, nous revenons sur les deux types de traits linguistiques pertinents pour la stylistique : les traits caractérisants et les traits identificatoires. Nous argumentons sur la nécessité de combiner ces deux types de traits, tout en mettant en relief leur importance respective dans les démarches stylistiques. Dans la partie 3.5, nous présentons une nouvelle méthode pour mesurer la complexité syntaxique, basée également sur la grammaire de dépendance. Les méthodes d'extraction supervisée pour les patrons caractérisants (l'inversion du sujet, le clivage et la répétition) sont présentées dans la section 3.6, ainsi que les mécanismes d'extraction et les discussions des résultats. La partie 3.7 porte sur la présentation de la méthode d'extraction non-supervisée des patrons identificatoires : nous présentons d'abord le mécanisme d'identification et d'extraction, avant de préciser les grandes étapes du processus. Cette partie comprend également une discussion sur les paramètres utilisés (la notion de profondeur des patrons notamment), sur les limites de la méthode et sur sa comparaison avec la méthode d'extraction linéaire.

Ce chapitre contient donc une présentation systématique de la méthodologie utilisée dans la thèse, avant son application à des corpus caractéristiques.

3.2 Les relations syntaxiques : linéarité ou hiérarchie

Nous avons présenté dans l'état de l'art quelques études qui exploitent des paramètres syntaxiques pour la stylométrie, mais nous avons également constaté que ce type d'approche dépend largement de la disponibilité d'outils informatiques performants adaptés à la tâche. Dans cette partie, nous évoquons d'abord la notion de paramètres profonds, qui vont au-delà de la linéarité des phrases. Nous présentons ensuite les notions de grammaire de dépendance et de grammaire de constituants¹⁴⁴ (Dubois et al. 2012), qui proposent chacune une méthode de représentation de la structure hiérarchique des phrases, par-delà leur représentation linéaire. Enfin, nous justifions notre choix, à savoir la grammaire de dépendance pour représenter la structure des phrases dans notre corpus.

En examinant les études présentées dans l'état de l'art, nous avons évoqué deux notions concernant les paramètres stylométriques : les paramètres superficiels et les paramètres profonds. Il n'existe pas de définition précise de la distinction entre ces deux catégories de paramètres, mais on peut admettre que les paramètres superficiels correspondent aux traits linguistiques marqués par leur facilité de repérage dans les textes, par exemple : la fréquence des mots¹⁴⁵ (Mosteller et al., 1984) ou la longueur des phrases¹⁴⁶ (Yule, 1939), etc. Les paramètres dits « profonds » demandent quant à eux une analyse plus poussée pour les repérer : ces paramètres correspondent aux

¹⁴⁴ Dubois, Jean, Mathée Giacomo-Marcellesi, Louis Guespin, Christiane Marcellesi, Jean-Baptiste Marcellesi, et Jean-Pierre Mével. 2012. *Le dictionnaire de linguistique et des sciences du langage*. Paris, France: Larousse.

¹⁴⁵ Mosteller, Frederick, et David Wallace. 1984. *Applied Bayesian and classical inference: the case of The Federalist papers*. New York, États-Unis d'Amérique, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord.

¹⁴⁶ Yule, G. Udny. 1939. « On Sentence-Length as a Statistical Characteristic of Style in Prose: With Application to Two Cases of Disputed Authorship ». *Biometrika* 30 (3/4): 363-90.

relations syntaxiques¹⁴⁷ (Sidorov et al., 2014), aux relations sémantiques¹⁴⁸ (Gamon 2004), etc. Le passage des paramètres superficiels vers les paramètres profonds sous-entend un changement de point de vue sur la phrase elle-même : les paramètres superficiels se concentrent souvent sur la linéarité de la phrase, ils considèrent la phrase comme une séquence d'unités linguistiques (par exemple, le nombre de mots dans une phrase, les fragments séquentiels de mots ou les n-grams linéaires que nous avons rencontrés dans l'étude de Legallois et al.¹⁴⁹, 2016). Les paramètres linguistiquement profonds, quant à eux, prennent aussi en compte les relations syntaxiques entre composants. On considère ainsi la phrase comme une structure hiérarchique dans laquelle les mots sont liés par des relations syntaxiques. Une des théories qui ont inspiré notre thèse est la syntaxe structurale de Lucien Tesnière. Selon ce linguiste :

« L'étude de la phrase, qui est l'objet propre de la syntaxe structurale, est essentiellement l'étude de sa structure, qui n'est autre que la hiérarchie de ses connexions.¹⁵⁰ » (Tesnière, 1988 [1959], p. 14)

Selon Igor Mel'cuk, le travail de Tesnière « a jeté les bases de toutes les explorations ultérieures dans ce domaine [à savoir la syntaxe de dépendance]¹⁵¹ » (Mel'cuk, 1988, p. 3). Tesnière pourrait être considéré comme le précurseur de la notion de grammaire de dépendance. Néanmoins, Michael Covington a montré que des modes

¹⁴⁷ Sidorov, Grigori, Francisco Velasquez, Efstathios Stamatatos, Alexander Gelbukh, et Liliana Chanona-Hernández. 2014. « Syntactic N-Grams as Machine Learning Features for Natural Language Processing ». *Expert Systems with Applications, Methods and Applications of Artificial and Computational Intelligence*, 41 (3): 853-60.

¹⁴⁸ Gamon, Michael. 2004. « Linguistic correlates of style: authorship classification with deep linguistic analysis features ». In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, 611-17. Geneva, Switzerland: COLING.

¹⁴⁹ Legallois, Dominique, Thierry Charnois, et Thierry Poibeau. 2016. « Repérer les clichés dans les romans sentimentaux grâce à la méthode des « motifs » ». *Lidil. Revue de linguistique et de didactique des langues*, no 53 (mai): 95-117.

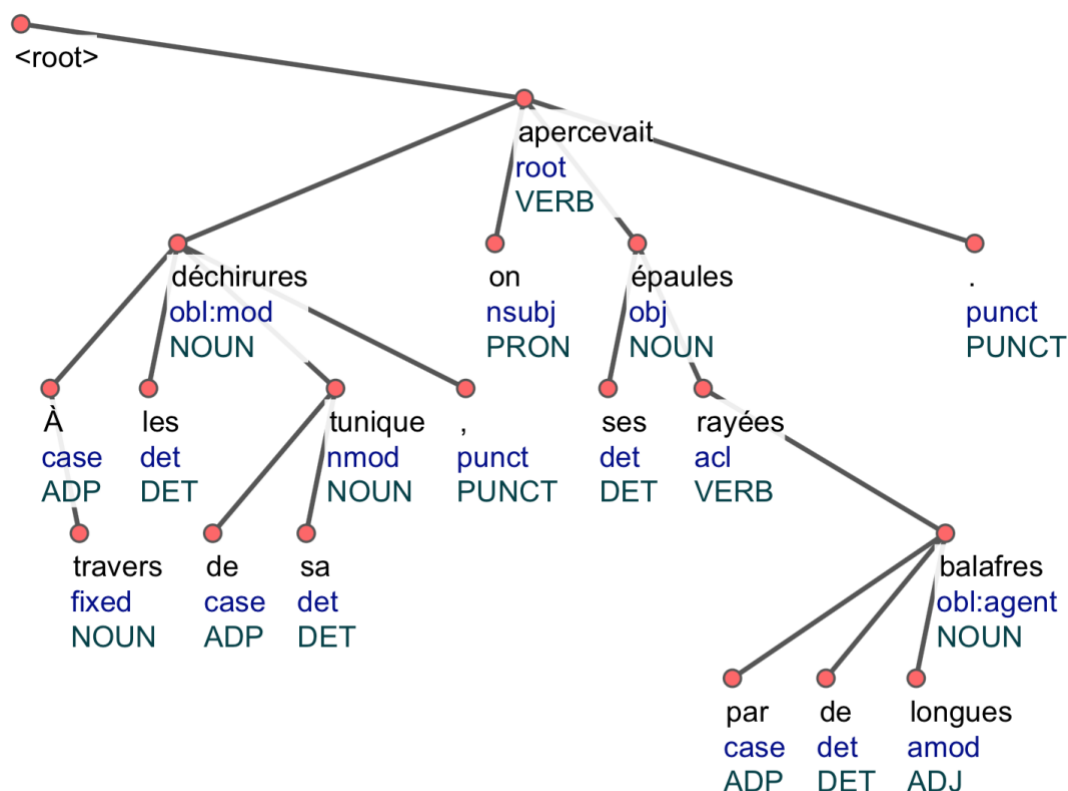
¹⁵⁰ Tesnière, Lucien. 1988 [1959]. *Éléments de syntaxe structurale*. 2e éd. rev. et Corr. Paris: Klincksieck, p. 14.

¹⁵¹ Mel'cuk, Igor A. 1988. *Dependency Syntax: Theory and Practice*. Albany (N.Y.): State University of New York Press.

de représentation similaire avaient été proposés dès le Moyen Âge¹⁵² (Covington, 2009).

La grammaire de dépendance, ou la théorie de dépendance, veut dire que les unités linguistiques (les mots et les ponctuations) sont reliées les unes aux autres par des liens dirigés. Le verbe est considéré comme le centre de la structure de la phrase. Toutes les autres unités syntaxiques sont directement ou indirectement liées au verbe en termes de liens dirigés, appelés dépendances¹⁵³ (Mel'cuk, 1988). Ainsi, les relations syntaxiques dans la phrase sont souvent représentées sous la forme d'un système hiérarchique et arborescent. Par exemple, pour la phrase de Flaubert, tirée de *Salammbô* :

3. *À travers les déchirures de sa tunique, on apercevait ses épaules rayées par de longues balafres.*



¹⁵² Covington, Michael A. 2009. *Syntactic Theory in the High Middle Ages: Modistic Models of Sentence Structure*. Cambridge: Cambridge University Press.

¹⁵³ Melcuk, Igor A. 1988. *op. cit.*

Figure 3.1 Arbre syntaxique de dépendance pour l'exemple de Flaubert.

La grammaire de dépendance permet de dévoiler les relations syntaxiques entre les composants de la phrase, par-delà sa représentation linéaire. Ainsi, la prise en compte de ces relations syntaxiques permet d'obtenir des représentations linguistiquement plus profondes et plus complètes que de simples séquences linéaires des mots, qui ne rendent pas compte de la syntaxe.

La grammaire de dépendance n'est pas la seule théorie à avoir formalisé les structures syntaxiques. Dans les années 1950, Noam Chomsky a proposé la *Phrase Structure Grammar* (appelé également la grammaire de constituants) dans le cadre de sa grammaire transformationnelle¹⁵⁴ (Chomsky, 1957). La grammaire de constituant est un type de grammaire dans laquelle c'est la notion de syntagme (ou constituant) qui joue le premier rôle. L'analyseur cherche à identifier les éléments syntaxiques qui forment des groupes logiques (syntagme nominal, syntagme verbal...) hiérarchiquement, jusqu'à former la structure syntaxique complète de la phrase. Celle-ci repose sur les relations entre constituants. Prenons comme exemple la phrase suivante pour montrer à quoi ressemble une analyse syntaxique des constituants :

4. *Sur la route reste un lion.*

¹⁵⁴ Chomsky, Noam. 1957. *Syntactic structures*. The Hague, Pays-Bas, France.

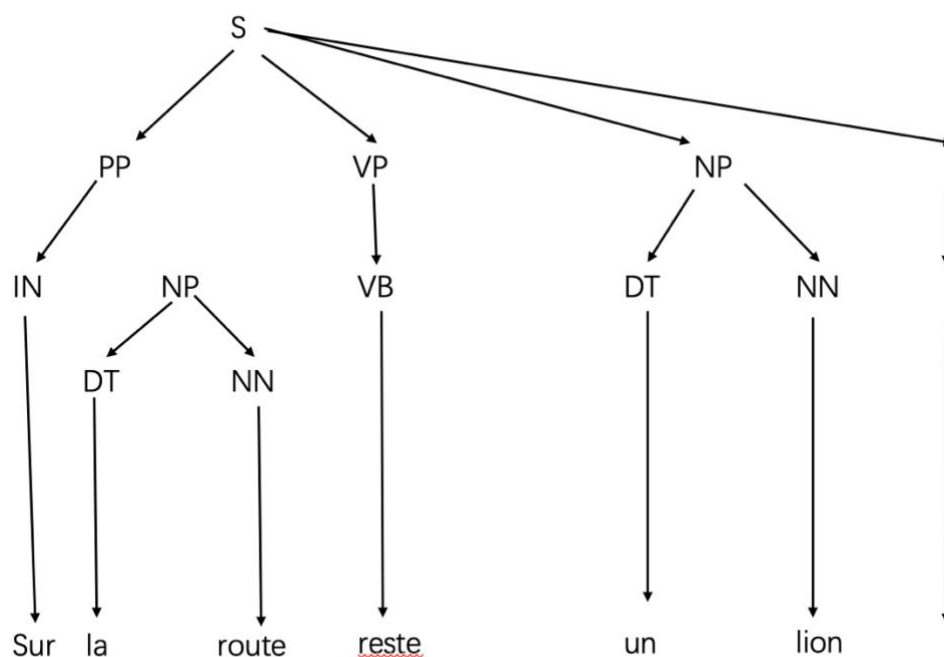


Figure 3.2 Arbre syntaxique de constituants.

On voit sur cet exemple que les relations ne sont pas directement entre les mots, mais que ceux-ci sont regroupés dans ce qu'on appelle des syntagmes, avec des étiquettes particulières (NP pour « noun phrase », ou groupe nominal, VP pour « verb phrase », ou groupe verbal, etc.).

Dans cette thèse, nous avons adopté la notion de grammaire de dépendance plutôt que celle de grammaire de constituants pour les deux raisons suivantes :

1. Les relations de dépendance sont plus utiles pour notre méthode que la notion de constituants. Nous allons présenter dans la section suivante une méthode d'extraction supervisée, c'est-à-dire une méthode qui nous permet d'extraire des figures de style. La reconnaissance des figures de style dépend souvent de la détection des relations de dépendance entre deux ou plusieurs mots dans la phrase. Par exemple, si nous voulons extraire l'inversion du sujet dans l'exemple précédent : « Sur la route reste un lion », on voit que le sujet « lion » et le verbe « reste » sont dans deux constituants (syntagmes) différents et il n'y

a pas de relation directe entre les deux mots. Mais si nous choisissons de représenter les relations syntaxiques par des relations de dépendance, la modélisation et l'extraction de cette inversion du sujet deviennent plus faciles :

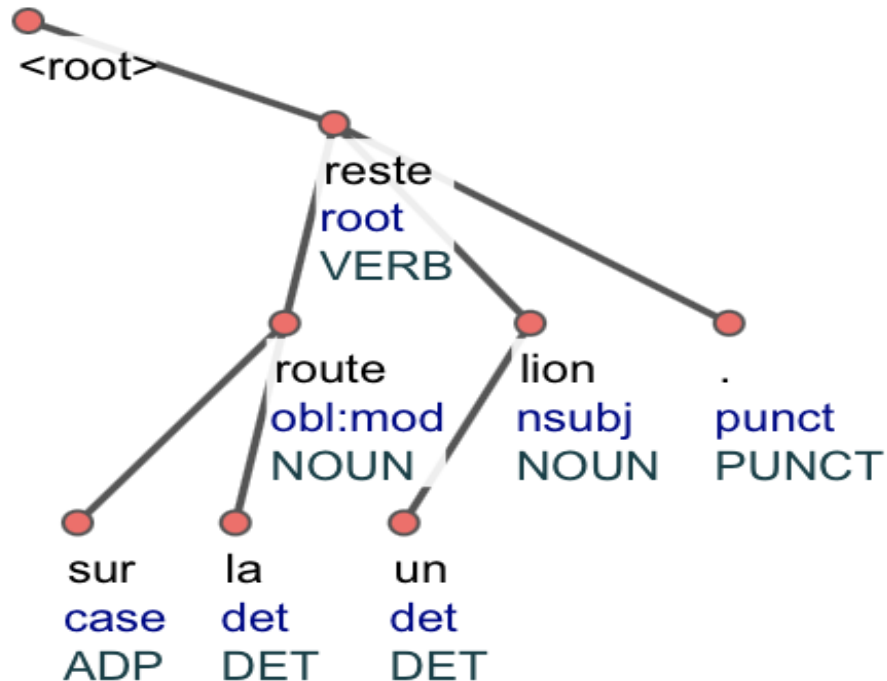


Figure 3.3 Arbre syntaxique de dépendance.

Le verbe et le sujet de cette phrase sont directement liés, avec mention de la fonction grammaticale du nom : « nsubj », c'est-à-dire sujet nominal. Par conséquent, pour extraire cette inversion du sujet, il suffit de remarquer que le « nsubj » est derrière le verbe dont il dépend. Certes, aucune de ces deux grammaires n'est supérieure à l'autre dans la mesure où elles représentent les mêmes relations syntaxiques (elles diffèrent juste par leur façon de représenter l'analyse), cependant, pour notre recherche, la grammaire de dépendance semble plus adaptée que la grammaire de constituants.

2. Les analyses syntaxiques en dépendances sont plus économiques du point de vue de la durée de traitement. L'étude de Cer et al. (2010) montre que les analyseurs syntaxiques en constituants mettent presque 60 fois plus de temps

que les analyseurs en dépendance¹⁵⁵. Nous avons constaté le même phénomène sur notre corpus : l'analyseur syntaxique en dépendances (UDPipe¹⁵⁶) analyse le roman *Les Travailleurs de la Mer* de Victor Hugo en 47 secondes¹⁵⁷ alors que StanfordNLP¹⁵⁸ a pris à peu près 4 heures et demie pour faire l'analyse syntaxique en constituants de ce roman.

Nous avons présenté dans cette partie deux formalismes permettant de représenter les structures syntaxiques d'une phrase : la grammaire de dépendance et la grammaire de constituants. L'exploitation des relations syntaxiques constitue un progrès pour la stylométrie parce que, comme nous l'avons évoqué dans l'état de l'art, des paramètres linguistiques plus profonds donnent souvent de meilleurs résultats que des paramètres linguistiques superficiels. Les méthodes syntaxiques que propose cette thèse se fondent sur la grammaire de dépendance, parce que celle-ci s'adapte mieux à notre démarche et à notre objectif.

¹⁵⁵ Cer, Daniel, Marie-Catherine de Marneffe, Dan Jurafsky, et Chris Manning. 2010. « Parsing to Stanford Dependencies: Trade-offs between Speed and Accuracy ». In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valletta, Malta: European Language Resources Association (ELRA).

¹⁵⁶ Straka, Milan, et Jana Straková. 2017. « Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe ». In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 88-99. Vancouver, Canada: Association for Computational Linguistics.

¹⁵⁷ Tous les résultats des durées de traitement dans la thèse ont été obtenus avec un ordinateur personnel iMac, Intel Core i5, 3.4GHz, 24 GB 2400 MHz DDR4.

¹⁵⁸ Manning, Christopher, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, et David McClosky. 2014. « The Stanford CoreNLP Natural Language Processing Toolkit ». In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 55-60. Baltimore, Maryland: Association for Computational Linguistics.

3.3 Comparaison des performances des différents analyseurs syntaxiques en dépendances pour la langue française.

Après avoir choisi la grammaire de dépendance comme outil d'analyse syntaxique pour notre méthode stylo-métrique, l'étape suivante consiste à choisir un analyseur syntaxique en dépendances dont la performance est la plus satisfaisante pour la langue française. La sélection d'UDPipe comme l'analyseur syntaxique pour nos démarches stylistiques se fonde non seulement sur les résultats publiés par les concepteurs du logiciel, mais aussi sur la vérification de ces performances par rapport aux échantillons de phrases annotées manuellement par nous-mêmes et représentatives de notre objet d'étude. Nous présentons d'abord les évaluations réalisées par les développeurs de trois analyseurs syntaxiques courants de la langue française : Talismane, StanfordNLP et UDPipe, ensuite, nous présentons les performances de ces trois outils sur deux échantillons de phrases (deux échantillons comprenant chacun 50 phrases extraites aléatoirement de *Les Travailleurs de la Mer* de Victor Hugo et de *Madame Bovary* de Flaubert). Les deux échantillons de phrases annotées manuellement sont présentés dans l'Annexe 1.

Talismane est un analyseur syntaxique en dépendances développé par Assaf Urieli¹⁵⁹ (2013). Le développeur précise que cet outil suit les étapes classiques pour l'analyse syntaxique en dépendances : « le découpage en phrases, la segmentation en mots, l'étiquetage (attribution d'une catégorie morphosyntaxique), et le parsing (repérage et étiquetage des dépendances syntaxiques entre les mots) »¹⁶⁰. L'algorithme

¹⁵⁹ Urieli, Assaf. 2013. « Robust French Syntax Analysis : Reconciling Statistical Methods and Linguistic Knowledge in the Talismane Toolkit ». Phdthesis, Université Toulouse le Mirail - Toulouse II.

¹⁶⁰ <http://redac.univ-tlse2.fr/applications/talismane.html>, consulté le 30 avril 2021.

est expliqué dans un article d’Urieli et al.¹⁶¹ (2013) : nous reprenons ici les résultats de l’évaluation de la performance de Talismane publiés dans cet article.

Parseur	LAS ¹⁵ Dev	UAS ¹⁶ Dev	LAS Test	UAS Test	Temps de calcul
Berkeley	86,5	90,8	86,8	91,0	12m46s
MSTParser	87,5	90,3	88,2	90,9	14m39s
MaltParser	86,9	89,4	87,3	89,7	1m25s
Talismane (faisc 1)	86,8	90,2	87,2	90,6	7m56s
Talismane (faisc. 2)	87,3	90,4	88,0	91,0	14m51s
Talismane (faisc. 5)	87,8	90,7	88,3	91,1	38m26s
Talismane (faisc. 10)	88,0	90,8	88,4	91,1	80m36s
Talismane (faisc. 20)	88,1	90,8	88,5	91,1	157m53s

Tableau 3.1 Exactitude (score LAS) et temps de calcul par parseur.

Les analyses en dépendances sont généralement évaluées à l’aide de deux mesures globales : le score d’attachement étiqueté (LAS, labeled attachment score) est la précision de tous les arcs ayant un gouverneur (ou tête) et une étiquette grammaticale correctement annotée. Le score d’attachement sans étiquette (UAS, unlabeled attachment score) est la précision de tous les arcs ayant le gouverneur correct, quelle que soit l’étiquette¹⁶² (Kübler, 2008). Dans le tableau, « dev » et « test » sont deux sous-corpus du French TreeBank¹⁶³.

Étant donné que notre méthode d’extraction se fonde sur l’identification des relations de dépendance et du gouverneur, le score LAS est le plus important pour nous. Or, nous constatons dans le tableau 3.1 que le meilleur score obtenu par Talismane est 88,5 % (sur le sous-corpus « test »), si nous prenons en compte tous les arcs ayant le gouverneur et l’étiquette corrects. Cependant, sur deux des romans dans notre corpus :

¹⁶¹ Urieli, Assaf, et Ludovic Tanguy. 2013. « L’apport du faisceau dans l’analyse syntaxique en dépendances par transitions : études de cas avec l’analyseur Talismane ».

¹⁶² Kübler, Sandra, Ryan McDonald, et Joakim Nivre. 2008. « Dependency Parsing ». *Synthesis Lectures on Human Language Technologies* 2 (1): 1-127.

¹⁶³ Pour savoir plus sur French Treebank, voir <http://ftb.linguist.univ-paris-diderot.fr>.

Les Travailleurs de la Mer de Victor Hugo et *Madame Bovary* de Flaubert, après la vérification manuelle de deux échantillons comprenant chacun 50 phrases, nous obtenons une exactitude inférieure à ce qui est affiché sur le tableau : 82,5 % au lieu de 88,5 %.

Intéressons-nous maintenant à l'analyseur de Stanford. La version 0.2.0 du StanfordNLP obtient (en 2018) un score LAS de 86,38 % pour le corpus French-GSD et 88,51 % pour le corpus French-Sequoia¹⁶⁴ :

TREEBANK	TOKENS	SENTENCES	WORDS	UPOS	XPOS	UFEATS	ALLTAGS	LEMMAS	UAS	LAS	CLAS	MLAS	BLEX
UD_French-GSD	99.67	95.4	99.37	96.97	99.37	96.61	95.77	96.97	89.21	86.38	82.43	78.08	79.54
UD_French-Sequoia	99.85	90.61	99.65	98.43	99.65	97.8	97.36	97.42	90.42	88.51	85.13	82.27	82.33

Tableau 3.2 Exactitude du StanfordNLP pour l'analyse syntaxique en dépendances.

Le corpus French-GSD contient 16 341 phrases et 389 376 tokens¹⁶⁵, le corpus French-Sequoia contient 3099 phrases et 68 615 tokens¹⁶⁶. Nous ne présentons pas les performances sur le corpus French-spoken (corpus de français oral) vu que notre propre corpus ne comprend que des œuvres écrites. Nous avons testé également StanfordNLP sur les deux échantillons de phrases, en faisant la comparaison avec ce que nous avons annoté manuellement : l'exactitude de StanfordNLP s'élève à 85 % sur nos échantillons.

Enfin, UDPipe publie sur son site Internet les performances sur trois corpus en langue française : French-GSD, French-ParTUT et French-Sequoia¹⁶⁷ :

¹⁶⁴ <https://stanfordnlp.github.io/stanfordnlp/performance.html>, consulté le 30 avril 2021.

¹⁶⁵ https://universaldependencies.org/treebanks/fr_gsd/, consulté le 30 avril 2021.

¹⁶⁶ https://universaldependencies.org/treebanks/fr_sequoia/, consulté le 30 avril 2021.

¹⁶⁷ <http://ufal.mff.cuni.cz/udpipe/models>, consulté le 7 septembre 2020.

Treebank	Mode	Words	Sents	UPOS	XPOS	UFeats	AllTags	Lemma	UAS	LAS	MLAS	BLEX
French-GSD	Raw text	98.8%	93.6%	95.8%	-	95.5%	94.5%	96.6%	87.1%	84.3%	74.4%	76.3%
French-GSD	Gold tok	-	-	97.1%	-	96.6%	95.6%	97.8%	89.0%	86.3%	76.0%	77.7%
French-GSD	Gold tok+mor	-	-	-	-	-	-	-	91.1%	89.2%	83.9%	84.2%
French-ParTUT	Raw text	99.4%	100.0%	94.6%	94.2%	91.9%	90.6%	95.0%	87.4%	83.5%	66.4%	71.5%
French-ParTUT	Gold tok	-	-	95.3%	94.8%	92.6%	91.3%	95.6%	88.0%	84.3%	67.4%	72.0%
French-ParTUT	Gold tok+mor	-	-	-	-	-	-	-	91.0%	89.0%	82.1%	83.4%
French-Sequoia	Raw text	99.1%	87.5%	96.1%	-	95.0%	94.1%	96.9%	84.8%	82.1%	72.5%	75.1%
French-Sequoia	Gold tok	-	-	97.1%	-	95.9%	95.0%	97.8%	86.8%	84.1%	74.5%	76.8%
French-Sequoia	Gold tok+mor	-	-	-	-	-	-	-	89.7%	88.1%	83.8%	84.0%

Tableau 3.3 Exactitude d’UDPipe pour l’analyse syntaxique en dépendances.

La comparaison avec les deux échantillons d’annotation manuelle de notre corpus révèle une exactitude de 89,3 %, supérieure à celles publiées par les développeurs sur les textes bruts des trois corpus en langue française (84,3 %, 83,5 % et 82,1 %). Ainsi, UDPipe est l’outil le plus performant parmi les trois analyseurs syntaxiques en dépendances que nous avons choisi de mettre en comparaison.

Nous avons présenté dans cette partie les performances de trois analyseurs syntaxiques en dépendances pour la langue française. Il existe évidemment d’autres analyseurs qui accomplissent cette tâche d’analyse syntaxique, les trois que nous avons choisis sont censés être les plus performants d’après les résultats publiés récemment. Il est important de remarquer et d’expliquer deux phénomènes avant de présenter dans la partie suivante notre méthodologie.

Premièrement, d’après notre vérification manuelle, deux des analyseurs n’atteignent pas l’exactitude publiée par leurs développeurs. Nous croyons que cela est dû à la différence entre le corpus sur lequel ont été entraînés ces analyseurs et notre corpus, qui est composé de romans du 19^e siècle. Par exemple, le corpus Séquoia sur lequel sont testés StanfordNLP et UDPipe se compose de phrases qui proviennent de : « Europarl français, [du] journal l’Est Républicain, [de] Wikipédia Fr et [de] documents

de l'Agence Européenne du Médicament¹⁶⁸ » (Candito et al. 2012), soit des textes plus récents que ceux de notre corpus. Il est donc normal qu'en appliquant le même analyseur à des romans français du 19^e siècle, il y ait une différence de performance.

Deuxièmement, pendant le processus de vérification manuelle, nous avons constaté que les erreurs de chaque analyseur sont souvent homogènes. Par exemple, si UDPipe confond « obj » (le COD) avec « obl : mod » (le modifieur) dans une phrase de Victor Hugo, il commettra la même erreur dans les autres phrases de Victor Hugo, et il commettra également cette même erreur dans les phrases d'un autre auteur. Ceci réduit potentiellement considérablement les conséquences de cette erreur parce que : 1. Si l'utilisateur n'exploite pas « obj » et « obl : mod », alors l'erreur peut être ignorée ; 2. Si l'utilisateur a besoin de ces deux étiquettes grammaticales, la correction peut se faire très rapidement en suivant la régularité des occurrences des erreurs.

Nous pouvons donc en conclure que les outils informatiques ont toujours des limites et il est important d'évaluer leurs avantages et leurs défauts vis-à-vis de notre propre projet de recherche (les analyseurs syntaxiques ne sont jamais exacts à 100 %). Ces erreurs restent toutefois marginales, et il faut en regard apprécier les nouvelles possibilités qu'apportent ces analyseurs, en permettant d'automatiser l'analyse syntaxique. Ainsi, dans la thèse, les performances d'UDPipe nous permettent de repérer des patrons syntaxiques et stylistiques (les inversions du sujet, les répétitions et les mises en relief par le clivage) que nous allons présenter dans ce chapitre. Ceci ne serait pas possible sans analyseur linguistique performant.

¹⁶⁸ Candito, Marie, et Djamé Seddah. 2012. « Le corpus Sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical ». In *TALN 2012 - 19e conférence sur le Traitement Automatique des Langues Naturelles*. Grenoble, France.

3.4 Les patrons caractérisants et les patrons identificatoires

La thèse exploite deux types de traits syntaxiques, pour avoir une vue générale des caractéristiques stylistiques des textes de notre corpus. Les deux types de traits ont été présentés et discutés en détail dans le Chapitre 2. Il s'agit de deux types de paramètres linguistiques qui fondent la base de deux approches stylistiques différentes, d'après Stanley Fish. Les traits caractérisants sont supposés avoir en eux-mêmes une fonction stylistique immédiatement visible par le lecteur ; les traits identificatoires n'ont pas toujours une signification stylistique, mais ils permettent de différencier les textes grâce à leur prééminence statistique. Nous avons vu (à la fin de la section 2.2.6 du Chapitre 2) qu'il serait impossible de demander à un seul type de traits d'accomplir les deux fonctions stylistiques, à savoir celle d'identifier et de caractériser le style. C'est pour cette raison que notre méthode d'analyse a recours à ces deux types de traits syntaxiques, pour permettre d'appliquer les deux variétés d'analyse à notre corpus : les patrons syntaxiques caractérisants exploitent la possibilité de repérer les figures de style à même les textes ; les patrons syntaxiques identificatoires extraient des structures syntaxiques caractéristiques d'une manière non-supervisée, d'après leur fréquence d'occurrences remarquable dans les textes. Dans cette partie nous présentons la nécessité de combiner les deux types de patrons pour notre étude stylistique sur les romans français du 19^e siècle.

La partie 3.6 est consacrée à l'identification et à l'extraction des trois patrons syntaxiques caractérisants : l'inversion du sujet, la répétition et la mise en relief par le clivage. Les trois patrons caractérisants sont essentiellement des figures de style mettant en jeu la syntaxe. Nous avons choisi d'explorer les patrons caractérisants à travers la notion de figure de style. En effet, les figures de style pourraient être considérées comme un exemple idéal de la notion d'« écart » que nous avons adoptée : « le style se définit d'une part, comme l'écart de la norme linguistique par rapport aux formes

standards, et d'autre part, comme la récurrence de l'apparition de ces écarts¹⁶⁹ » (Molinié, 1994). L'idée de considérer les figures de style comme un écart à la norme pourrait remonter à Quintilien, quand il écrit que les figures sont « un changement rationnel de sens ou de langage par rapport à la forme ordinaire et simple [*in sensu vel sermone aliqua a vulgari et simplici specie cum ratione mutatio*]¹⁷⁰ » (Quintilien, 1921, III : 353). Cette définition a été souvent reprise dans l'histoire de la stylistique, par exemple, dans la définition d'Henry Peacham pour les figures de style : « Une figure est une forme de mots, d'oraison ou de phrase, renouvelée par l'art, différente de la manière vulgaire et de la coutume d'écrire ou de parler »¹⁷¹ (Henry Peacham, 2001). Dans la définition donnée par Littré, les figures de style se définissent encore par la comparaison : « Certaines formes de langage (...) donnent au discours plus de grâce et de vivacité, d'éclat et d'énergie¹⁷² » (Suhamy, 2020, p.9).

Ces définitions considèrent les figures de style comme un écart à une norme dans les textes et mettent l'accent sur le côté formel de ces unités linguistiques en permettant aux lecteurs de détecter leurs occurrences. Les figures de style, comme patrons caractérisants, n'ont pas besoin de justifier leur spécificité : elles ont justement été promues au rang de figures de style de par leur fonction stylistique. La nécessité d'étudier la signification des figures de style (au-delà de leur simple repérage) a toutefois été démontrée par des stylisticiens comme Henry Suhamy, qui écrit dans la conclusion *Des Figures de Style* :

« L'étude des figures n'est pas une fin en soi. Elle permet de mieux comprendre le sens et la forme des énoncés, de déceler certains codages et

¹⁶⁹ Molino, Jean. 1994. « Pour une théorie sémiologique du style », dans Molinié, Georges Organisateur de réunion, Pierre, Cahné Organisateur de réunion, Colloque international Qu'est-ce que le style ?, et École doctorale Sciences du langage. 1994. *Qu'est-ce que le style ? : actes du colloque international*. Paris, France: Presses universitaires de France.

¹⁷⁰ Quintilian. 1921. *The Instituto Oratoria*. With an English Translation by H.E. Butler. London: Heinemann.

¹⁷¹ Peacham, Henry. 2001. *The Garden of Eloquence (1593)*. Ann Arbor: Scholar Press.

¹⁷² Suhamy, Henri. 2020. *Les figures de style*. Vol. 14e éd. Que sais-je ? Paris cedex 14: Presses Universitaires de France, p. 9.

travestissements conventionnels de l'expression. Elle compense l'étroitesse vétilleuse de ses visées par l'étendue pluridisciplinaire de ses champs d'application : la rhétorique, la poésie, la littérature, mais aussi l'étymologie, la grammaire, la psychologie du langage, la communication.¹⁷³ » (Suhamy, 2020, p.119)

Ou encore, dans l'ouvrage de Jeanne Fahnestock, *Rhetorical Figures in Science*, l'auteur confirme leur fonction constitutive de la signification du texte :

« Les figures de style occupent un territoire examiné sur ce continuum de possibilités expressives. Elles ont traditionnellement été utilisées (et donc identifiées) comme des moyens particulièrement efficaces de dire ce qui peut également être dit de bien d'autres manières. Elles sont, en d'autres termes, idéalement, mais pas exclusivement constitutives des multiples significations d'un texte. Elles font simplement ce qu'il est possible de faire avec d'autres moyens, ou même font au mieux ce qu'elles sont traditionnellement identifiées à faire.¹⁷⁴ » (Fahnestock, 1999, p.22)

Notre étude de la fonction stylistique des figures de style sera non-exhaustive. L'inversion, la répétition et le clivage ne dévoilent que partiellement des caractéristiques stylistiques des textes dans le corpus : ces figures révèlent les préférences des auteurs à utiliser ces trois outils stylistiques (et syntaxiques) au lieu de recourir à « d'autres moyens », ou d'« autres manières ».

¹⁷³ Suhamy, Henri. 2020. *Les figures de style*. Vol. 14e éd. Que sais-je ? Paris cedex 14: Presses Universitaires de France, p. 119.

¹⁷⁴ « The figures of speech occupy examined territory on this continuum of expressive possibilities. They have traditionally been used (and hence identified) as particularly effective ways of saying what can also be said in many other ways. They are, in other words, ideally but not exclusively constitutive of a text's many meanings, simply doing better what it is possible to do with other means, or even doing best what they are traditionally identified as doing ». Fahnestock, Jeanne. 1999. *Rhetorical Figures in Science*. Oxford University Press, p. 22.

L'identification des figures de style est un des sujets de recherche de la stylométrie. Dans le mémoire de Jakub Gawryjolek¹⁷⁵ (2009), l'auteur propose une méthode d'extraction supervisée pour extraire les figures de style liées à la notion de répétition. Cette méthode supervisée se base sur la création d'un outil d'annotation appelé JANTOR (Java ANnotation Tool Of Rhetoric), qui permet l'annotation manuelle et automatisée de fichiers au format HTML. Un analyseur syntaxique est au préalable appliqué sur les phrases, avant de lancer le processus d'extraction. En combinant les résultats de l'analyse syntaxique en dépendances et de l'annotation automatique et manuelle, l'auteur montre qu'il obtient des résultats prometteurs.

La thèse de Marie Dubremetz¹⁷⁶ (2017) présente une méthode d'extraction non-supervisée des figures de style de répétition : le chiasme¹⁷⁷, l'épanaphore¹⁷⁸ et l'épiphore¹⁷⁹. Des données représentatives sont annotées et les extractions sont réalisées par entraînement sur ces données annotées manuellement. L'auteur montre que même avec un corpus annoté très incomplètement, un système de détection de figures répétitives peut être obtenu, avec une précision raisonnable. Elle applique le système à quatre types de textes différents : discours politique, fictions, titres d'articles et de romans et citations. L'évaluation montre que le système est robuste aux changements de genre et que les fréquences des trois figures rhétoriques varient selon le genre.

Les deux études ci-dessus présentent deux essais de détection automatique des figures de style. La méthode d'extraction que nous présentons pour l'inversion du sujet, la répétition et le clivage est aussi une méthode d'extraction supervisée. Cependant,

¹⁷⁵ Gawryjolek, Jakub Jan. 2009. « Automated Annotation and Visualization of Rhetorical Figures », mai. <https://uwspace.uwaterloo.ca/handle/10012/4426>.

¹⁷⁶ Dubremetz, Marie. 2017. « Detecting Rhetorical Figures Based on Repetition of Words: Chiasmus, Epanaphora, Epiphora ». <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-334486>.

¹⁷⁷ « Figure de style consistant à inverser l'ordre des termes dans les parties symétriques de deux membres de phrase de manière à former un parallèle ou une antithèse ». <https://www.cnrtl.fr/definition/chiasme>, consulté le 28 juillet, 2021.

¹⁷⁸ « Figure de style consistant en la répétition d'une même formule au début de phrases ou de segments de phrase successifs, dans la même structure syntaxique ». <https://fr.wikipedia.org/wiki/Épanaphore>, consulté le 28 juillet, 2021.

¹⁷⁹ « Figure de style consistant en la répétition, à la fin de deux ou de plusieurs groupes de phrases ou de vers qui se succèdent, d'un même mot ou d'un même groupe de mots ». <https://fr.wikipedia.org/wiki/Épiphore>, consulté le 28 juillet, 2021.

notre méthode n'exige pas de disposer d'annotations préalables avant de lancer le processus d'extraction, elle ne requiert qu'une analyse syntaxique en dépendances qui nous permet de modéliser et d'extraire les trois figures de style qui nous intéressent. De plus, contrairement aux deux études ci-dessus, notre méthode ne se limite pas au repérage des figures de style de répétition : l'inversion du sujet et le clivage montrent que la grammaire de dépendance est utile pour détecter les figures de style mettant aussi en jeu l'ordre des mots.

L'objectif de la thèse n'est pas de proposer une étude systématique de l'extraction automatique de figures de style. Ainsi, la méthode que nous proposons dans les parties suivantes n'est pas exhaustive (au sens où il est impossible de modéliser et extraire toutes les figures de style automatiquement et parfaitement) ; il ne s'agit pas non plus d'une étude générale sur les figures de style de tous niveaux linguistiques (lexicaux, syntaxiques ou sémantiques). Nous voulons simplement montrer que l'analyse syntaxique en dépendances, sur laquelle se fondent toutes les méthodes d'extraction de la thèse, permet d'identifier des figures de style dont les fonctions stylistiques dévoilent non seulement les caractéristiques stylistiques du texte, mais aussi la préférence personnelle des auteurs pour ces figures.

Cependant, comme nous allons le constater à travers l'application de notre méthode sur le corpus (discutée dans les chapitres 4 et 5), la généralisation des caractéristiques stylistiques et des préférences personnelles est difficile, en partie parce que nous avons pu automatiser l'extraction de trois types de figures de style seulement. C'est pour cette raison que nous proposons aussi d'identifier, en parallèle et toujours à partir de l'analyse syntaxique, les patrons identificatoires sur la base de leur spécificité statistique. Les patrons identificatoires (appelés aussi les patrons syntaxiques hiérarchiques, parce qu'ils sont fondés sur l'analyse des relations hiérarchiques dans la phrase) sont des structures syntaxiques dont les fréquences d'occurrences sont statistiquement significatives pour un certain texte par rapport à d'autres. Ces patrons possèdent en quelque sorte les mêmes significations statistiques que les mots

fonctionnels (souvent caractéristiques d'un auteur par rapport à un autre, comme on l'a vu précédemment) tout en étant linguistiquement et sémantiquement interprétables.

3.5 Une nouvelle méthode pour mesurer la complexité syntaxique

L'objectif de cette étude sur la complexité syntaxique est de fournir une vue générale sur les caractéristiques de la syntaxe, avant d'étudier les patrons syntaxiques. Cette vue macroscopique peut donner une première idée de la spécificité du style d'un auteur.

Notre méthode pour mesurer la complexité syntaxique se fonde, comme toutes les autres méthodes proposées dans cette thèse, sur la structure hiérarchique de la phrase (par opposition à sa structure linéaire directement accessible). La nécessité d'explorer les relations syntaxiques pour identifier les différentes propositions dans les phrases complexes a été confirmée par les études précédentes sur ce sujet. Givón (2009, p.4) par exemple relie la notion de complexité syntaxique à « l'organisation hiérarchique » dans les phrases :

« L'essentiel du travail fondateur d'Herbert Simon (1962) sur l'architecture de la complexité est qu'une complexité accrue est, au niveau le plus général, une organisation hiérarchique accrue ; c'est-à-dire une augmentation du nombre de niveaux hiérarchiques au sein d'un système¹⁸⁰ ». (Givón, 2009, p. 4)

3.5.1 Les méthodes précédentes pour mesurer la complexité syntaxique

La méthode la plus « superficielle » pour mesurer la complexité syntaxique est sans doute celle se limitant à observer la longueur de phrases. G. Udney Yule (1939) utilise

¹⁸⁰ « The gist of Herbert Simon's (1962) seminal work on the architecture of complexity is that increased complexity is, at the most general level, increased hierarchic organization; that is, an increase in the number of hierarchic levels within a system ». Givón, Talmy. 2009. *The Genesis of Syntactic Complexity: Diachrony, Ontogeny, Neuro-Cognition, Evolution*. Amsterdam Philadelphia: John Benjamins Pub. Co.

ainsi la longueur de phrases comme paramètre stylométrique pour différencier avec succès les œuvres d'auteurs différents¹⁸¹. D'autres tentatives sur la longueur de phrases sont faites par Williams¹⁸² (1940) et par Wake¹⁸³ (1957). Les méthodes fondées sur ce paramètre sont dites superficielles parce qu'elles considèrent les phrases longues comme syntaxiquement plus complexes que les phrases courtes, et qu'elles ne nécessitent que de compter le nombre de mots dans chaque phrase. La longueur des phrases est effectivement un paramètre important, mais elle ne répond que partiellement à la question de la complexité syntaxique.

Biber, et al. (1998, p. 178) se réfèrent à l'idée de T-unit, à savoir « une clause indépendante avec toutes ses clauses dépendantes¹⁸⁴ » pour mesurer la complexité syntaxique. Le T-unit est essentiellement un paramètre pour mesurer la complexité syntaxique par le nombre de propositions subordonnées dans une phrase. Hallen et Shakespear (2002) reprennent ce paramètre et utilisent le T-unit pour interpréter les poèmes d'Emily Dickinson. Ils concluent que :

« La longueur de T-unit peut servir de mesure de la complexité syntaxique et de la maturité cognitive chez un écrivain ou un auteur. Les écrivains moins expérimentés ont tendance à joindre des phrases plus courtes dans un style paratactique, tandis que les écrivains expérimentés ont tendance à combiner des phrases en clauses principales plus longues avec des clauses dépendantes subordonnées dans un style hypotactique¹⁸⁵ ». (Hallen, et al., 2002, p. 91)

¹⁸¹ Yule, G. Udny. 1939. « On Sentence-Length as a Statistical Characteristic of Style in Prose: With Application to Two Cases of Disputed Authorship ». *Biometrika* 30 (3/4): 363-90.

¹⁸² Williams, C. B. 1940. « A Note on the Statistical Analysis of Sentence-Length as a Criterion of Literary Style ». *Biometrika* 31 (3/4): 356-61.

¹⁸³ Wake, William C. 1957. « Sentence-Length Distributions of Greek Authors ». *Journal of the Royal Statistical Society. Series A (General)* 120 (3): 331-46.

¹⁸⁴ Biber, Douglas, Susan Conrad, and Randi Reppen. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge Univ. Press, 2012. Internet resource.

¹⁸⁵ Hallen, Cynthia L, et Jennifer Shakespear. 2002. « The T-Unit as a Measure of Syntactic Complexity in Emily Dickinson's Poems ». *Emily Dickinson Journal* 11 (1): 91-103.

Graesser, et al. (2005) proposent une combinaison de différents paramètres pour mesurer la complexité syntaxique. Les facteurs qui contribuent à la « complexité syntaxique » comprennent la mesure de la quantité de modificateurs dans les groupes nominaux, la longueur de la phrase avant le verbe principal et la similitude des phrases entre elles. À l'aide de ces paramètres, ils analysent la complexité de différents textes, déclarant, par exemple, que « les paramètres révèlent que *Little Women* [le roman de Louisa May Alcott publié en 1868] est moins complexe que *The Great Gatsby* [le roman de Francis Scott Key Fitzgerald publié en 1925]¹⁸⁶ » (Graesser, et al., 2004).

Szmrecsanyi (2004) pense que la complexité syntaxique peut également être mesurée par le nombre de mots sous un nœud (à savoir une branche d'un arbre syntaxique). D'autres possibilités incluent le comptage d'autres éléments tels que les conjonctions de subordination, les verbes non finis, etc. Szmrecsanyi se sert d'un petit corpus de 50 phrases, étiqueté et analysé, de source écrite ou parlée. Il essaie les différentes méthodes dont il a parlé : la longueur de phrases, un indice combiné de complexité syntaxique et le nombre de nœuds dans un arbre syntaxique. Il constate « que les trois mesures structurelles classent les données de manière étonnamment conforme¹⁸⁷ » (Szmrecsanyi, 2004, p. 1036).

¹⁸⁶ Graesser, Arthur C, Danielle S Mcnamara, Max M Louwerse, et Zhiqiang Cai. 2004. « Coh-Matrix: Analysis of Text on Cohesion and Language ». *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc.* 36 (2): 193.

¹⁸⁷ Szmrecsanyi, Benedikt, Gérard Purnelle, Gérard Fairon, et Anne Dister. 2004. « On Operationalizing Syntactic Complexity ». *Le Poids Des Mots. Proceedings of the 7th International Conference on Textual Data Statistical Analysis*. Louvain-La-Neuve, March 10-12, 2004, 1032-39, p. 1036)

3.5.2 Présentation de la méthode

Notre méthode pour mesurer la complexité syntaxique se fonde sur les résultats de l'analyse syntaxique en dépendances. Prenons comme exemple la phrase suivante de Victor Hugo dans *Les Travailleurs de la Mer* :

5. *Dans les îles de la Manche, un hiver où il gèle à glace est mémorable, et la neige fait évènement.*

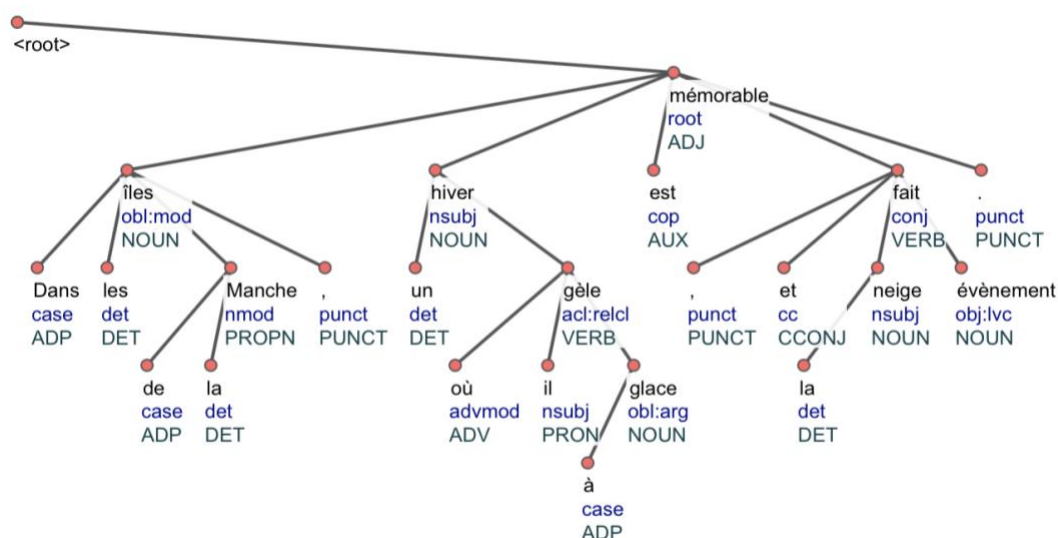


Figure 3.4 Arbre syntaxique de dépendance pour la phrase de Victor Hugo.

Nous considérons les relations de dépendance de cette phrase comme émanant du nœud central «root». L'expansion de la phrase se manifeste de deux façons : l'expansion en profondeur et l'expansion en largeur. Logiquement, pour parcourir un arbre-phrase, le temps d'analyse dépend du temps de son parcours (avec une priorité prédéfinie pour d'abord s'intéresser aux branches gauches ou droites). Ainsi, la complexité de cette structure syntaxique est fonction du nombre de branches que l'arbre possède. Pour évaluer la complexité de la phrase, nous pouvons ainsi chercher d'une

part la branche la plus longue (la profondeur de la structure syntaxique) et d'autre part, l'endroit où il y a le plus de branches (la largeur de l'arbre).

Nous proposons d'utiliser l'étendue de cette expansion syntaxique comme indice de complexité syntaxique. Pour mesurer l'étendue de la structure syntaxique, il suffit de mesurer la profondeur et la largeur de l'arbre syntaxique. Dans notre exemple, la branche verticale la plus profonde s'étend à 5 niveaux : à partir de « mémorable », passer par « hiver », « gèle », « glace », pour arriver enfin à « à », alors que la largeur maximum est atteinte par la ligne horizontale qui traverse « Dans », « les », « Manche », « , », « un », « gèle », « , », « et », « neige », « événement ». Ainsi, la largeur de la structure syntaxique est 10. Nous multiplions la profondeur et la largeur pour créer un indice de complexité syntaxique. Nous nommons cet indice « LPIndice », une nouvelle mesure de la complexité syntaxique basée sur la complexité de la structure de l'arbre syntaxique.

Reprenons l'exemple de Hugo, dans *Les Travailleurs de la mer* : « Dans les îles de la Manche, un hiver où il gèle à glace est mémorable, et la neige fait événement. » Nous remarquons que la branche la plus profonde se trouve dans la deuxième proposition de cette phrase : « un hiver où il gèle à glace est mémorable », cette proposition contient une proposition subordonnée (« où il gèle à glace »). Selon la grammaire de dépendance, les subordinations sont considérées comme dépendantes de la proposition principale, ainsi, les subordinations rallongent verticalement l'arbre syntaxique de dépendance. La profondeur syntaxique que nous mesurons dans une phrase est en fait la proposition qui possède le plus de dépendants verticaux, manifestée majoritairement par les structures subordonnées.

De façon similaire, nous avons gardé le paramètre « largeur » dans les résultats parce que ce paramètre révèle les syntagmes souvent juxtaposés ou coordonnés dans une phrase. Dans notre exemple, « Dans les îles de la manche » est un syntagme prépositionnel juxtaposé avec la proposition principale « un hiver où il gèle à glace est mémorable » ; « et la neige fait événement » est en coordination avec la proposition

principale. Tous les deux élargissent l'étendue horizontale de l'arbre syntaxique. Ainsi, le paramètre « largeur » pourrait être utilisé comme un indice servant à mesurer les phénomènes de coordination ou de juxtaposition dans une phrase.

La combinaison des trois paramètres : « profondeur », « largeur » et « LPIndice » donnent une vue générale de la complexité syntaxique.

3.6 Les extractions des patrons caractérisants

Nous avons évoqué notre choix d’avoir recours à la grammaire de dépendance pour analyser les phrases de notre corpus, nous avons également montré qu’UDPipe obtenait les meilleures performances pour cette tâche (cf. section 3.3). Dans la partie précédente, nous avons argumenté sur la nécessité de prendre en considération à la fois les patrons caractérisants et les patrons identificatoires dans notre démarche stylistique. Ainsi, notre approche visant à identifier et à extraire les patrons syntaxiques caractérisants est fondée sur les résultats de l’analyse syntaxique en dépendance réalisés par UDPipe en appliquant le modèle « french-gsd-ud-2.5-191206.udpipe¹⁸⁸ ». Le choix de ce modèle entraîné sur le corpus French-GSD est lié au fait que son score LAS (labeled attachment score, voir la note 19) est le meilleur (84,3 %) par rapport au modèle French-Sequoia (82,1 %) et French-ParTuT (83,5 %) ¹⁸⁹.

Nous ne présentons dans cette partie que les méthodes pour identifier et extraire les patrons caractérisants. Les résultats d’extraction ainsi que les analyses qui s’ensuivent seront présentés dans les chapitres 4 et 5, où ces méthodes seront appliquées d’abord sur les romans de Victor Hugo, ensuite sur les romans des quatre auteurs de notre corpus. Nous précisons pour chaque patron caractérisant les étapes à suivre pour l’identifier, l’algorithme pour l’extraire et enfin le nettoyage manuel. Chaque résultat est finalement discuté et mis en perspective.

¹⁸⁸ <https://ufal.mff.cuni.cz/udpipe/1/models>, consulté le 4 mai, 2021.

¹⁸⁹ *Idem.*

3.6.1 L'identification et l'extraction des inversions du sujet

Nous parlons d'inversion du sujet quand le sujet est placé derrière le verbe dans une proposition ou une phrase¹⁹⁰ (Tellier et al. 2006). Repérer les inversions du sujet est possible grâce à l'analyse syntaxique en dépendances. D'après les règles de la grammaire de dépendance, le verbe et son sujet sont directement liés par une relation syntaxique (contrairement à la grammaire de constituants, voir la section 3.2 pour un exemple). Le verbe dans cette construction peut être « root » (la racine¹⁹¹ de la phrase, cf. exemple 6), il peut aussi s'agir du verbe d'une des propositions d'une phrase complexe (exemple 7) :

6. *Tout à côté **se penche un vieux pommier** malade pansé avec un bandage de paille et de terre glaise.* (Flaubert)
7. *Ce qui est certain, c'est que, d'ordinaire, après les vainqueurs **viennent les voleurs.*** (Flaubert)

¹⁹⁰ Tellier, Christine, et Daniel Valois. 2018. *Constructions méconnues du français. Constructions méconnues du français*. Paramètres. Montréal : Presses de l'Université de Montréal.

¹⁹¹ <https://universaldependencies.org/u/dep/root.html>, consulté le 4 mai, 2021.

Tout à côté se penche un vieux pommier malade pansé avec un bandage de paille et de terre glaise .

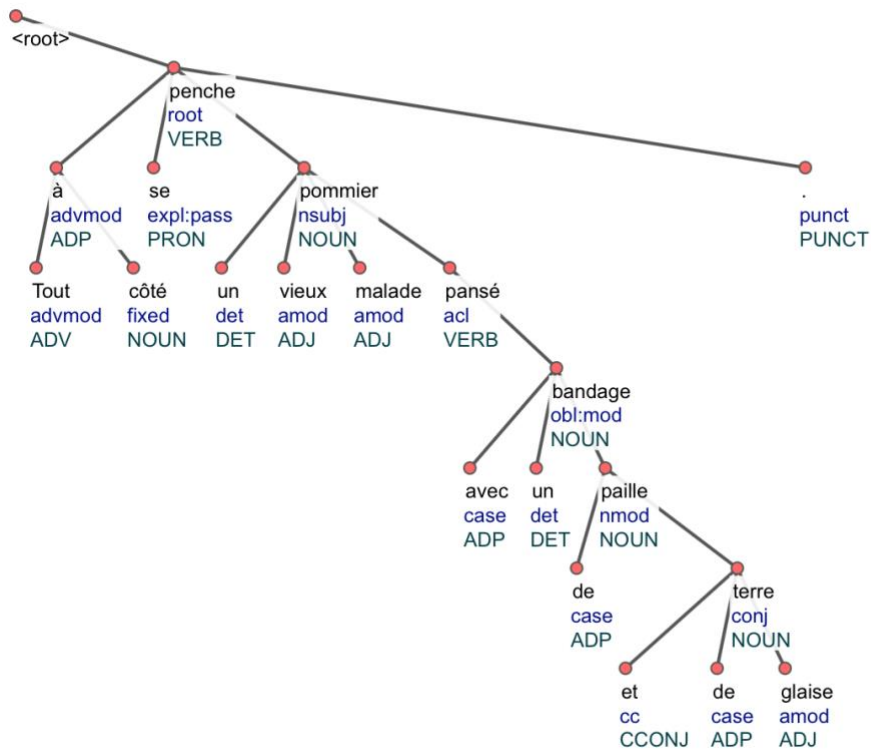


Figure 3.5 Arbre syntaxique de dépendance pour l'exemple 6.

Ce qui est certain , c' est que , d' ordinaire , après les vainqueurs viennent les voleurs .

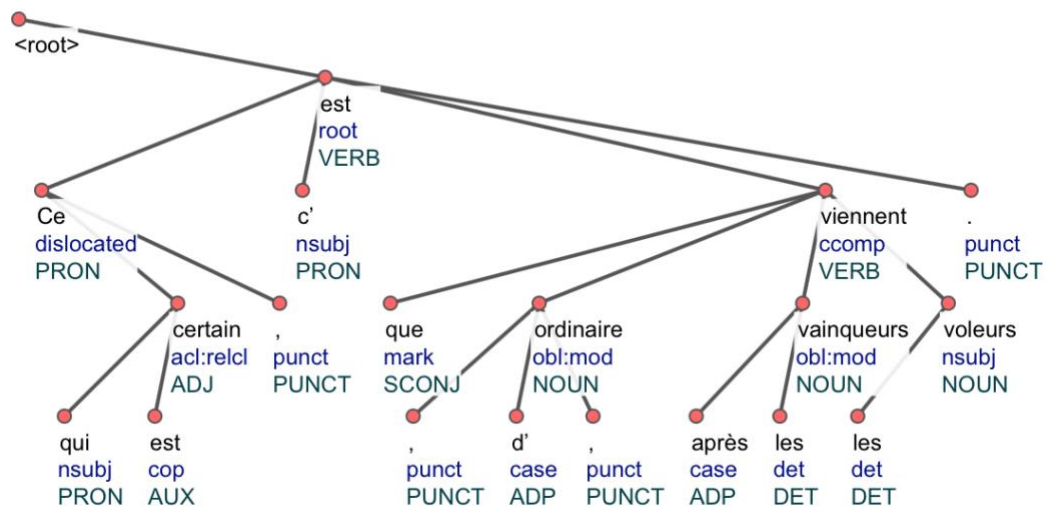


Figure 3.6 Arbre syntaxique de dépendance pour l'exemple 7.

Le sujet du verbe a l'étiquette de « nsubj » (le « nsubj » marque les sujets nominaux d'une clause), ou dans le cas du verbe au passif, « nsubj : pass » (sujet d'une phrase au passif) :

8. *De cette cave sort Lacenaire.* (Hugo)
9. *Au bas de cette plaque était percée une ouverture tout à fait pareille à la bouche d'une boîte aux lettres.* (Flaubert)

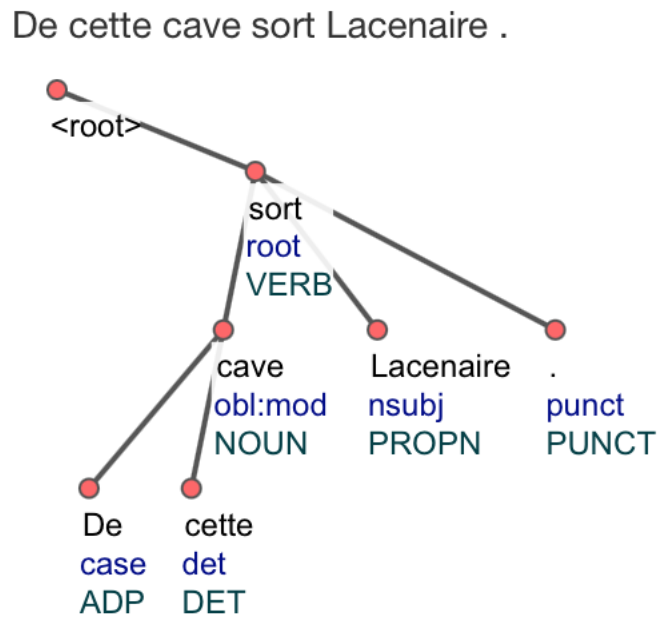


Figure 3.7 Arbre syntaxique de dépendance pour l'exemple 8.

À le bas de cette plaque était percée une ouverture tout à fait pareille à la bouche d' une boîte à les lettres .

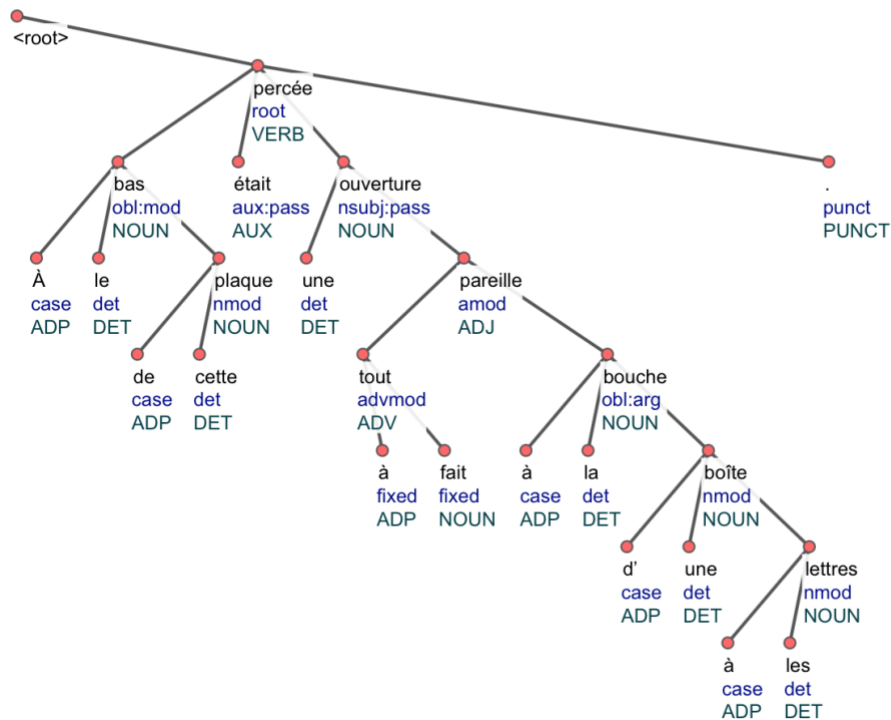


Figure 3.8 Arbre syntaxique de dépendance pour l'exemple 9.

La première étape consiste à parcourir les relations de dépendance dans la phrase jusqu'à trouver le mot qui porte l'étiquette « nsubj » ou l'étiquette « nsubj : pass » ; il faut ensuite suivre le lien pour arriver au verbe dont dépend ce mot. La dernière étape consiste à vérifier si le sujet est placé après le verbe dans la phrase. Ces trois étapes permettent de trouver toutes les occurrences d'inversion du sujet dans les phrases (Voir l'Annexe 3 pour les explications de l'algorithme d'extraction).

● Nettoyage des résultats et discussion

Les inversions du sujet sont assez fréquentes dans notre corpus. Cependant, toutes les inversions du sujet ne sont pas stylistiques, la plupart de nos résultats sont essentiellement les inversions grammaticales (cf. ci-dessous pour une typologie plus précise). Aussi, des nettoyages, soit automatiques, soit manuels ont dû être faits sur les résultats. Nous avons supprimé les inversions suivantes en raison de leurs fonctions

non-stylistiques, d'après l'ouvrage de Christine Tellier et Daniel Valois¹⁹² (Tellier et al. 2006) :

1. **Les questions avec inversion du sujet** : il existe plusieurs manières de poser une question, mais le choix d'inverser le sujet semble neutre et n'est donc pas stylistiquement marqué. Dans notre approche, la suppression des questions est réalisée principalement de manière automatique : nous supprimons toutes les phrases qui se terminent par un point d'interrogation dans les résultats. Notons toutefois que toutes les phrases interrogatives ne se terminent pas avec un point d'interrogation, par exemple :

*10. **Pouvait-on** croire qu'une personne aussi... charmante n'eût pas d'ami sérieux ! (Flaubert)*

Il existe aussi des phrases interrogatives qui se terminent par un point ou un point-virgule, celles-ci sont supprimées manuellement.

2. **Les incises** : les incises sont les inversions non-stylistiques les plus fréquentes dans les résultats. Nous considérons que les incises sont peu indicatives du style en raison de leur caractère contraint par le contexte. Pour supprimer les incises, nous avons élaboré une liste des verbes couramment utilisés dans les incises et avons supprimé toutes les inversions dont le verbe se trouve dans cette liste (la suppression se fait automatiquement par un script écrit en Python). Cette liste est incluse dans l'Annexe 2.
3. **Les subordinations relatives** : cette catégorie regroupe également une grande partie des phrases de nos résultats. Les inversions du sujet sont courantes dans les propositions relatives dont le sujet est plus long que le verbe :

¹⁹² Tellier, Christine, et Daniel Valois. 2018. *Op. cit.*

*11. Le Génie des airs secouait sa chevelure bleue, embaumée de la
senteur des pins, et l'on respirait la faible odeur d'ambre
qu'exhalaient les crocodiles couchés sous les tamarins des fleuves.*
(Chateaubriand)

Cette catégorie d'inversion est considérée comme stylistique, cependant, nous pensons que l'usage de cette figure est plus ou moins figé en français moderne. La suppression de ces inversions n'est malheureusement pas automatisable parce que le pronom relatif (« que », « dont », « où », etc.) est un indice ambigu. L'analyse automatique des subordinées relatives n'est pas suffisante pour nos besoins à l'heure actuelle. Cette partie du nettoyage a donc dû être faite manuellement.

4. **Les autres inversions grammaticales** : à part les subordinations relatives, d'autres occurrences d'inversions du sujet ont aussi été constatées dans des usages plus ou moins figés dans la grammaire. Par exemple :

*12. Le danger n'est ni à Londres, comme le croit Robespierre, ni à Berlin,
comme le croit Danton ;* (Hugo)

« Comme » n'est pas un adverbe qui exige l'inversion du sujet, mais cet usage est malgré tout relativement figé en français. Ces cas, encore plus difficiles à modéliser que les subordinations relatives, sont aussi supprimés manuellement.

5. **Les adverbes qui exigent les inversions obligatoires ou facultatives** : les adverbes comme « encore », « à peine » ou « aussi » exigent l'inversion du sujet quand ils sont placés à la tête d'une proposition. Nous avons utilisé la liste d'adverbes proposée par le Bureau de la Traduction des Travaux publics et

Services gouvernementaux de Canada ¹⁹³ pour détecter les occurrences d'inversions dans nos résultats. Néanmoins, ces adverbes peuvent aussi apparaître dans les phrases avec inversion du sujet, sans que ce soit l'adverbe qui en soit la cause. L'analyse ne peut donc être complètement automatisée et une vérification manuelle est exigée dans ce cas.

Voici les cinq catégories d'inversion que nous avons supprimées dans nos résultats. La plupart de ces cas sont purement grammaticaux (questions, adverbes, incises, etc.), la suppression automatique de ces occurrences est souvent possible. Cependant, nous avons montré aussi qu'il existe des catégories d'inversions traditionnellement considérées comme stylistiques, mais devenues plus ou moins grammaticales dans l'utilisation courante. Pour ces catégories, la suppression est souvent manuelle.

Dans cette partie, nous avons présenté les étapes nécessaires pour identifier les inversions du sujet, puis notre algorithme permettant d'extraire toutes les inversions du sujet. Nous avons ensuite discuté de cinq grandes catégories d'inversions du sujet qui ne sont pas considérées comme stylistiques. Le nettoyage des résultats n'est pas tout à fait automatique : la vérification manuelle est toujours demandée pour certaines catégories d'inversion dont la modélisation est encore difficile en raison des limites techniques. Notre but, ici, est de dévoiler une partie des caractéristiques stylistiques d'un texte par l'étude des figures de style d'ordre syntaxique. L'étude de la typologie des régularités d'occurrences de ces figures, ainsi que leur analyse, seront présentées dans les deux chapitres suivants.

¹⁹³

https://www.btb.termiumplus.gc.ca/clefsfp-srch?lang=fra&srchtxt=possible&cur=69&nmbr=111&lettr=indx_catlog_i&page=9opkVN9D2Afc.html, consulté le 4 mai 2021.

3.6.2 L'identification et l'extraction des phrases clivées

La mise en relief par un présentatif, ou le clivage, est un phénomène stylistique qui consiste à « accorder à un certain constituant de la phrase plus d'importance qu'aux autres¹⁹⁴¹⁹⁵ » (Dessaintes, 1960) (Müller-Hauser, 1943). Les procédés pour mettre en relief un ou plusieurs constituants dans une phrase sont très divers, ils peuvent avoir lieu à tous les niveaux linguistiques : lexical, syntaxique ou sémantique. La thèse ne porte que sur les phénomènes liés à la syntaxe, aussi n'exploite-t-elle que des mises en relief au niveau syntaxique. Nous présentons dans cette partie un algorithme qui détecte et extrait automatiquement les clivages. Nous présentons ensuite une discussion autour de ce sujet et enfin, le nettoyage des résultats qui n'est malheureusement pas complètement automatisé.

Au niveau syntaxique, il existe deux figures de style majeures pour mettre en relief une partie de la phrase : le détachement et le clivage :

13. *Paris, je ne saurais vivre ailleurs.* (détachement)

14. *Si vous êtes dans un désert, c'est mon ami qui en fait le charme.*

(Chateaubriand) (clivage)

L'exemple 13 montre une mise en relief par détachement : « Paris », l'objet indirect de cette phrase a été détaché de la proposition principale pour être déplacé en tête de la phrase, suivi d'une virgule. Nous avons rencontré une grande difficulté à détecter cette structure syntaxique dans les résultats d'analyse syntaxique en dépendances parce que l'analyseur a du mal à définir la fonction de la partie détachée et déplacée (et fournit donc des analyses erronées) : cette partie est souvent considérée

¹⁹⁴ Dessaintes, Maurice. 1960. *Éléments de linguistique descriptive en fonction de l'enseignement du français*. Namur : La Procure.

¹⁹⁵ Müller-Hauser, M.-L. 1943. *La mise en relief d'une idée en français moderne*. Genève.

comme une proposition indépendante dont les relations syntaxiques sont fautives et très diverses dans les résultats. En raison des limites techniques, nous n'avons pas trouvé une bonne méthode capable de bien modéliser et d'extraire cette structure de détachement syntaxique.

Le deuxième exemple montre une mise en relief sous la forme d'un clivage. Cette construction syntaxique consiste à détacher un terme de la phrase en l'insérant dans l'expression « c'est... qui/que », « c'est » introduit le terme extrait, « qui/que » le reste de la phrase. « Qui » est réservé à la fonction sujet, « que », à toutes les autres fonctions. Bien que notre parseur UDPipe ne reconnaisse pas toujours cette construction (l'élément entre « c'est... qui/que » est souvent étiqueté fautivement à cause d'une autre possibilité d'interprétation de cette même structure que nous allons expliquer dans les paragraphes suivants), il est possible d'extraire cette structure formellement, sans passer par UDPipe.

● **Identification des clivages**

L'identification des clivages est relativement facile. Pour éliminer la variation autour du verbe « être » (temps, personne, mode...), nous remplaçons d'abord chaque mot par son lemme, nous détectons ensuite l'existence d'une des quatre structures suivantes : « ce être », « ce ne être », « être ce » ou « ne être ce » suivi par « que » ou « qui » dans la phrase.

Cette simplicité de l'identification pourrait provoquer une ambiguïté, car il est possible que le « que » ou « qui » qui apparaît après le présentatif ne soit pas forcément lié à l'autre partie de ce clivage, il pourrait s'agir simplement d'un pronom relatif d'une autre proposition de la phrase. Par exemple :

15. *Ce sont de jeunes filles fraîches et jolies **qui** présentent aux coups de ces marteaux énormes les petits morceaux de fer **qui** sont rapidement transformés en clous.* (Stendhal)

Le deuxième « qui » ne renvoie évidemment pas à « ce », mais « aux petits morceaux de fer ». Ou bien on peut aussi avoir affaire à une simple comparaison liée par « que » :

16. *Ce n'est donc pas la même, dit Gisquette, **que** celle **qu'**on a donnée il y a deux ans, le jour de l'entrée de monsieur le légat, et où il y avait trois belles filles faisant personnages...* (Hugo)

Ce manque de précision est logique, d'une part parce que nous avons choisi de considérer les deux parties de cette construction (« ce être » et « qui/que ») indépendamment, d'autre part parce que l'extraction a été réalisée au niveau linguistique superficiel sans tenir compte des étiquettes de relation syntaxique, qui, comme nous avons déjà évoqué, ne sont pas toujours correctes. Les explications de l'algorithme d'extraction se trouvent dans l'Annexe 3.

- **Nettoyage des résultats et discussions.**

Nous présentons ici la plus grande difficulté dans le repérage du clivage. Dans la langue française, la séquence « c'est X qui/que V » possède deux séquences morphologiquement identiques : soit une proposition clivée (exemple 17), soit une proposition relative (phrase 18) :

17. *Tu sais que **c'est** la nuit **que** je te délivre.* (Chateaubriand)

18. *Ce fut une grande affaire, **qui** lui coûta cher.* (Stendhal)

Il faut donc dresser les caractéristiques inhérentes à chacune de ces séquences, afin de pouvoir discriminer des cas, qui pourraient présenter une certaine difficulté d'analyse. Sophie-Anne Brault-Scappini a discuté de cette difficulté d'analyse dans sa thèse ¹⁹⁶ (Brault-Scappini, 2007). Si les deux phrases ci-dessus ne causent pas d'ambiguïté, la phrase suivante nous fait effectivement hésiter :

19. C'est une femme, se disait-il, qui me rend toutes les folies de la jeunesse !

(Stendhal)

Ici, est-ce que « une femme » est l'attribut du « ce » ou occupe-t-elle la fonction du sujet dans la phrase subordonnée ? Dans le premier cas, nous devons interpréter cette phrase de cette manière : *c'est une femme (mais pas un homme) qui me rend toutes les folies de la jeunesse*. L'accent ici est mis sur « une femme », alors que dans le deuxième cas, nous devrions l'interpréter comme : *c'est une femme, elle me rend toutes les folies de la jeunesse, [elle est XXX (elle s'appelle XXX)]*. Dans ce cas-là, ce n'est pas « une femme » qui est mise en relief, le vrai accent de cette phrase est dans la proposition suivante « qui me rend toutes les folies de la jeunesse » qui décrit les qualités de cette femme.

Rappelons-nous que dans la langue française, un élément ne peut pas avoir deux fonctions différentes. Ainsi, dans notre exemple, « une femme » devrait être soit l'attribut du pronom démonstratif « ce », soit le sujet de la proposition suivante. Pourtant, aucun marqueur formel ne peut nous aider à prédire la fonction de « une femme ». Il existe des études qui cherchent à modéliser et à différencier les deux

¹⁹⁶ Brault-Scappini, Sophie-Anne. 2007. « Étude du dispositif d'extraction en « c'est...qu », différenciation entre une relative en « c'est...qu » et une proposition clivée ». *L'information grammaticale* 114 (1): 53-56.

séquences morphologiquement identiques¹⁹⁷ (Brault-Scappini, 2007), mais la méthode la plus fiable est de remettre cette phrase dans son contexte.

En raison de cette ambiguïté sémantique s'ancrant dans la syntaxe, le travail de vérification manuelle et de nettoyage devient primordial. La phrase elle-même n'est souvent pas suffisante, relocaliser cette phrase dans le contexte est la seule manière fiable pour déterminer la fonction de « c'est... qui/que ».

Nous avons présenté dans cette partie une méthode pour identifier et extraire les clivages. La simplicité d'extraction produit des résultats qui demandent un travail manuel de vérification intense. En comparaison avec la méthode d'extraction des inversions du sujet (dans la partie précédente), l'extraction des clivages montre que plus nous utilisons les traits linguistiques superficiels, moins les résultats sont précis. L'utilisation des séquences de lemmes (un paramètre superficiel) produit un effet négatif sur la précision des résultats ; le fait qu'une proposition clivée et une subordination relative partagent la même construction syntaxique complique encore l'identification et l'extraction de ce patron caractérisant. Même si l'extraction des clivages n'est pas tout à fait automatisée, les chapitres qui suivent montrent que cette figure de style est souvent utile pour révéler les caractéristiques stylistiques des textes de notre corpus.

¹⁹⁷ *Idem.*

3.6.3 L'identification et l'extraction des répétitions

La détection des répétitions syntaxiques est possible grâce à l'analyse des séquences d'étiquettes morphosyntaxiques. La méthode la plus simple pour repérer les répétitions syntaxiques consiste en effet à parcourir les séquences d'étiquettes morphosyntaxiques dans les phrases pour vérifier s'il existe des séquences répétées. Nous allons montrer le manque de précision de cette méthode linéaire qui ne prend pas en compte les délimitations syntagmatiques, qui sont pourtant primordiales pour l'interprétation sémantique des structures répétées (voir la section suivante pour les exemples). Nous présentons ensuite une nouvelle méthode, basée sur l'analyse syntaxique en dépendances, pour extraire cette figure de style d'une manière plus précise.

3.6.3.1 La méthode linéaire

La méthode linéaire repère les séquences d'étiquettes morphosyntaxiques répétitives dans une phrase. La première étape de cette méthode consiste à associer à chaque mot une étiquette morphosyntaxique :

20. J'entraînai la fille de Simaghan aux pieds des coteaux qui formaient des golfes de verdure, en avançant leurs promontoires dans la savane. (Chateaubriand)

PRON VB DT NN ADP NP ADP DT NN ADP DT NN PRON VB ADP DT NN
ADP NN PUNCT ADP VB DT NN ADP DT NN PUNCT

Ensuite, il faut définir la taille des séquences prises en compte pour le repérage des répétitions, avant de lancer l'extraction. Prenons l'exemple des séquences avec la

taille 4. Dans notre cas, nous obtenons un seul patron répétitif conformant à nos critères, « VB DT NN ADP » :

PRON **VB DT NN ADP** NP ADP DT NN ADP DT NN PRON **VB ADP DT NN**
ADP NN PUNCT ADP VB DT NN ADP DT NN PUNCT
VB DT NN ADP

Les deux fragments de textes derrière cette structure morphosyntaxique sont : « entraînai la fille de » et « formaient des golfes de ». Ils ne sont pas sémantiquement interprétables à cause de la délimitation syntagmatique problématique (c'est-à-dire que les séquences repérées ne correspondent pas à des syntagmes pleins). Ce manque de précision s'explique par le fait que la taille des séquences morphosyntaxiques est prédéfinie et que cette prédéfinie de taille ne prend pas en compte la délimitation syntagmatique.

Pour la première raison, il serait difficile de trouver une taille parfaite pour détecter les répétitions, par exemple les 4-grams ne fonctionnent pas dans l'exemple précédent, mais ils peuvent bien détecter les répétitions sémantiquement interprétables dans l'exemple suivant :

*21. Suspendu sur les cours des eaux, groupés sur les rochers et sur les montagnes, dispersés dans les vallées, des arbres **de toutes les formes, de toutes les couleurs, de tous les parfums**, se mêlent, croissent ensemble, montent dans les airs à des hauteurs qui fatiguent les regards. (Chateaubriand)*

Pour la deuxième raison, la délimitation syntagmatique n'est possible qu'avec des analyses syntaxiques plus profondes. Nous avons montré avec les méthodes d'extraction des inversions du sujet et des clivages qu'avec un niveau d'extraction superficiel, nous obtenons généralement les résultats imprécis. C'est pourquoi nous

proposons dans la partie suivante une nouvelle méthode visant à améliorer la précision de repérage de cette figure de style.

3.6.3.2 La méthode d'extraction syntaxique

- **Identification des répétitions**

Notre nouvelle méthode d'extraction se base sur l'analyse syntaxique en dépendances. Les structures répétitives possèdent non seulement des séquences d'étiquettes syntaxiques identiques, mais elles ont aussi des relations syntaxiques identiques. Prenons l'exemple de la phrase suivante de Victor Hugo :

22. *Vous pouvez raisonner un dogue, étonner un taureau, fasciner un boa, effrayer un tigre, attendre un lion.*

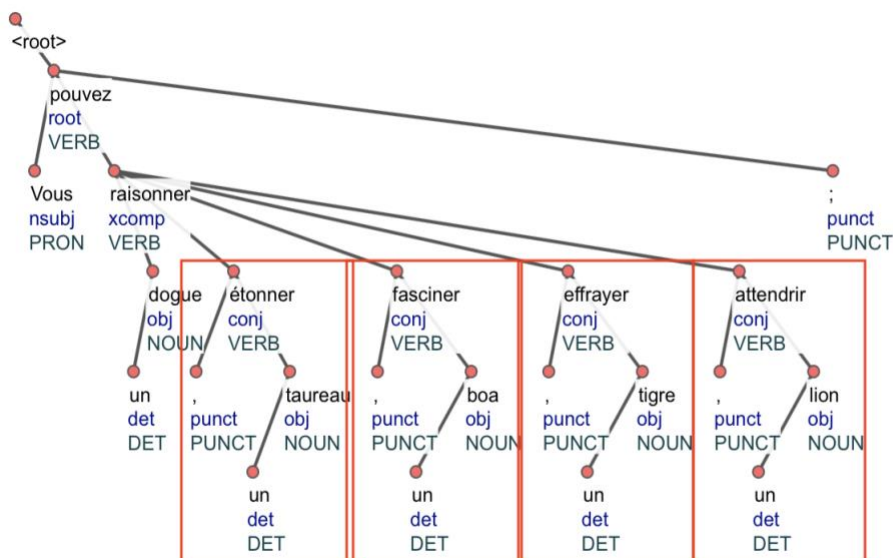


Figure 3.9 Les patrons en parallélisme dans un arbre syntaxique de dépendance.

Nous pouvons constater que les quatre structures parallèles sont à la fois identiques morphosyntaxiquement (« PUNCT VB DT NN ») et syntaxiquement. Sur l'arbre de dépendance syntaxique, nous voyons quatre sous-branches qui se répètent sans altération. Ainsi, l'identification des répétitions syntaxiques devrait prendre en compte non seulement les séquences morphosyntaxiques, mais aussi les relations syntaxiques.

Les avantages de notre méthode se manifestent de trois façons :

1. Nous pouvons éviter le problème d'avoir à définir a priori une taille idéale pour détecter les répétitions syntaxiques. Dans notre méthode, au lieu de prédéfinir la taille des séquences morphosyntaxiques, nous définissons la profondeur des structures syntaxiques à extraire, c'est-à-dire la profondeur des relations syntaxiques de dépendance verticales. Par exemple, les quatre structures syntaxiques soulignées dans la figure ci-dessus ont chacune pour profondeur 3, parce que les sous-branches de l'arbre syntaxique s'étendent verticalement sur trois niveaux. La structure syntaxique peut avoir autant d'éléments horizontaux que nécessaire pour garder la complétude syntagmatique. Prédéfinir la profondeur et non pas la taille des séquences linéaires nous permet d'extraire plus de patrons sémantiquement interprétables, parce que le choix de la profondeur d'extraction est moins libre que le choix de la taille des séquences morphosyntaxiques (la profondeur syntaxique d'une phrase se situe en général entre 2 et 6, alors que la taille des séquences morphosyntaxiques peut varier d'entre 2 et la longueur totale de la phrase. Nous verrons par ailleurs que seules deux profondeurs sont idéales pour reconnaître les répétitions syntaxiques dans notre corpus (voir la section suivante pour une discussion sur le choix des profondeurs d'extraction, en 3.7.2).
2. Le nettoyage des résultats est plus simple. Il suffit en pratique de supprimer les patrons qui ne sont pas contigus : tous les autres patrons sont sémantiquement interprétables.

3. Nos résultats ont une plus grande précision que ceux obtenus avec la méthode linéaire. Selon un test comparatif, au lieu de 10,306 occurrences de 4-grams répétitifs dans les romans de Chateaubriand, nous avons 342 occurrences de structures syntaxiques répétitives avec la profondeur d'extraction 3.

- **Les étapes de l'algorithme d'extraction**

1. Analyser toutes les phrases de notre corpus avec UDPipe.
2. Extraire tous les patrons syntaxiques hiérarchiques de profondeur 3 et 4 de manière non-supervisée. Nous avons décidé de n'extraire que les patrons de profondeurs 3 et 4, parce que la profondeur 2 donne des patrons trop courts qui ont rarement une signification stylistique (par exemple, DT NN, les profondeurs d'extractions sont discutées en détail dans la section 3.7.2). Nous n'avons pas utilisé les profondeurs 5 et plus non plus, parce que les structures hiérarchiques de ces tailles ne se répètent presque jamais dans notre corpus : elles ne sont donc pas statistiquement significatives.
3. Ensuite, nous éliminons les structures syntaxiques non-contiguës (les structures avec un ou plusieurs éléments manquants (pour des syntagmes ayant une profondeur complète supérieure à 3 ou 4). Remarquons que les structures syntaxiques supprimées à cause de discontinuités « réapparaissent » quand nous élargissons la profondeur d'extraction : ainsi, le fait d'utiliser plusieurs profondeurs d'extraction apporte de meilleurs résultats (voir la section 3.7.2 pour les exemples).
4. Enfin, nous utilisons un script en Python pour parcourir chaque phrase et sa structure syntaxique. Le script cherche, parmi toutes les structures syntaxiques d'une phrase, celles qui se répètent au moins deux fois.

Les explications de l'algorithme d'extraction se trouvent en Annexe 3.

● Nettoyage des résultats et discussion

Le nettoyage des résultats est relativement plus facile que celui des clivages et des inversions parce qu'il n'existe pas d'ambiguïté sémantique (comme pour les clivages) ni d'usages figés non-stylistiques (comme pour les inversions). Le travail de nettoyage consiste simplement à combiner les résultats et à calculer les fréquences globales au niveau du corpus.

Nous voyons que l'analyse syntaxique apporte une meilleure précision à l'extraction des structures syntaxiques répétitives. L'extraction des trois figures de style fondées sur la syntaxe nous permet d'avoir une première idée des tendances de chaque auteur de notre corpus à employer ces outils stylistiques dans leurs créations littéraires. Nous allons étudier, d'une manière plus profonde, les typologies et les régularités d'occurrences de ces trois figures de style syntaxiques dans les Chapitres 4 et 5. Certes, ces trois patrons caractérisants ne sont encore ni suffisants, ni assez généraux pour caractériser un style littéraire à eux seuls, mais nous voulons ici mettre en relief le fait que l'exploitation des relations syntaxiques peut nous aider à modéliser et à extraire certaines figures de style fondées sur la syntaxe. Nous complétons ce manque de généralité par l'introduction, dans la partie suivante, de la notion de patron syntaxique identificateur.

3.7 L'extraction des patrons identificatoires

Dans la partie précédente, nous avons présenté des méthodes d'extraction supervisée pour identifier et extraire trois patrons caractérisants (les inversions du sujet, les clivages et les répétitions). Ces trois figures de style, fondées sur la syntaxe, peuvent dévoiler une partie des caractéristiques stylistiques des textes. Néanmoins, l'identification supervisée des figures de style est largement limitée par la possibilité de modélisation (certaines figures ne sont pas repérables en se fondant uniquement sur la syntaxe) ainsi que par la performance des analyseurs syntaxiques. En raison de ces limites, notre étude sur les figures de style ne peut pas donner une vue générale des caractéristiques syntaxiques et stylistiques des textes de notre corpus. C'est pour cette raison que nous avons argumenté (dans le Chapitre 2, mais aussi dans la partie 3.4), sur la nécessité d'adopter une méthodologie « opportuniste », qui prenne en compte à la fois les patrons caractérisants et les patrons identificatoires. Les premiers nous aident à comprendre la préférence de chaque auteur pour certains outils stylistiques et syntaxiques, alors que les deuxièmes révèlent les caractéristiques syntaxiques spécifiques qui permettent de bien distinguer un texte des autres. Les deux types de patrons nous permettent de pratiquer les deux variétés de stylistiques dont parlait Stanley Fish (voir 2.2.6 du Chapitre 2).

Les patrons identificatoires désignent les patrons syntaxiques qui n'ont pas toujours une signification stylistique, mais possèdent « la fonction identificatoire¹⁹⁸ » (Rastier, 2001) permettant de différencier les œuvres. Par conséquent, il n'est pas nécessaire de prédéfinir les caractéristiques de ces traits avant de les repérer dans les textes. Les extractions non-supervisées pourraient sortir tous ces patrons sans discrimination avant d'étudier leur fréquence grâce à des calculs de spécificité (les patrons syntaxiques statistiquement les plus spécifiques d'après leur fréquence dans le

¹⁹⁸ Rastier, F. 2001. *Arts et sciences du texte*. Paris cedex 14, France: Presses Universitaires de France.

texte seraient alors considérés comme identificatoires). Cependant, bien que les patrons identificatoires n'aient pas toujours de significations stylistiques, ils doivent être linguistiquement corrects (c'est-à-dire correspondre à un syntagme plein). Un fragment syntaxique qui ne respecte pas la délimitation syntagmatique n'est pas interprétable et par conséquent, ne peut être relié à aucune fonction stylistique. Nous présentons dans cette partie une méthode pour extraire ces patrons identificatoires d'une manière non-supervisée, en respectant les délimitations syntagmatiques grâce à l'analyse syntaxique en dépendances. Nous commençons par présenter la méthode d'extraction linéaire de Legallois et al.¹⁹⁹ (2016), nous discutons ensuite de son mécanisme et de ses limites. Ensuite, nous présentons le mécanisme et l'algorithme de notre méthode qui pourrait être considérée comme une amélioration de la méthode linéaire. Les discussions des erreurs et les étapes d'extraction seront également présentées dans cette partie. Enfin, nous faisons une comparaison de notre méthode avec la méthode linéaire pour évaluer notre méthode par rapport à l'autre.

3.7.1 La méthode d'extraction linéaire des patrons identificatoires

La méthode d'extraction linéaire, présentée dans Legallois et al. (2016) propose d'extraire des séquences d'unités morphosyntaxiques dans une phrase. Dans l'article Legallois et al.²⁰⁰ (2018), les auteurs proposent un mélange d'étiquettes en combinant les lemmes pour les mots fonctionnels et les étiquettes morphosyntaxiques pour les autres mots (un étiquetage sémantique plus avancé est aussi évoqué. Pour la facilité la présentation du mécanisme utilisé, nous avons choisi de ne pas tenir compte

¹⁹⁹ Legallois, Dominique, Thierry Charnois, et Thierry Poibeau. 2016. « Repérer les clichés dans les romans sentimentaux grâce à la méthode des « motifs » ». *Lidil. Revue de linguistique et de didactique des langues*, n° 53 (mai): 95-117.

²⁰⁰ Legallois, Dominique, Thierry Charnois, et Meri Larjavaara, éd. 2018. *The grammar of genres and styles: from discrete to non-discrete units*. Berlin, Allemagne: De Gruyter Mouton.

d'éventuelles étiquettes sémantiques sur les mots dans nos exemples). Le but de cet étiquetage est de rendre les patrons plus interprétables. Ainsi, pour la phrase suivante :

23. Dans les îles de la Manche, un hiver où il gèle à glace est mémorable, et la neige fait évènement. (Hugo)

on obtient :

dans le NN de le NP PUNCT un NN où il VB à NN être ADJ PUNCT et le NN faire NN PUNCT

Une fenêtre glissante d'une taille prédéfinie (par exemple 4) repère toutes les séquences de 4 unités :

dans le NN de
le NN du
NN du NP
Du NP PUNCT
...
NN faire NN PUNCT

La fréquence de tous les patrons pour un texte donné est utilisée dans les calculs statistiques de spécificité, afin de repérer les plus spécifiques sur le plan statistique.

Nous remarquons dans cet exemple deux problèmes de cette méthode linéaire :

1. La difficulté d'interprétation : un grand nombre de patrons n'ont pas de signification, ni grammaticale ni syntaxique, par exemple : « le NN de le ». La méthode ne prend pas en compte la délimitation syntagmatique.
2. Le problème de la taille idéale des patrons : il est difficile de savoir et de justifier quelle est la taille idéale pour extraire les patrons. Quelle taille de structures

syntaxiques est la plus caractéristique du style ? Si aucune taille n'est prédéfinie, le nombre de patrons explose et l'analyse est alors très difficile.

Pour résoudre ces deux problèmes, notre méthode propose d'explorer les structures syntaxiques hiérarchiques derrière la représentation linéaire d'une phrase. La possibilité de cette exploration a été discutée dans la section 3.2. La grammaire de dépendance est une des méthodes qui permet de mettre en évidence la structure syntaxique hiérarchique d'une phrase reliée par les chaînes de relations syntaxiques. Notre méthode d'extraction hiérarchique se base sur les analyses syntaxiques en dépendances.

3.7.2 La méthode d'extraction hiérarchique des patrons syntaxiques

Nous présentons, dans cette partie, la méthode permettant d'identifier et d'extraire les patrons syntaxiques hiérarchiques. Les étapes d'extraction, l'explication de l'algorithme, le choix du paramètre d'extraction, la comparaison des résultats avec la méthode linéaire et l'analyse des erreurs d'extraction sont présentés de manière détaillée.

● Le mécanisme d'extraction des patrons syntaxiques

Les phrases ont toutes une structure syntaxique hiérarchique, qui ne correspond que de manière très imparfaite à leur structure linéaire. Les outils d'analyse syntaxique permettent de révéler les relations syntaxiques derrière la structure linéaire des phrases : celles-ci peuvent être visualisées sous forme de structures arborescentes²⁰¹. Par conséquent, pour extraire les patrons syntaxiques, prendre en considération les relations

²⁰¹ Matthews, Peter H. 2007. *Syntactic Relations: A Critical Survey*. Cambridge: Cambridge Univ. Press.

syntaxiques entre les mots donne souvent des résultats plus précis (l'extraction des répétitions dans la section 3.6.3 en est un bon exemple). L'essentiel de notre méthode consiste à extraire les patrons syntaxiques dont les composants sont liés par des relations syntaxiques plutôt que par l'ordre séquentiel de la phrase. Notre méthode est proche de celle de Sidorov et al. (2014) dans laquelle les auteurs nomment ces patrons syntaxiques les « sn -grams²⁰² ».

Pour expliquer notre méthode, prenons l'exemple de la même phrase de Victor Hugo que dans l'exemple précédent : « Dans les îles de la Manche, un hiver où il gèle à glace est mémorable, et la neige fait événement ». La visualisation des relations syntaxique correspond à :

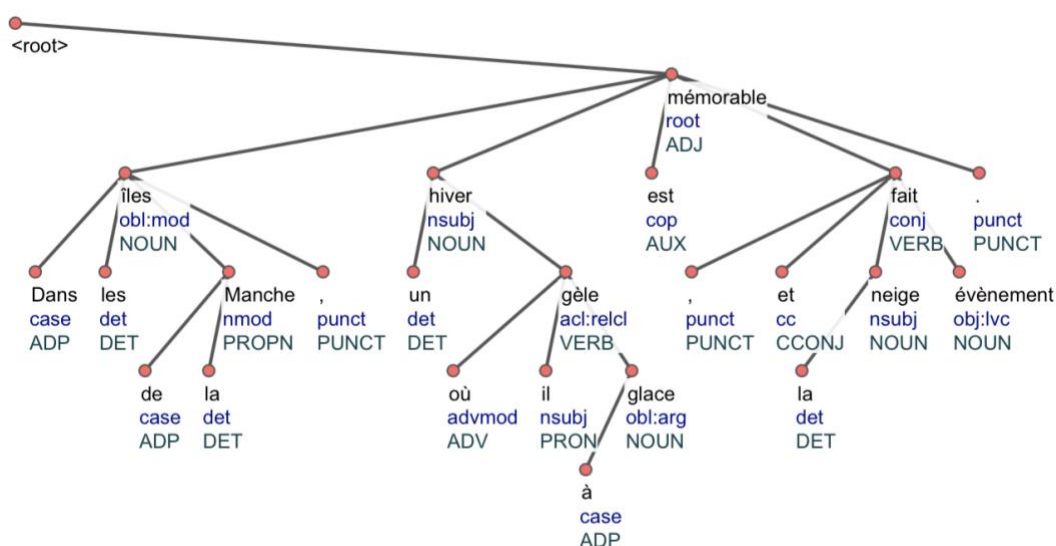


Figure 3.10 Arbre syntaxique de dépendance pour la phrase de Victor Hugo.

Au lieu d'extraire les séquences de mots d'une longueur prédéfinie, notre méthode propose d'extraire les structures syntaxiques hiérarchiques d'une certaine profondeur

²⁰² Sidorov, Grigori, Francisco Velasquez, Efstathios Stamatatos, Alexander Gelbukh, et Liliana Chanona-Hernández. 2014. « Syntactic N-Grams as Machine Learning Features for Natural Language Processing ». *Expert Systems with Applications, Methods and Applications of Artificial and Computational Intelligence*, 41 (3): 853-60.

(c'est-à-dire l'étendue verticale du patron sur l'arbre syntaxique ; le choix de la profondeur d'extraction est discuté dans « **Les choix du paramètre d'extraction** », à la section 3.7.2). L'avantage de cette méthode est qu'elle réduit considérablement le nombre de patrons dans les résultats par rapport aux méthodes ne mettant pas en jeu la syntaxe, et que les patrons sont en principe plus fréquemment interprétables vu qu'ils respectent des relations syntaxiques précises.

Dans notre exemple, si nous prédéfinissons la profondeur d'extraction à 3, nous obtenons quatre patrons, soulignés dans le graphique suivant (au lieu de 19 si on procède à l'extraction linéaire des 4-grams dans la phrase) :

Dans les îles de la Manche , un hiver où il gèle à glace est mémorable , et la neige fait évènement .

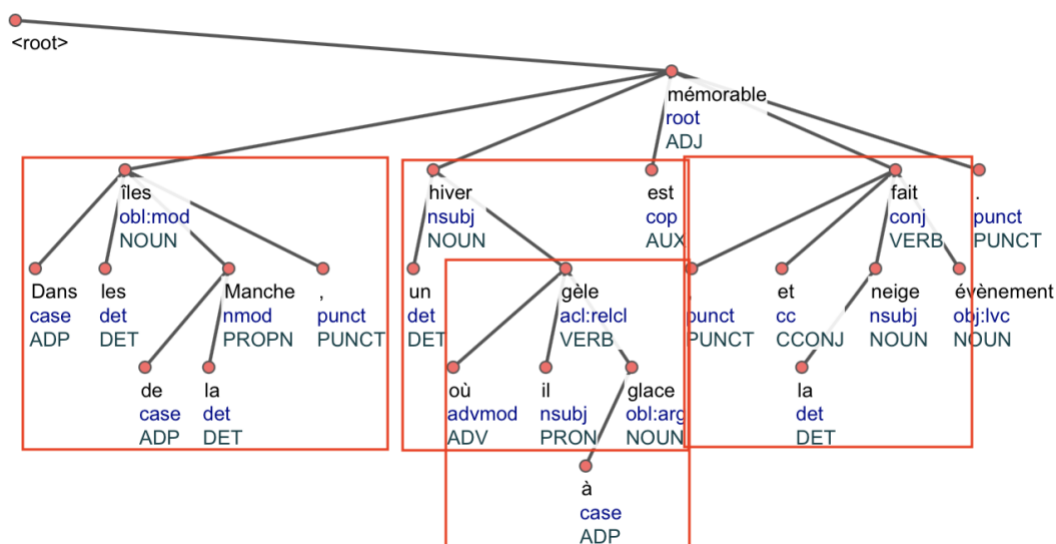


Figure 3.11 Les quatre patrons syntaxiques avec la profondeur d'extraction 3.

Dans les îles de la Manche,

un hiver où il gèle glace

où il gèle à glace

, et la neige fait évènement

Notons que l'extraction des patrons syntaxiques identificatoires ne prend pas en compte l'élément « root » de la phrase, parce que les patrons qui contiennent « root » sont soit des phrases complètes (pour les phrases dont la construction syntaxique est simple), soit des patrons non-contigus (en raison de la profondeur d'extraction insuffisante pour extraire toute la phrase). Bref, les patrons syntaxiques qui comprennent l'élément « root » ne peuvent produire des patrons complets qu'à condition d'extraire toute la phrase comme patron, ce qui n'est pas idéal pour notre recherche de patrons identificatoires. Nous supprimons ensuite les patrons qui ne sont pas interprétables syntaxiquement en raison de problème de discontinuité, et nous obtenons finalement trois patrons comme résultat (« Dans les îles de la Manche, », « où il gèle à glace » et « , et la neige fait évènement »), en supprimant le deuxième patron dont la construction syntaxique n'est pas complète du fait d'un mot manquant. Nous réécrivons ensuite ces patrons avec la méthode d'étiquetage décrite dans l'article de Legallois et al. (2018) : nous gardons les mots fonctionnels sous la forme de lemmes et attribuons aux autres mots leur étiquette morphosyntaxique. Ainsi, nous obtenons les trois patrons syntaxiques suivants :

Dans le NN de le NP PUNCT

où il VB à NN

PUNCT et le NN faire NN

● Les étapes d'extraction

Nous présentons dans cette partie les étapes d'extraction des patrons syntaxiques hiérarchiques selon notre méthode.

Étape 1 :

Nettoyage des textes du corpus (les romans des quatre auteurs du 19^e siècle retenus pour notre étude). Le nettoyage consiste à supprimer les préfaces qui ne sont pas écrites par l'auteur du roman, ainsi que les notes ajoutées par les éditeurs. Les mentions comme « Chapitres X » et les « Livres X » sont également supprimées en raison de leur fréquence élevée, mais peu significative. L'objectif du nettoyage est de préparer les textes bruts pour l'analyse syntaxique en dépendances.

Étape 2 :

Utilisation d'UDpipe pour l'analyse syntaxique en dépendances de chacun des textes dans le corpus. Le modèle d'analyse que nous avons utilisé est « french-gsd-ud-2.5-191206.udpipe ». Pendant la rédaction de la thèse, un nouveau modèle : « french-gsd-ud-2.6.udpipe » a été mis en ligne par les développeurs d'UDPipe. Nous jugeons d'après quelques tests de performance que le nouveau modèle n'apporte pas d'amélioration notable, aussi avons-nous continué avec le modèle 2.5 pour les analyses syntaxiques de la thèse. Les résultats sont produits selon le format du Treebank French-GSD de Universal Dependency²⁰³.

²⁰³ https://universaldependencies.org/treebanks/fr_gsd/, consulté le 4 mai 2021.

Étape 3 :

Préparation des résultats d'analyse syntaxique pour l'extraction. Il s'agit de petits nettoyages qui consistent à réanalyser les articles contractés comme « au », « aux », « des », « du », etc.

UDPipe liste les articles contractés ainsi que leurs formes séparées. Dans l'exemple suivant, il suffit de supprimer la ligne 19 :

```
19 15-16 au ADP - - 17 case - - -  
20 15 à à ADP - - 17 case - - -  
21 16 le le DET - - Definite=Def|Gender=Masc|Number=Sing|PronType=Art 17 det _ -
```

Figure 3.12 Exemple des articles contractés dans les résultats d'UDPipe.

Étape 4 :

Extraction des patrons syntaxiques : Nous utilisons un script écrit en Python pour extraire les patrons syntaxiques qui correspondent aux niveaux hiérarchiques choisis par l'utilisateur. Nous avons emprunté partiellement la méthode de Legallois et al. (2018), en gardant les mots correspondant aux classes lexicales fermées (articles, prépositions, pronoms, etc.) sous forme de lemme, et en remplaçant les autres mots par l'étiquette morphosyntaxique correspondante. Le résultat pour chaque phrase se présente sous la forme suivante : la phrase originale, les patrons identifiés sous leur forme textuelle et sous forme de séquence d'étiquettes d'après les règles de réécriture précisées ci-dessus.

Étape 5 :

Retour au texte : il est très important de pouvoir retourner au texte original pour localiser les patrons dans les phrases. C'est ainsi que nous pouvons ensuite analyser les effets stylistiques des patrons en contexte. Il nous est possible d'obtenir l'affichage des

patrons avec le texte correspondant. Voici un fragment du retour au texte pour le patron « le NN de le NN » :

```
2 les fenêtres de le rez-de-chaussée
3 # text = on ne sait quels hideux emplâtres de planches clouées bouchent les
  fenêtres du rez-de-chaussée ;
4
5 Les anciens de le pays
6 # text = Les anciens du pays racontent, mais ces faits-là appartiennent au
  passé, que la population catholique de l'archipel normand a été autrefois, bien
  malgré elle, plus en communication encore avec le démon que la population
  huguenote.
7
8 les gens de le pays
9 # text = Cette maison a ce qu'elle veut, dirent les gens du pays.
10
11 les naturels de le pays
12 # text = Ils étonnent les naturels du pays.
```

Figure 3.13 Exemple du retour au texte du patron « Le NN de le NN ».

L'algorithme d'extraction ainsi que les explications sont présentés en Annexe 3.

- **Les choix du paramètre d'extraction**

Nous avons dit précédemment que la profondeur d'extraction est prédéfinie à 3 pour les sous-arbres syntaxiques. Dans cette partie, nous justifions ce choix empirique, en examinant les résultats de différents essais sur le roman *Les Travailleurs de la Mer* de Victor Hugo pour déterminer la profondeur d'extraction idéale.

Nous avons d'abord éliminé la profondeur 1 et les profondeurs de 5 et plus. La profondeur 1 nous donne essentiellement les mots au lieu des structures syntaxiques, tandis qu'une profondeur de plus de 5 donne très peu de résultats et semble donc peu pertinente pour une analyse statistique (peu de phrases ont une telle complexité et donc une telle profondeur). Beaucoup de phrases simples n'auraient aucun patron correspondant à de tels critères :


```

# text = Ce mot était son nom.

# text = Il s'appelait Gilliatt.

# text = Il resta longtemps immobile, regardant ce nom, ces petits pieds, cette
neige, puis continua sa route, pensif.

# text = Gilliatt habitait la paroisse de Saint-Sampson.

# text = Il n'y était pas aimé.

# text = Il y avait des raisons pour cela.

# text = D'abord il avait pour logis une maison « visionnée ».

```

Figure 3.14 Exemples des phrases avec la profondeur syntaxique inférieure à 5.

De plus, les profondeurs importantes génèrent des patrons trop longs, qui s'étendent sur plusieurs syntagmes, voire plusieurs propositions. Les patrons trop longs sont difficiles à la fois pour l'interprétation et pour la généralisation des régularités stylistiques d'un auteur :

```

# text = Il arrive quelquefois, à Jersey ou à Guernesey, qu'à la campagne, à la
ville même, passant dans quelque coin désert ou dans une rue pleine d'habitants,
vous rencontrez une maison dont l'entrée est barricadée ;
, à Jersey ou à Guernesey , qu' à la campagne , à la ville même , passant dans
quelque coin désert ou dans une rue pleine d' habitants
, à PROPN ou à PROPN , que à le NOUN , à le NOUN ADJ , VERB dans DET NOUN ADJ ou
dans un NOUN ADJ de NOUN
, vous rencontrez une maison dont l' entrée est barricadée
, vous VERB un NOUN dont le NOUN être VERB

```

Figure 3.15 Exemple des patrons de la profondeur d'extraction 5 pour une phrase de Victor Hugo.

Nous pouvons donc laisser les profondeurs 5 et supérieures de côté.

Nous allons maintenant comparer l'application des profondeurs 2, 3 et 4 sur le roman *Les Travailleurs de la Mer* de Victor Hugo.

Pour la profondeur 2, nous avons 27,539 patrons contigus (sans manque d'éléments dans la construction syntaxique). Nous avons cependant décidé de ne pas garder cette profondeur d'extraction car les patrons obtenus sont peu interprétables dans une perspective d'analyse stylistique. En effet, les patrons obtenus sont relativement courts, ils ne fournissent pas assez d'informations pour caractériser le style d'un auteur. Certes, il nous serait possible de trier encore les patrons pour ne garder que les patrons qui contiennent plus d'un certain nombre de composants ; cependant, la façon la plus économique pour avoir les patrons plus longs est simplement d'élargir la profondeur d'extraction.

et ADJ
de le NN
ce NN
de NN
le NN
de un ADJ NN
de NN
par de le NN ADJ

Tableau 3.4 Exemples des patrons de la profondeur d'extraction 2.

Pour la profondeur 3, nous avons pu extraire 11,686 patrons contigus dans le roman *Les Travailleurs de la Mer*, les patrons ont une longueur idéale, comprise entre

3 à 10 mots. Parmi eux, les patrons les plus fréquents, ceux que nous allons utiliser dans les analyses statistiques, ont une longueur comprise entre 3 et 5 mots.

L'extraction avec la profondeur 4 fournit 5,231 patrons contigus dans ce roman particulier. Les patrons sont logiquement plus longs que ceux de profondeur 3, mais ils ne sont pas très fréquents : par exemple, celui qui est apparu le plus de fois dans ce roman a une fréquence de 13. Ces faibles fréquences limitent leur pouvoir caractérisant.

Patrons	Fréquences
qui VB le NN de le NN	13
le NN de le NN de le NN	9
le NN ADJ et ADJ	8
à le NN de le NN de le NN	8
de NN ADJ et ADJ	7
le NN de le NN de NP	7
VB le NN de le NN	7

Tableau 3.5 Liste des patrons les plus fréquents de la profondeur d'extraction 4.

Après ces tests d'extraction avec les profondeurs différentes, nous pouvons constater que la profondeur 3 est la meilleure solution pour notre méthode. Cela ne veut cependant pas dire que les résultats d'extraction avec les autres profondeurs soient complètement à négliger. Au contraire, nous pouvons constater, par exemple, que la profondeur 4 fournit des patrons plus interprétables : les patrons sont plus proches d'un syntagme complet, ils ont souvent une construction syntaxique plus large que celle des patrons de profondeur 3, aussi, sont-ils plus faciles à interpréter et plus informatifs. Intégrer la profondeur 4 (en plus de la 3) dans notre méthode, malgré la faible fréquence

des patrons trouvés, pourrait bien être une solution utile pour augmenter le nombre de patrons stylistiquement significatifs dans notre résultat.

- **Comparaison des résultats avec la méthode linéaire**

Nous avons évoqué que l'avantage de notre méthode est de pouvoir produire moins de patrons au total et d'augmenter le nombre de patrons interprétables dans les résultats par rapport à la méthode d'extraction linéaire. Dans cette partie, nous allons montrer les avantages et les désavantages de notre méthode en procédant à une comparaison avec la méthode linéaire.

Avec la profondeur 3, nous avons pu extraire 11,686 patrons contigus dans le roman *Les Travailleurs de la Mer* de Victor Hugo. La méthode linéaire, elle, produit plus de 90,000 patrons différents (séquences de 5-grams, c'est-à-dire composées de cinq mots ou ponctuations). L'avantage de notre méthode est donc confirmé : la méthode est plus précise, ce qui facilite les analyses ultérieures, à la fois sur les plans statistiques et stylistiques.

Nous avons enfin calculé les durées de traitement de notre méthode en comparaison avec celles de la méthode linéaire sur deux des romans de Victor Hugo. Nous pouvons constater dans les deux tableaux suivants que notre méthode nécessite beaucoup plus de temps de traitement. Cependant, sur le roman de plus long de notre corpus (*Les Misérables*), la durée de traitement sur un ordinateur personnel est acceptable²⁰⁴.

²⁰⁴ Les résultats des durées de traitement ont été obtenus avec un ordinateur personnel iMac, Intel Core i5, 3.4GHz, 24 GB 2400 MHz DDR4.

	<i>Les Travailleurs de la Mer</i>	<i>Les Misérables</i>
<i>Changement d'étiquettes</i>	0.032518s	0.135056s
<i>Extraction(4-grams)</i>	0.044364s	0.186202s
<i>Total méthode séquentielle</i>	0.076882s	0.321258s
<i>Changement d'étiquettes</i>	11.850256s	45.844821s
<i>Extraction</i>		
<i>Total méthode hiérarchique</i>	11.850256s	45.844821s

Tableau 3.6 Durées de traitement des deux méthodes d'extraction.

Nous avons ensuite fait une vérification manuelle pour compter les patrons syntaxiquement interprétables dans les résultats. À partir du roman *Les Travailleurs de la Mer*, nous avons extrait deux échantillons de 100 patrons (les 100 plus fréquents et 100 autres choisis aléatoirement). Les chiffres de patrons interprétables sont présentés dans le tableau ci-dessous :

		<i>Les travailleurs de la Mer</i>
Méthode linéaire (5-grams)	Sur les 100 plus fréquents	41/100
	Sur les 100 aléatoires	26/100
	Moyen	33.5/100
Notre Méthode hiérarchique (profondeur 3)	Sur les 100 plus fréquents	92/100
	Sur les 100 aléatoires	90/100
	Moyen	91/100

Tableau 3.7 Nombres de patrons sémantiquement interprétables des deux méthodes d'extraction.

Nous obtenons 91 patrons interprétables sur 100 dans nos résultats (il s'agit de patrons syntaxiquement interprétables, qui n'ont pas forcément une fonction stylistique).

- **Examen des erreurs d'extraction**

Nous avons montré que notre méthode produit moins de patrons, mais plus de patrons interprétables que la méthode linéaire. Cependant, d'autres questions se posent à la vue des résultats.

Du fait de notre méthode fondée sur les relations syntaxiques, nous perdons logiquement en chemin une partie des patrons qui ont la même représentation linéaire,

mais dont les relations syntaxiques ne sont pas identiques. Par exemple, le patron le plus fréquent dans le roman *Les Travailleurs de la Mer* : « le NN de le NN » identifie 338 séquences avec notre méthode, mais si nous lançons une recherche linéaire autour de cette structure (donc sans tenir en compte des relations de dépendance), nous avons 1302 occurrences dans le texte. Pour examiner ce phénomène, prenons le résultat de l'analyse d'une des phrases du corpus :

sent_id = 183

patron = le NN de le NN

text = Ils étonnent les naturels du pays.

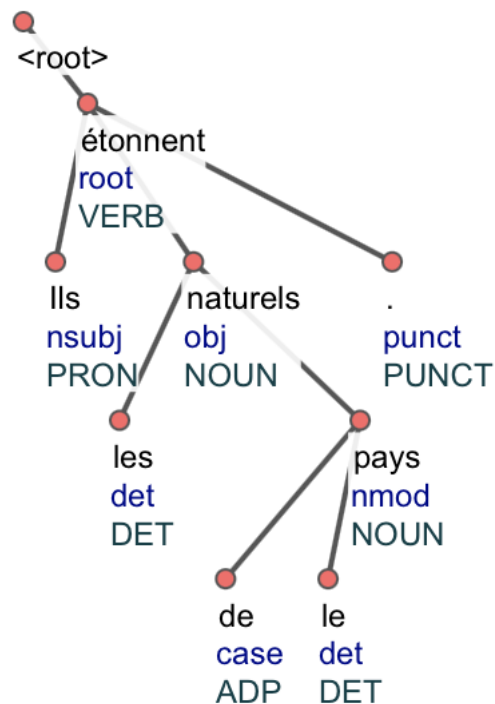


Figure 3.16 Arbre syntaxique pour sent_id = 183.

« Les naturels du pays » ici est un exemple des patrons qui se conforme à notre critère d'extraction. Prenons ensuite trois exemples qui ne seront pas présents dans nos résultats, chacun pour une raison particulière :

sent_id = 45

patron = le NN de le NN

text = Il y a aux murs du papier qui se décolle.

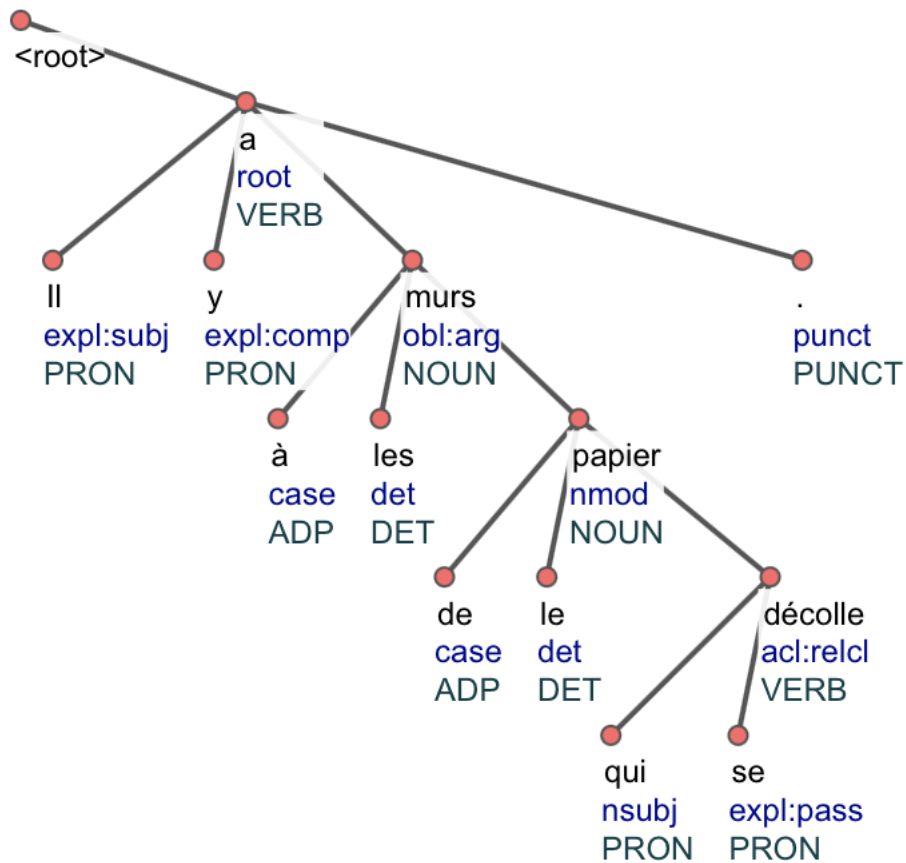


Figure 3.17 Arbre syntaxique pour sent_id = 45.

Cette phrase englobe trois problèmes typiques :

1. Le patron, « le NN de le NN », correspondant ici à « les murs du papier », est bien dans la phrase, mais il se trouve au sein d'un patron plus grand, « aux murs des papiers qui se décolle », qui est le « vrai » objet de notre extraction (l'algorithme se focalise sur des syntagmes complets, intégrant toutes leurs extensions, ce qui exclut « les murs du papier », qui n'est pas complet).
2. Ensuite, on a aussi ici affaire à une erreur de l'analyseur syntaxique UDPipe, l'objet direct du verbe « avoir » est « papier » alors l'analyseur syntaxique a identifié « murs » comme objet.
3. Il y a enfin un troisième problème dans cet exemple : le patron « le NN de le NN » ne correspond pas à un syntagme unique ici : la séquence comporte en fait deux éléments : « aux murs » (un circonstanciel) et « du papier » (l'objet). L'analyse syntaxique (s'il n'y avait pas eu d'erreur) devrait révéler qu'il s'agit d'un patron trans-syntagmatique incomplet, ce dont ne tient évidemment pas compte la méthode linéaire (qui retient donc cette séquence comme une séquence valide, ce qu'elle est effectivement en surface, sur la base des seules étiquettes morphosyntaxiques). Néanmoins, notre méthode échoue aussi ici à cause de la mauvaise analyse de notre analyseur syntaxique.

Nous avons montré dans cet exemple trois raisons pour lesquelles cette phrase n'est pas dans nos résultats. Examinons un autre exemple :

```
# sent_id = 47
```

```
# patron = le NN de le NN
```

```
# text = L'épaississement des toiles pleines de mouches indique la paix  
profonde des araignées.
```

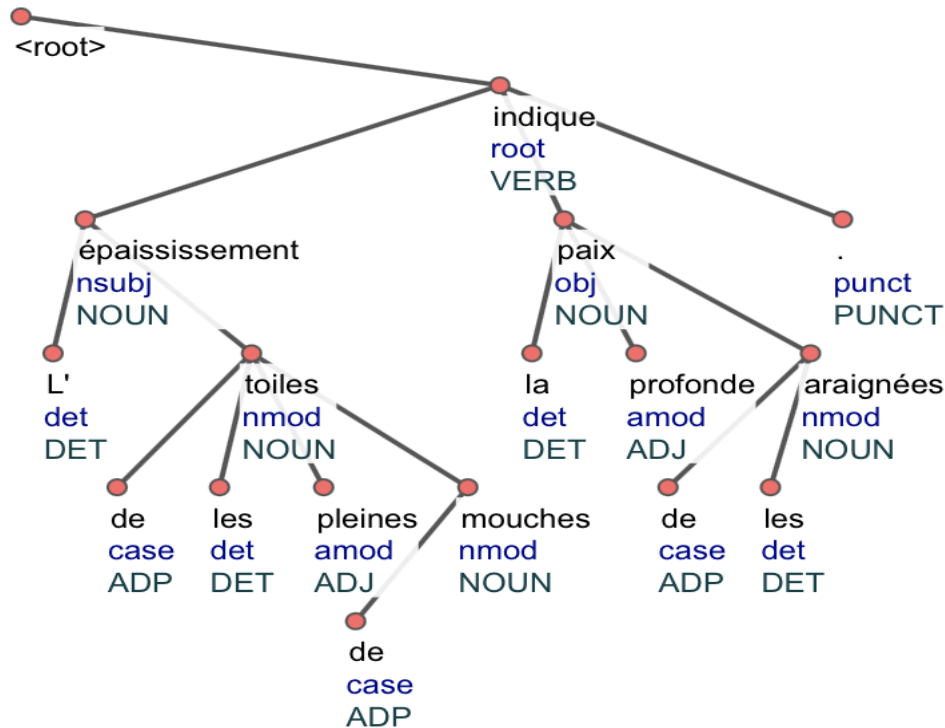


Figure 3.18 Arbre syntaxique pour sent_id = 47.

Cet exemple montre un autre problème, qui n'est pas lié à l'analyseur syntaxique. Le patron « le NN de le NN », qui correspond à la séquence « l'épaississement des toiles » se trouve dans une structure plus grande. En effet, le sous-arbre avec « épaississement » comme tête permet d'obtenir le patron en profondeur 3 suivant : « l'épaississement des toiles pleines mouches », mais même ce grand patron englobant « le NN de le NN » ne sera pas présent dans notre résultat parce que la préposition « de » est manquante (car ne correspondant pas à la profondeur d'extraction prédéfinie), ce qui le rend incomplet syntaxiquement. La seule manière d'avoir le patron complet avec « épaississement » comme tête est de sélectionner la profondeur 4, qui nous permet d'extraire le patron suivant : « l'épaississement des toiles pleines de mouches ». C'est pourquoi cette séquence est absente de nos résultats avec la profondeur 3, bien que présente dans le texte.

sent_id = 55

patron = le NN de le NN

text = Les populations campagnardes et maritimes ne sont pas tranquilles à l'endroit du diable.

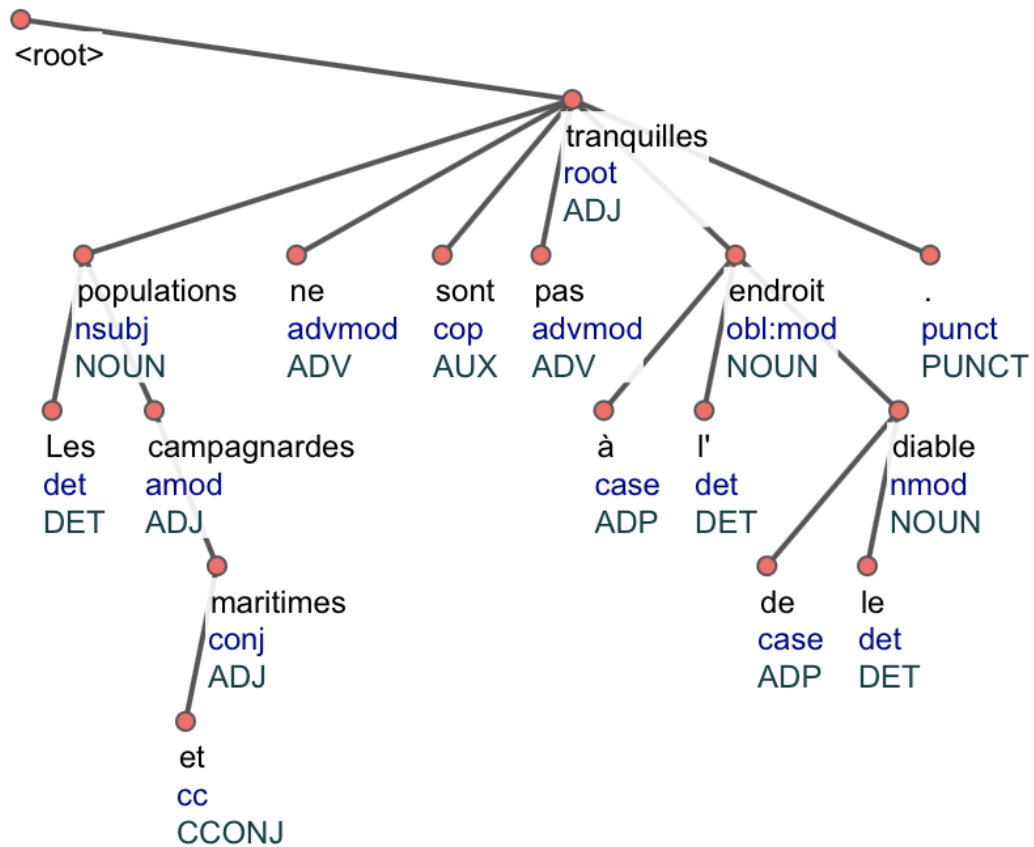


Figure 3.19 Arbre syntaxique pour sent_id = 55.

Dans ce dernier exemple, l'analyse syntaxique est correcte. Le patron « l'endroit du diable » est complet syntaxiquement, mais n'apparaît pas dans nos résultats. En effet,

l'analyse montre que la préposition fait partie du groupe syntaxique et c'est donc la séquence « à l'endroit du diable » (à le NN de le NN) qui apparaîtra dans les résultats. C'est un facteur dont il faut tenir compte pour certains syntagmes, en particulier les syntagmes prépositionnels comme on vient de le voir.

Nous avons examiné plusieurs cas limites, qui ne sont pas conformes, d'une façon ou d'une autre, à notre critère d'extraction. Nous pouvons conclure que le nombre d'occurrences réduit de nos résultats d'extraction est dû, d'une part à des erreurs de l'analyseur syntaxique UDPipe, d'autre part, au mécanisme d'extraction de notre méthode. Certaines séquences que nous pouvons trouver linéairement dans le texte ne sont pas présentes dans nos résultats parce qu'elles font souvent partie d'une structure syntaxique plus large (la méthode ne retient pas les séquences incluses dans une autre séquence de plus haut niveau). Selon les exemples que nous avons montrés, réduire ou augmenter la profondeur d'extraction ne serait pas une solution favorable parce que l'élément manquant n'est pas toujours d'un autre niveau : dans le dernier exemple, l'élément manquant est la préposition « à », qui est bien située au même niveau que « le » et « diable » : la modification de la profondeur d'extraction ne changerait pas le résultat final.

La raison à ce phénomène est principalement due à la modélisation des données : les phrases sont transformées en structures arborescentes avant l'extraction, ce qui (à dessein) empêche de les voir d'une manière linéaire. Aussi la recherche d'un patron donné à la surface du texte (une séquence d'étiquettes morphosyntaxiques) donnera-t-elle toujours plus de résultats que notre méthode. Nous pouvons ainsi conclure que notre méthode réduit considérablement le nombre de patrons dans les résultats, et elle augmente aussi le nombre de patrons interprétables, au prix d'un certain « silence ». Ce qui nous renvoie à une autre question : est-ce que les résultats de notre méthode sont supérieurs à ceux de la méthode linéaire ? Nous essayons ci-dessous de présenter nos arguments.

Premièrement, la réduction importante du nombre de patrons et, dans le même temps, la meilleure interprétabilité de ces patrons facilitent les analyses statistiques et littéraires. Contrairement à la méthode linéaire, il faut considérablement moins de temps pour trier les patrons et identifier ceux qui sont pertinents.

Deuxièmement, même si certaines occurrences d'un patron particulier n'apparaissent pas dans les résultats, ce processus n'est pas aléatoire : les séquences non identifiées le sont parce qu'elles ne sont pas conformes à nos critères d'extraction, et les critères d'extraction sont identiquement appliqués à tous les textes du corpus. Par conséquent, les analyses stylistiques auxquelles nous pouvons procéder avec cette méthode demeurent homogènes.

Troisièmement, les patrons passés sous silence avec notre méthode ne sont pas toujours définitivement perdus, en particulier lorsqu'il s'agit de séquences incluses dans des patrons plus larges. Dans le dernier exemple ci-dessus, nous pouvons voir que le patron « le NN de le NN » n'est pas extrait parce qu'il figure en fait dans une structure plus grande, en l'occurrence « à le NN de le NN ».

3.8 Conclusion

Ce chapitre consacré à la présentation de la méthodologie de la thèse adopte un point de vue combinatoire sur la stylistique. Nous avons présenté dans ce chapitre les méthodes d'extraction supervisée et non-supervisée, afin de pouvoir explorer les caractéristiques à la fois qualitatives et quantitatives des textes, par le truchement de deux types de patrons syntaxiques : les patrons caractérisants et les patrons identificatoires. Une nouvelle méthode pour mesurer la complexité syntaxique d'un texte est présentée dans la section 3.5.

Nous avons montré que les figures de style sont à la fois accessibles et caractéristiques du style des textes, comme l'ont montré les théoriciens cités dans la section 3.4. Les figures de style sont considérées comme des écarts dans la forme du texte, qui sont étroitement liés au fond du texte. D'après la méthode d'extraction supervisée que nous avons mise au point, nous constatons que le recours à l'analyse syntaxique donne souvent de résultats plus précis qu'une simple analyse superficielle : une analyse superficielle produit souvent des résultats qui demandent ensuite un nettoyage manuel intense. Un des défauts de notre approche est sans doute son manque de généralité : les figures de style que nous avons modélisées et extraites ne donnent qu'une vue partielle du style du texte. Il existe un nombre considérable de figures de style que nous ne pouvons pas extraire d'une manière automatique. C'est pour cette raison que nous avons décidé d'introduire en plus une méthode d'extraction non-supervisée, pour identifier des patrons syntaxiques qui, malgré leur manque de signification sur le plan stylistique, donnent plus d'information sur les caractéristiques syntaxiques du texte grâce à leur spécificité statistique.

La méthode d'extraction non-supervisée propose d'extraire les patrons syntaxiques d'après la grammaire de dépendance. Nous considérons qu'il s'agit d'une amélioration par rapport à la méthode d'extraction linéaire parce que l'extraction est plus précise et les résultats sont plus interprétables. Nous avons ensuite montré de façon empirique

que les profondeurs d'extraction 3 et 4 sont les meilleures en corpus. Une discussion sur les résultats d'extraction montre les raisons pour lesquelles nous obtenons toujours moins de patrons hiérarchiques que la méthode linéaire.

Ce chapitre a donc présenté la méthodologie de la thèse, qui consiste à explorer la capacité des structures syntaxiques à caractériser le style. Dans les deux chapitres suivants, nous allons appliquer ces méthodes d'extraction, supervisée et non-supervisée, d'abord sur les neuf romans de Victor Hugo et ensuite sur les romans des quatre auteurs de notre corpus pour pouvoir estimer la précision et la robustesse de notre méthode.

Chapitre 4 Caractérisation du style romanesque de Victor Hugo par la méthode des patrons syntaxiques

4.1 Introduction

Nous présentons dans ce chapitre l'application de nos méthodes d'extraction supervisée et non-supervisée de patrons syntaxiques au corpus des neuf romans de Victor Hugo. L'objectif est de montrer la capacité des patrons syntaxiques à caractériser les styles littéraires.

Les patrons syntaxiques caractérisants et identificatoires, qui ont été présentés en détail dans chapitre 3, sont des éléments que nous appelons « linguistiquement profonds » du fait de leur structure syntaxique hiérarchique, basée sur la grammaire de dépendance. Cette partie de la thèse a pour objectif de mettre en évidence l'importance de combiner ces deux catégories de patrons syntaxiques afin de produire des analyses à la fois quantitative et qualitative du style de Victor Hugo. Nous mettons également l'accent sur une des idées principales de la thèse : un seul patron suffit rarement à caractériser un style littéraire, alors que la méthode combinatoire, comprenant l'exploitation à la fois statistique et littéraire de l'ensemble des patrons syntaxiques, peut révéler efficacement les caractéristiques stylistiques des textes, qui confirment ou complètent les critiques littéraires sur le style de Hugo.

La partie 4.2 est consacrée à un bref parcours examinant les principaux points mis en avant par les critiques littéraires sur le style de Victor Hugo. La notion d'hétérogénéité est par exemple une des caractéristiques les plus proéminentes quand on parle du style de Hugo, surtout dans les romans rédigés pendant et après l'exil de l'auteur. Cette hétérogénéité se manifeste matériellement par les phénomènes stylistiques formels tels que l'antithèse (procédé stylistique qui consiste à opposer, dans

la même phrase, deux mots ou groupes de mots de sens contraire afin de mettre une idée en relief par un effet de contraste²⁰⁵), la métaphore qui relie souvent les mots abstraits aux mots concrets et la répétition abondante. Guidés par les critiques littéraires, mais ayant toujours l'intention de ne pas les suivre aveuglément, nous essayons de tester la capacité des patrons syntaxiques à apporter de nouveaux éléments révélateurs du style hétérogène de Hugo.

Nous évaluons, dans la section 4.3, la capacité de notre méthode à mesurer la complexité syntaxique, basée sur la grammaire de dépendance. La méthode consiste à mesurer la profondeur et la largeur de l'arbre syntaxique de dépendance correspondant à la phrase en cours d'analyse. Un indice représentant la profondeur et la largeur d'un arbre syntaxique est introduit pour quantifier la complexité syntaxique. La profondeur de dépendance (indiquant souvent les propositions en subordination) et la largeur de dépendance (indiquant souvent les propositions en coordination et en juxtaposition) sont des mesures efficaces pour caractériser le style d'un auteur à un niveau macroscopique.

L'application de la méthode d'extraction supervisée pour l'inversion, le clivage et la répétition est présentée dans la section 4.4. Les patrons caractérisants sont supposés avoir une fonction stylistique, mais ils sont en fait souvent peu informatifs en raison de leur faible fréquence. C'est pour cette raison que nous insistons sur le danger de recourir à un seul type de patron pour caractériser un style littéraire.

Les patrons identificatoires complètent les patrons caractérisants en fournissant une vue plus diversifiée de la syntaxe de Hugo. La partie 4.5 montre la capacité des patrons identificatoires à alimenter les analyses statistiques et qualitatives. Nos analyses de chaque patron spécifique et de l'ensemble des patrons spécifiques confirment ce que disent les critiques littéraires sur le style hétérogène de Hugo, marqué par la présence des structures antithétiques, métaphoriques et répétitives.

²⁰⁵ <https://www.larousse.fr/dictionnaires/francais/antithese/4320>, consulté le 14 juin 2021.

4.2 Le style romanesque de Victor Hugo

4.2.1 Aperçu sur le style de Victor Hugo

Considéré comme un des plus grands écrivains de la littérature française, Victor Hugo a fait l'objet de nombreuses études littéraires et stylistiques. La plupart des critiques sont élogieuses sur le style hugolien. André Gide considère Victor Hugo comme « le plus puissant assembleur d'images et le plus sûr maître de la syntaxe²⁰⁶ » (Gide, 1949, p. XXXII). Sartre parle de « cet homme étonnant » qui « possède je ne sais quelle puissance surhumaine²⁰⁷ » (Sartre, 1972 [1971], III, p. 283). Flaubert exprime son admiration pour Hugo en écrivant que « Hugo [...] enfoncera tout le monde²⁰⁸ » (Flaubert, 1926 [1887], II, p. 164).

Le style de Hugo est un sujet vaste qui ne se résume pas en quelques mots. La diversité qui existe dans les qualités de style hugolien ne nous permet pas de produire une conclusion simple valant pour l'ensemble de ses œuvres. Cette complexité stylistique se traduit non seulement par le grand éventail de genres littéraires que l'auteur a abordés, mais aussi par la période d'exil qui divise les créations de Hugo en deux parties, tout en déclenchant un potentiel changement de style dans ses œuvres après l'exil. La thèse cherche à repérer les caractéristiques stylistiques de Hugo dans ses romans, plus précisément, les caractéristiques qui peuvent être révélées par les phénomènes syntaxiques.

Le roman n'était pas le premier choix de Victor Hugo pour s'adresser au public. Comme explique Anne Ubersfeld dans *Le Roi est le Bouffon*, le roman détournait Hugo

²⁰⁶ Gide, André Éditeur scientifique Préfacier. 1949. *Anthologie de la poésie française*. Paris, France: Gallimard, p. XXXII.

²⁰⁷ Sartre, Jean-Paul. 1972 [1971]. *L'idiot de la famille*. Paris, France: Gallimard, p. 283.

²⁰⁸ Flaubert, Gustave. 1926 [1887]. *Correspondance*. Paris, France: L. Conard, p. 164.

de « sa véritable parole²⁰⁹ » (Ubersfeld, p.11). Judith Wulf fait la même remarque en prenant le roman pour « un genre secondaire » pour Hugo, par rapport à la poésie :

« Par opposition à la poésie, qui constitue le socle de son identité et dessine, tout au long de sa vie, la ligne de son destin de créateur, en marge du drame, qui concentre ses espoirs de renommée et de fortune et constitue un laboratoire dialogique, le roman fait pâle figure²¹⁰. » (Wulf, 2014, p. 18)

Pourtant, Wulf argumente sur la nécessité d'étudier les romans de Hugo parce qu'ils acquièrent un rôle important pendant l'exil. Le fait que le roman ait été promu de « choix » à « nécessité » avec l'exil est un bon point de départ pour comprendre le changement de style dans les écrits de Hugo, mais aussi le changement de son point de vue sur la littérature en général.

La promotion du rôle de roman dans les écrits de Hugo est due à l'exil. L'exil empêche Hugo de s'adresser directement au public comme dramaturge. C'est à ce moment-là que l'auteur commence à recourir aux textes narratifs :

« Lorsque la publication d'un écrit devient le principal moyen d'existence de l'œuvre, c'est l'ampleur de la diversité qui pose désormais problème : chercher à traduire, dans une même unité, toutes les voix de l'histoire collective, même lorsqu'elles sont discordantes, risque de les trahir²¹¹ ». (Wulf, 2014, p. 17)

²⁰⁹ Ubersfeld, Anne (1921-2010). 2001. *Le roi et le bouffon : étude sur le théâtre de Hugo de 1830 à 1839* (Éd. rev.) / Anne Ubersfeld, p. 11.

²¹⁰ Wulf, Judith. 2014. *Étude sur la langue romanesque de Victor Hugo: le partage et la composition*. Paris, France: Classiques Garnier, p.18

²¹¹ Wulf, Judith. 2014. *Op.cit.*, p.17

À « l'ampleur de la diversité » après l'exil, s'oppose la conception du style de Hugo avant l'exil, où ce dernier cherche à « restaurer », « harmoniser » et « combler ²¹² » (Hugo, 1985 [1827], p. 24) pour rendre les textes plus lisibles. Ainsi, le style dans les romans d'exil de Hugo est caractérisé par la diversité qui se manifeste non seulement dans le fond (la diversité des voix), mais aussi dans la forme (la diversité des techniques d'écriture), alors que son style avant l'exil par contraste pourrait être caractérisé par une certaine absence de diversité²¹³. Les différences entre les œuvres de jeunesse de Hugo et ses œuvres pendant et après l'exil ont également été constatées par Brombert, qui pense « qu'aucun des romans de jeunesse de Hugo ne peut prétendre être un grand roman²¹⁴ ». (Brombert, 1985, p. 23)

En effet, la diversité, ou « l'hétérogénéité²¹⁵ » (Wulf, 2014) sont les mots clés pour comprendre le style de Hugo dans les romans de l'exil :

« À la dramaturgie de la vaine parole succède alors le roman du bavardage, l'autre « bouteille à la mer » qui, renonçant à l'énonciation englobante du lyrisme et à l'unité de perspective du drame, se jette à corps perdu dans une totalité de divergence et court le risque de renoncer à l'efficience d'une cohérence surplombante, pour mieux gagner l'espoir de donner la parole à tous²¹⁶ » (Wulf, 2014, p. 20).

« Crise et critique vont désormais de pair, érigeant un événement en posture puis en principe d'écriture, entraînant avec elle toute une série de formes qui valorisent l'altérité, l'hétérogénéité, comme autant de forces de résistance à

²¹² Hugo, Victor. 1985 [1827]. « Préface de Cromwell » dans *Œuvres complètes*. Édité par Guy Rosa, Jacques Seebacher, et Groupe Hugo. 16 vol. Paris, France : R. Laffont, p. 24.

²¹³ Pour la notion d'harmonie dans les premiers romans de Victor Hugo, voir Grossman, Kathryn M. 1986. *The Early novels of Victor Hugo: towards a poetics of harmony*. Genève, Suisse : Droz.

²¹⁴ Brombert, Victor, et Héloïse Neefs. 1985. *Victor Hugo et le roman visionnaire*. Paris, France: PUF, p. 23.

²¹⁵ Wulf, Judith. 2014. *Op.cit.*

²¹⁶ *Ibid.*, p. 20.

la catégorisation schématique du discours dogmatique²¹⁷ » (Wulf, 2014, p. 92).

La première citation met en contraste le roman avec la dramaturgie hugolienne. Victor Hugo a toujours eu la volonté de s'adresser à tous par le truchement de la langue, malgré le changement de médium de communication, de la dramaturgie au roman. À l'occasion des funérailles de Balzac, Hugo a bien fait comprendre qu'il considérait le roman comme un genre qui prend « toutes les formes et tous les styles²¹⁸ » (Hugo, 1985 [1850], p. 316-318). Ainsi, son roman porte cette même volonté qui se manifeste par l'altérité et l'hétérogénéité, ayant pour but de toucher la totalité du public.

La notion d'hétérogénéité est vaste et abstraite, comme caractéristique stylistique. La réconciliation de cette notion d'hétérogénéité avec notre approche formelle devrait être matérialisée par des formes d'énonciations hétérogènes. L'étude de Judith Wulf relève d'une manière exhaustive la matérialisation de l'hétérogénéité dans les textes. Bien que la plupart de ces phénomènes concernent le niveau sémantique, nous présentons quelques phénomènes d'ordre syntaxique, qui nous semblent aussi concrétiser cette notion d'hétérogénéité, emblématique du style de Hugo.

4.2.2 La matérialisation de l'hétérogénéité dans les œuvres de Victor Hugo

● L'antithèse

Une des manifestations linguistiques et formelles de l'hétérogénéité est l'antithèse²¹⁹. Ce phénomène est fréquent et caractéristique des romans de Hugo.

²¹⁷ *Ibid.*, p. 92.

²¹⁸ Voir l'allocution de Hugo à l'occasion des funérailles de Balzac, le 21 août 1850, Hugo, Victor. 1985. *Œuvres complètes*. Édité par Guy Rosa, Jacques Seebacher, et Groupe Hugo. 16 vol. Paris, France : R. Laffont, VII, p. 316-318.

²¹⁹ <https://www.larousse.fr/dictionnaires/francais/antithese/4320>, consulté le 14 juin 2021.

Ferdinand Brunot et Charles Bruneau constatent que « l'antithèse abonde dans les vers de Victor Hugo²²⁰ » (Brunot et al., 1948, p. 236). Hugo lui-même met l'accent sur l'importance de cette figure de contraste en se référant à la notion de « génie²²¹ » :

« Les plus hauts génies, les intelligences encyclopédiques aussi bien que les esprits épiques, Aristote aussi bien qu'Homère, Bacon aussi bien que Shakespeare, détaillent l'ensemble pour le faire comprendre, et ont recours aux oppositions, aux contrastes et aux antinomies²²² » (Hugo, 1985 [1827], p. 699).

Il développe ensuite la notion d'antithèse à travers une suite de comparaison : « La dissection fait le chirurgien, l'analyse fait le philosophe ; l'antithèse est le grand organe de la synthèse ; c'est l'antithèse qui fait la lumière²²³ » (Hugo, 1985 [1827], p. 699). Les antithèses hugoliennes ont pour but de « récuser la logique exclusive du vrai et du faux²²⁴ » (Dufour, 2004, p. 289). Ainsi, il s'agit d'un phénomène littéraire qui rend la signification incertaine, il fait surgir une ouverture dans laquelle les interprétations restent diverses et hétérogènes.

● La métaphore

La métaphore hugolienne est une des caractéristiques les plus discutées par les critiques littéraires. La métaphore est un outil efficace pour « rendre sensible par une image des idées²²⁵ » (Hugo, 1985 [1827], p. 16). Une des particularités des métaphores

²²⁰ Brunot, Ferdinand, et Charles Bruneau. 1948. *L'époque romantique. 1 vol. Histoire de la langue française, des origines à 1900* 12. Paris: A. Colin, p. 236.

²²¹ Pour une discussion sur la notion de « génie », voir chapitre 2, partie 2.2.1

²²² Hugo, Victor. 1985 [1827]. « Préface de mes œuvres et post-scriptum de ma vie » dans *Œuvres complètes*. Édité par Guy Rosa, Jacques Seebacher, et Groupe Hugo. 16 vol. Paris, France : R. Laffont, p. 699.

²²³ *Idem*.

²²⁴ Dufour, Philippe. 2004. *La pensée romanesque du langage*. Paris : Seuil, p. 289.

²²⁵ Hugo, Victor. 1985 [1827]. « Préface de Cromwell » dans *Œuvres complètes*. Édité par Guy Rosa, Jacques Seebacher, et Groupe Hugo. 16 vol. Paris, France : R. Laffont, p. 16.

hugoliennes dans ses romans est la combinaison de mots abstraits avec des mots concrets : « Cette figure [la métaphore] est employée dans les romans de Hugo pour mettre en rapport des notations sensibles avec des éléments abstraits²²⁶ » (Wulf, 2014, p. 64). La combinaison des mots issus des catégories sémantiques différentes dans les métaphores hugoliennes crée souvent un effet de « rapprochement inattendu²²⁷ » (Huguet, 1905, p. 363) selon Edmond Huguet. Huguet explique ensuite le mécanisme des métaphores hugoliennes :

« Grâce à ces habitudes d'esprit et à ces conceptions, le poète [Hugo] peut facilement provoquer l'hallucination qui transfigure les objets, et, en leur laissant à peu près leur forme, les fait changer de nature²²⁸ » (Huguet, 1905, p. 364).

Ce mécanisme de changement de nature a pour objectif de créer une tension dans les textes :

« Ceux [les métaphores] qui ont été étudiés ne consistent pas à remplacer un signifiant par un autre, mais établissent une tension soit syntaxique, soit, plus généralement, prédicative. [...] De manière générale, l'utilisation du concret ne consiste pas uniquement à donner l'impression qu'un élément abstrait est concret, mais à structurer la compréhension d'un phénomène de manière sensible²²⁹ » (Wulf, 2014, p. 67).

Ainsi, la métaphore hugolienne n'est pas ornementale, elle consiste à produire une vision inédite et réorganisée du réel afin de le rendre accessible à ses lecteurs.

²²⁶ Wulf, Judith. 2014. *Op.cit.*, p. 64.

²²⁷ Huguet, Edmond. 1905. *Les métaphores et les comparaisons dans l'oeuvre de Victor Hugo*. Paris : Hachette, p. 363.

²²⁸ Huguet, Edmond. 1905. *Op.cit.*, p. 364.

²²⁹ Wulf, Judith. 2014. *Op.cit.*, p. 67.

● La répétition

Le style de Hugo est également marqué par l'utilisation abondante de structures répétitives. Les répétitions, que ce soit de mots, de structures syntaxiques ou de thèmes récurrents, sont un des phénomènes emblématiques de ses romans :

« Elles [les figures de répétition] organisent le texte en reliant des groupes qui ont les mêmes signifiants. Ce mode d'organisation est très développé dans les romans de Victor Hugo. Il se substitue à la structure linéaire²³⁰ » (Wulf, 2014, p. 76).

Une des figures de style emblématiques de la catégorie dont parle Wulf est le parallélisme. Selon Wulf, le parallélisme est un des exemples qui rompt la structure linéaire du texte en créant « un effet d'architecture » en profondeur et en épaisseur : « Pour créer cet effet d'architecture, Hugo joue sur des phénomènes de parallélismes et de contrastes proches de la structure du poème²³¹ » (Wulf, 2014, p. 177).

Les trois figures formelles que nous venons de présenter sont des manifestations de l'hétérogénéité du style des romans de Hugo, surtout pour ceux publiés après l'exil. Les trois phénomènes créent chacun un effet particulier : soit par la possibilité d'interprétation diverse du fait de la signification incertaine (l'antithèse) ; soit par la combinaison et le rapprochement de mots de sémantismes hétérogènes (la métaphore) ; soit par l'accumulation d'éléments visant à densifier le roman (la répétition). Selon Brombert, les phénomènes d'hétérogénéités créent une harmonie dans les œuvres de Hugo : « Les antithèses et oxymorons célèbres, les tropes séditieux, loin d'affirmer des

²³⁰ Wulf, Judith. 2014. *Op.cit.*, p. 76.

²³¹ Wulf, Judith. 2014. *Op.cit.*, p. 177.

antagonismes irréconciliables, fonctionnent en fait comme éléments harmonisateurs²³² » (Brombert, 1985, p. 10).

Notre thèse ne vise pas à donner une analyse complète des phénomènes stylistiques dans les romans de Victor Hugo. Nous proposons une méthode stylistique nouvelle qui peut être appliquée dans le cadre de recherches littéraires ou linguistiques. Par conséquent, l'objectif de ce chapitre est d'étudier la syntaxe hugolienne par les patrons caractérisants et identificatoires pour vérifier la capacité de notre méthode à déchiffrer les caractéristiques stylistiques de Hugo. Dans les sections suivantes, nous essayons de montrer l'intérêt de la syntaxe pour ce faire.

²³² Brombert, Victor, et Héloïse Neefs. 1985. *Op. cit.*, p. 10.

4.3 Mesurer la complexité syntaxique dans les romans de Victor Hugo

Avant de lancer l'application de la méthode supervisée et la méthode non-supervisée sur les romans de Victor Hugo, nous appliquons notre méthode pour analyser la complexité syntaxique de ses œuvres (via l'analyse des phrases).

4.3.1 Calcul de la complexité syntaxique des romans de Victor Hugo

Dans le tableau suivant, nous présentons la complexité syntaxique des phrases dans les neuf romans de Victor Hugo.

Romans	Largeur	Profondeur	LPIndice
BUG_JARGAL	6.20	4.69	29.08
HAN_DISLANDE	6.23	4.74	29.56
LE_DERNIER_JOUR_DUN_CONDAMNE	5.77	4.19	24.19
NOTRE_DAME_DE_PARIS	5.90	4.36	25.74
CLAUDE_GUEUX	6.08	4.67	28.42
LES_MISERABLES	6.21	4.48	27.82
LES_TRAVAILLEURS_DE_LA_MER	5.96	4.32	25.75
LHOMME QUI RIT	5.77	4.22	24.32
QUATRE_VINGT_TREIZE	5.69	4.18	23.79

Tableau 4.1 Complexités syntaxiques dans les neuf romans de Victor Hugo.

Les romans sont présentés dans ce tableau du haut en bas en fonction de leur date de publication (suivant un ordre chronologique) : *Bug-Jargal* est le premier roman de Victor Hugo publié en 1826 et *Quatrevingt-treize* (selon la graphie voulue par l'auteur), le dernier roman de Victor Hugo publié en 1874. Pour mieux visualiser l'évolution de la complexité syntaxique dans les écrits de Hugo, nous pouvons afficher les résultats dans la figure suivante.

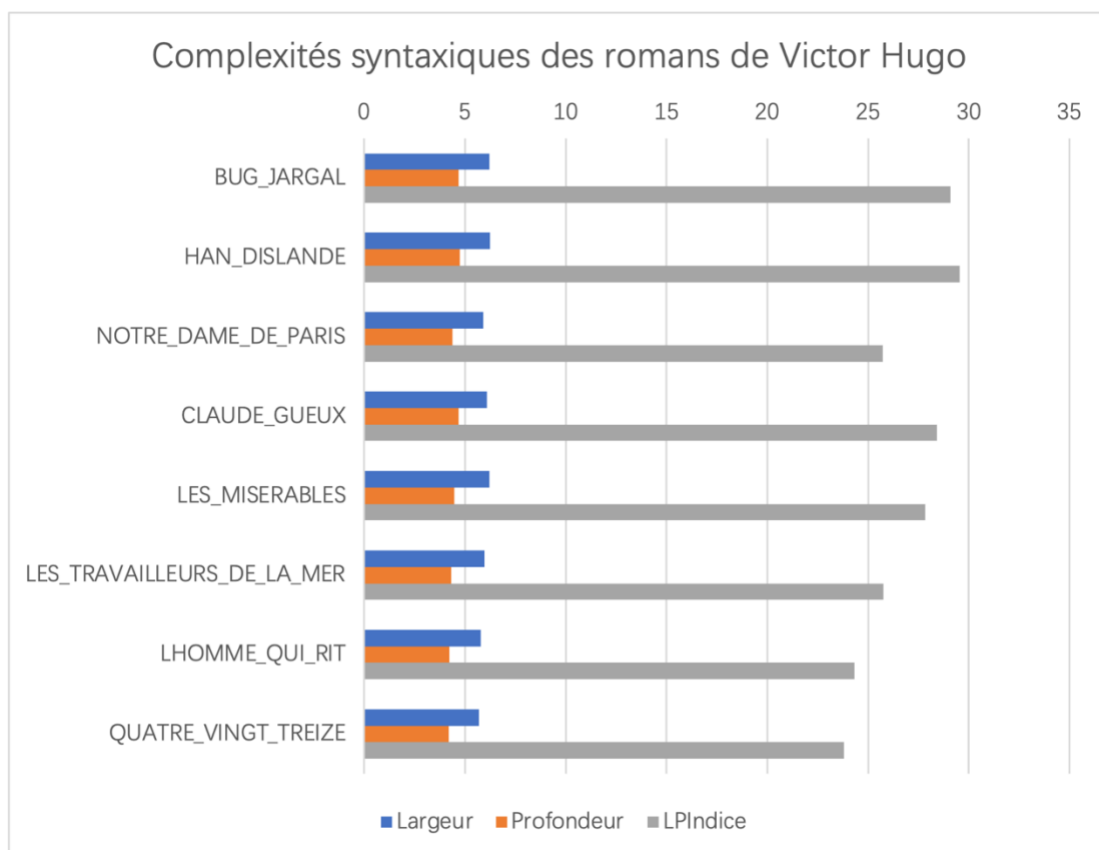


Figure 4.1 Complexités syntaxiques dans les neuf romans de Victor Hugo.

Bien que les chiffres obtenus pour les largeurs et les profondeurs ne présentent pas de différences importantes, le LPIndice sur la figure ci-dessus montre une réduction de la complexité syntaxique dans les derniers romans. Les deux premiers romans de Hugo (*Bug-Jargal* et *Han d'Islande*) contiennent des phrases dont la structure syntaxique est la plus complexe, alors que les deux derniers romans de Hugo (*L'Homme qui rit* et *Quatrevingt-treize*) contiennent des phrases moins complexes. *Le Dernier Jour d'un condamné* est un roman exceptionnel qui se compose d'une petite pièce de théâtre au début, puis d'un long monologue du « condamné ». Ainsi, ce roman se compose essentiellement des discours rapportés (les dialogues et les monologues). Nous pensons que c'est à cause de ces particularités, entre genre romanesque et genre théâtral, que les résultats obtenus sont si spécifiques sur le plan de la complexité syntaxique.

Dans le corpus Victor Hugo, nous avons constaté que *Han d'Islande* est le roman dont les phrases ont le plus tendance à contenir des subordinations (ce qui est visible à

travers l'importance de l'indice « profondeur » dans ce roman). Cette œuvre comprend également le plus de phrases coordonnées ou juxtaposées ; dans les quatre derniers romans de Hugo, la « profondeur » et la « largeur » diminuent progressivement jusqu'aux valeurs les moins importantes de *Quatrevingt-treize*, le dernier roman, où il y a le moins de phrases complexes en situation de subordination, coordination ou juxtaposition.

Le résultat que nous avons obtenu contredit partiellement ce qu'affirment les critiques littéraires sur le style des romans de Victor Hugo. Nous avons évoqué dans la partie précédente que, selon Judith Wulf, une des caractéristiques des romans de l'exil de Victor Hugo est l'hétérogénéité, à savoir, le roman comme rouage de différents genres littéraires, de différentes techniques stylistiques et de différentes voix. D'après nos résultats, nous avons plutôt constaté que les phrases hugoliennes, dans ses derniers romans, ont tendance à être moins complexes que dans ses romans d'avant l'exil : les phrases contiennent de moins en moins de propositions en situation de subordination, de coordination et de juxtaposition.

Pourtant, la complexité n'égale pas l'hétérogénéité. La question de l'hétérogénéité est un problème général qui se manifeste à travers plusieurs niveaux linguistiques. La méthode que nous proposons dans cette partie ne révèle qu'une partie des caractéristiques syntaxiques, à savoir la complexité syntaxique, sans entrer dans le détail de la composition des phrases. La caractéristique principale que nous avons constatée jusqu'ici est une moindre complexité des phrases de Hugo dans les romans d'après l'exil que dans les romans d'avant l'exil. Ce phénomène se traduit par une diminution générale des phrases complexes avec subordination, coordination et juxtaposition dans les derniers romans de Hugo. Nous explorons d'autres caractéristiques stylistiques sur la syntaxe de Hugo dans les parties suivantes.

4.4 Application de la méthode d'extraction supervisée sur le corpus Victor Hugo

4.4.1 L'inversion du sujet

Après avoir analysé la complexité syntaxique des romans de Victor Hugo, nous commençons cette partie par l'application de la méthode supervisée sur ce même corpus. Dans la section 3.6.1 (Chapitre 3), nous avons présenté la méthode d'extraction des inversions du sujet ainsi que le nettoyage des résultats pour ne garder que les inversions stylistiques. Nous nous concentrons dans cette partie sur l'analyse de la typologie des inversions et la variation du nombre d'occurrences de cette figure de style dans les différents romans de Hugo. Nous examinerons ensuite de plus près les occurrences afin de déterminer la fonction stylistique de cette figure. Précisons que la thèse n'est pas une étude systématique des inversions du sujet dans les œuvres de Hugo : nous n'avons pas pour objectif de proposer une analyse exhaustive de toutes les inversions dans ses textes. Le but de cette section, comme nous avons déjà évoqué dans l'Introduction de ce chapitre, est d'abord de vérifier si l'inversion est bien un marqueur du style de Hugo. Cette section est aussi un exemple de l'application concrète de notre méthode, qui sera reprise pour l'analyse des autres auteurs.

4.4.1.1 La définition et les fonctions stylistiques des inversions du sujet

Nous adoptons la définition de Galperin²³³ pour l'inversion stylistique :

« Contrairement à l'inversion grammaticale, l'inversion stylistique ne change pas le sens structurel de la phrase, c'est-à-dire que le changement dans la juxtaposition des membres de la phrase n'indique pas de sens structurel, mais a une fonction superstructurale. L'inversion stylistique vise à attacher un

²³³ Galperin, I. R. 1997. *Stylistics*. La Habana: s.n., p. 203.

stress logique ou une coloration d'émotion supplémentaire à la signification de surface de l'énoncé. Par conséquent, un motif d'intonation spécifique est le satellite inévitable de l'inversion ». (Galperin, p. 203)

D'après cette citation, l'inversion du sujet est considérée comme une sorte de mise en relief sans altération du sens de la phrase. Elle crée un effet stylistique en attirant l'attention du lecteur sur le point précis où l'écrivain met l'accent. Ullmann classe les effets stylistiques que peuvent créer les inversions dans trois catégories (notons que dans cette citation, X signifie le déclencheur d'inversion, V le verbe dans la construction inversée et S le sujet) :

« Ullmann qualifiait l'inversion du sujet nominal de « *procédé impressionniste* » : X constituant une sorte de « fond de tableau » thématique sur lequel vient s'inscrire V – un état ou l'image arrêtée du surgissement d'un phénomène –, puis seulement est énoncé le S support du prédicat – d'où des effets de suspense, d'insistance ou de nouveauté sur cet élément rhématique²³⁴ » (Fuchs, 2019).

Ainsi, nous constatons trois effets stylistiques que pourrait créer une inversion : les effets de suspense, d'insistance et de nouveauté. Le déclencheur d'inversion est un élément linguistique qui est à l'origine de la figure de style. Par exemple, dans :

24. *Dedans éclatait la mitraille.*

« Dedans » est un adverbe qui déclenche l'inversion du sujet. Selon Christine Tellier et Daniel Valois, le déclencheur est une des propriétés principales de l'inversion stylistique :

²³⁴ Fuchs, Catherine. 2019. « La postposition du sujet nominal : paramètres linguistiques et effets stylistiques ». In *L'Ordre des mots à la lecture des textes*, édité par Agnès Fontvieille-Cordani et Stéphanie Thonnérieux, 27-44. Textes & Langue. Lyon: Presses universitaires de Lyon.

« Nous identifierons quatre propriétés principales de l'inversion stylistique : 1) la présence obligatoire d'un élément déclencheur ; 2) les contraintes sur le type d'inversion permise ; 3) la position du SN inversé dans les phrases avec verbe à temps composé ; et 4) les types de complément compatibles avec le sujet inversé²³⁵ » (Tellier et al., 2018).

L'étude des déclencheurs d'inversion est un vaste sujet, parce que « s'il impacte de façon évidente le sens du verbe, le déclencheur interroge cependant par sa nature, sa fonction ou encore la raison pour laquelle il est antéposé dans l'énoncé²³⁶ » (Roig, 2017). L'étude de la typologie des inversions dans nos résultats se base sur la classification des différents déclencheurs.

4.4.1.2 Analyses quantitatives des inversions du sujet

Nous avons trouvé 110 occurrences d'inversions du sujet stylistiques dans les neuf romans de Victor Hugo. Bien que l'inversion du sujet soit courante dans les écrits de Hugo, la plupart des inversions ne sont pas stylistiques. À l'exception des inversions grammaticales, comme les incises et les phrases interrogatives, nous supprimons également les inversions dont l'usage est plus ou moins figé dans la langue française (voir la section 3.6.1 du Chapitre 3). Nous avons fait une classification des inversions selon les déclencheurs en jeu, placés devant la construction inversée verbe-sujet. Ainsi, la modélisation des inversions prend la forme suivante : Déclencheur+Verbe+Sujet.

²³⁵ Tellier, Christine, et Daniel Valois. 2018. *Constructions méconnues du français*. Paramètres. Montréal: Presses de l'Université de Montréal.

²³⁶ Roig, Audrey. 2017. « L'inversion du sujet (pro)nominal : un indice syntaxique ? » *Verbum* (Presses Universitaires de Nancy) XXXIX (2): 193.

Nous présentons d'abord les grandes catégories des inversions, avant d'entrer dans les détails. Quatre grandes catégories se trouvent dans nos résultats :

1. Groupe prépositionnel+Verbe+Sujet

25. *À cela se joignit la malveillance.* (Han d'Islande)

2. Groupe adverbial+Verbe+Sujet

26. *Dedans éclatait la mitraille, dehors se dressait la retirade.* (Quatrevingt-treize)

3. Groupe mis en relief+Verbe+Sujet

27. *C'est de cette façon, c'est par le hasard d'un mot d'orateur mal compris qu'est tombée la tête de madame Élisabeth.* (Quatrevingt-treize)

4. Concessif avec inversion du sujet.

28. *Tu es un drôle avec qui j'ai démangeaison de ripailler, **dût-il** m'en coûter un douzain neuf de douze tournois.* (Notre-Dame de Paris)

Ensuite, nous localisons les inversions catégorisées dans chacun des romans de Victor Hugo. Nous présentons leurs fréquences respectives dans ces romans dans le tableau suivant, les romans sont classés du haut en bas selon leur date de publication chronologique :

Romans	Concessif	Prépositionnel	Adverbial	Mise en relief
Bug-Jargal	1	2		
Han d'Islande	1	10		1
Le Dernier Jour d'un condamné			1	
Notre-Dame de Paris	5	7	4	1
Claude Gueux				
Les Misérables	9	10	4	1
Les Travailleurs de la mer		5	1	5
L'Homme qui rit		21		6
Quatrevingt-treize		10	2	3

Tableau 4.2 Fréquences absolues des inversions du sujet dans les romans de Hugo.

Nous pouvons constater qu'il y a une croissance du nombre d'occurrences pour les inversions concessives dans les romans de Hugo, mais cette tendance s'arrête complètement avec les trois derniers romans. La catégorie « Groupe mis en relief+Verbe+Sujet » est également de plus en plus utilisée dans les romans publiés après *Claude Gueux*. Pour éliminer l'effet de la longueur variable des textes sur le nombre d'occurrences, nous divisons les occurrences de chaque catégorie par le nombre total d'occurrences dans ce roman pour obtenir les valeurs relatives. Nous présentons dans la figure suivante les pourcentages des différentes catégories d'inversion pour chaque roman de Victor Hugo.

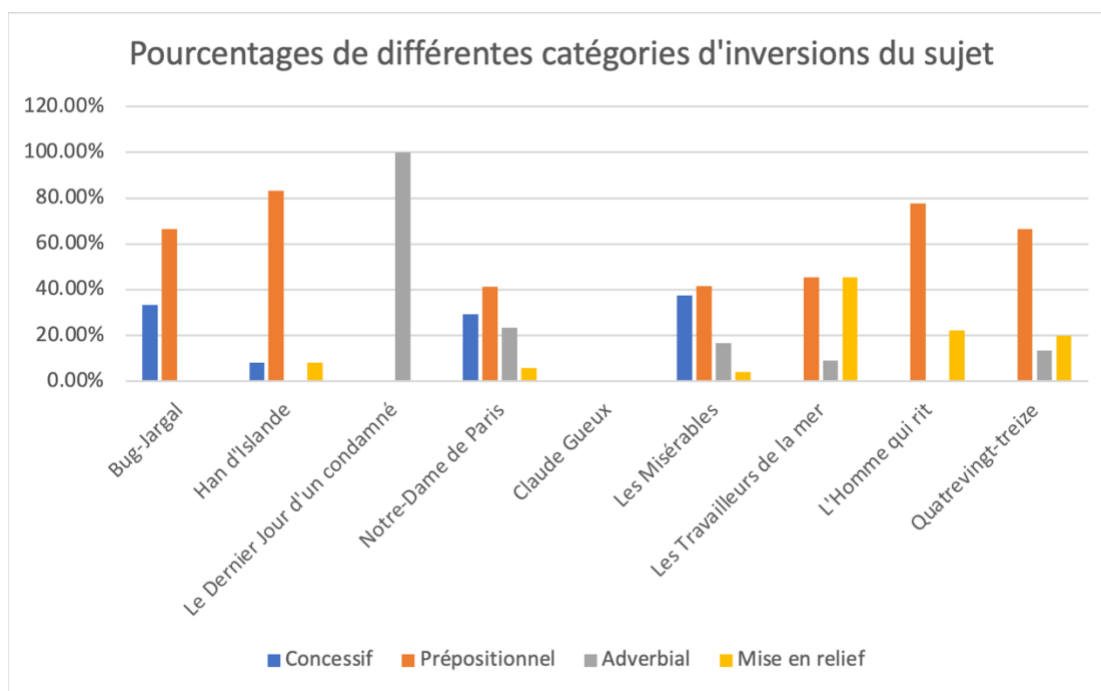


Figure 4.2 Pourcentages de différentes catégories d'inversions du sujet dans les romans de Hugo.

Si nous utilisons *Claude Gueux* pour couper les neuf romans en deux sous-corpus selon leur date de publication chronologique, nous constatons une évolution dans l'utilisation des différentes catégories d'inversion entre les deux périodes de publication, et nous résumons ici les caractéristiques quantitatives des inversions du sujet dans les romans de Hugo :

1. La catégorie d'inversion stylistique la plus courante est « Groupe Prépositionnel+Verbe+Sujet ». Cette catégorie est très présente dans les deux premiers et les deux derniers romans de Hugo.
2. *Notre-Dame de Paris* et *Les Misérables* sont les seuls romans à comprendre toutes les catégories d'inversions stylistiques énumérées. Les pourcentages de

répartition se ressemblent beaucoup, malgré une publication avec presque 30 ans de décalage.

3. La combinaison de la mise en relief et de l'inversion stylistique du sujet apparaît beaucoup plus souvent dans les écrits de la deuxième période de publication.
4. *Les Misérables* compte le plus d'occurrences d'inversions concessives (presque toutes les occurrences d'inversions concessives se trouvent dans *Les Misérables* et *Notre-Dame de Paris*). Les inversions concessives sont une des manières d'exprimer la concessif, mais tous les autres n'emploient pas cet usage, ainsi, elles pourraient être considérées comme un choix volontaire de Hugo qui se distingue des autres auteurs.
5. *Le Dernier Jour d'un condamné* ne contient qu'une seule occurrence d'inversion stylistique. Deux explications sont possibles : a) le roman est relativement plus court que les autres. Cette explication vaut également pour *Claude Gueux*, un roman extrêmement court avec seulement 581 phrases. b) Le genre du texte pourrait aussi jouer un rôle dans la détection des figures de style comme l'inversion du sujet. Le fait que notre résultat pour ce roman est exceptionnel indique la capacité potentielle des figures de style à différencier les genres littéraires.

4.4.1.3 Analyses qualitatives des inversions

Nous procédons dans cette partie à l'analyse qualitative des inversions du sujet dans les romans de Hugo. Cette partie consiste à vérifier si l'inversion reflète le style « hétérogène » de Victor Hugo, en examinant les cas d'antithèse, de métaphore, ou de répétition (fréquemment citées par les critiques pour justifier le caractère hétérogène du

style de l'auteur), ou d'autres caractéristiques qui n'auraient pas encore été repérées par les critiques littéraires.

Le tableau suivant présente les détails des patrons modélisés des inversions du sujet :

Catégorie	patron	Fréquence totale	Fréquence dans les premiers cinq romans	Fréquence dans les derniers quatre romans	
Groupe Prépositionnel+Verbe+Sujet	dérrière+verbe+sujet	5	2	3	
	à+verbe+sujet	28	12	16	
	dans+verbe+sujet	6	0	6	
	en dehors de+verbe+sujet	1	0	1	
	en deçà de+verbe+sujet	1	0	1	
	sous+verbe+sujet	4	0	4	
	de+verbe+sujet	7	1	6	
	devant+verbe+sujet	2	1	1	
	sur+verbe+sujet	9	4	5	
	près+verbe+sujet	3	1	2	
	après+verbe+sujet	1	0	1	
	entre+verbe+sujet	1	1	0	
	là+verbe+sujet	3	0	3	
	dedans+verbe+sujet	1	0	1	
	à côté+verbe+sujet	2	1	1	
	au-delà+verbe+sujet	2	2	0	
	Groupe adverbial+ Verbe+Sujet	dérrière+verbe+sujet	1	1	0
au-dessous+verbe+sujet		1	1	0	
au milieu+verbe+sujet		2	2	0	
au fond+verbe+sujet		1	0	1	
puis+verbe+sujet		1	1	0	
Groupe prépositionnel mis en relief+ Verbe+Sujet		C'être de...que+verbe+sujet	5	0	5
		C'être par...que+verbe+sujet	3	1	2
		C'être sur...que+verbe+sujet	3	1	2
		C'être dans...que+verbe+sujet	2	0	2
		C'être à...que+verbe+sujet	5	2	3
	C'être autour...que+verbe+sujet	1	1	0	
	C'être vers...que+verbe+sujet	1	0	1	
	C'être là...que+verbe+sujet	1	1	0	
	Groupe adverbial mis en relief+ Verbe+Sujet	être-ce	8	3	5
		être-il	5	3	2
devoir-il		3	3	0	
avoir-il		2	1	1	
arriver-on		1	0	1	
Pouvoir-on		1	0	1	
Concessif					

Tableau 4.3 Modélisation des inversions du sujet dans les romans de Hugo.

Nous pouvons constater qu'il existe des patrons qui apparaissent uniquement dans les quatre derniers romans de Victor Hugo, et d'autres qui n'apparaissent que dans les cinq premiers romans. Nous présentons ces patrons spécifiques à une certaine période dans le tableau suivant :

patrons spécifiques pour les 5 premiers romans	patrons spécifiques pour les 4 derniers romans
entre+verbe+sujet	dans+verbe+sujet
au-delà+verbe+sujet	en dehors de+verbe+sujet
derrière+verbe+sujet	en deçà de+verbe+sujet
au-dessous+verbe+sujet	sous+verbe+sujet
au milieu+verbe+sujet	après+verbe+sujet
puis+verbe+sujet	là+verbe+sujet
C'être autour...que+verbe+sujet	dedans+verbe+sujet
C'être là...que+verbe+sujet	au fond+verbe+sujet
devoir-il	C'être de...que+verbe+sujet
	C'être dans...que+verbe+sujet
	C'être vers...que+verbe+sujet
	arriver-on
	Pouvoir-on

Tableau 4.4 Patrons d'inversion spécifiques pour les deux périodes de publication de Hugo.

Dans ce qui suit, nous nous concentrons principalement sur les quatre derniers romans d'Hugo vu que, selon les critiques littéraires, c'est dans les romans d'après l'exil que se manifeste le plus souvent le style « hétérogène » de l'auteur.

- **Emploi de multiples déclencheurs d'inversion.**

Une des caractéristiques des inversions stylistiques de Victor Hugo se manifeste par l'emploi de multiples déclencheurs. Par exemple, pour le patron « dans+verbe+sujet », spécifique pour les quatre derniers romans :

29. *En bas, dans le fer à cheval au pied de la tribune, se tenaient les huissiers.*
(Quatrevingt-treize)

Ici, trois circonstanciels de lieu sont utilisés pour mettre l'accent sur l'endroit précis où « se tenaient les huissiers ». Les trois circonstanciels et la construction inversée créent un rythme équilibré, un type de versification : le premier circonstanciel (« en bas ») est suivi de trois hexasyllabes²³⁷ (« dans le fer à cheval », « au pied de la tribune », « se tenaient les huissiers »). Ou encore :

30. *C'est de cette façon, c'est par le hasard d'un mot d'orateur mal compris qu'est tombée la tête de madame Élisabeth.* (Quatrevingt-treize)

Dans cet exemple, le deuxième circonstanciel de manière est ajouté pour prolonger « cette façon ». L'effet d'insistance est créé par l'attention mise sur l'explication détaillée de la « façon » non seulement par le deuxième circonstanciel explicatif, mais aussi par l'emploi du clivage. Le rythme syllabique de la phrase met encore davantage l'accent sur l'unité la plus courte du circonstanciel explicatif : « mal compris ».

L'empilement de plusieurs déclencheurs est très courant dans les patrons que nous extrayons du corpus de Victor Hugo : c'est un des phénomènes les plus caractéristiques que nous repérons en analysant les inversions de l'auteur. Nous constatons, d'après les exemples ci-dessus, que ce phénomène consiste à renforcer l'accent de phrase, déjà mis sur le déclencheur par l'antéposition. Cet effet d'insistance est créé non seulement par les précisions ou explications sémantiques qu'apportent les déclencheurs supplémentaires, mais aussi par le truchement de la réorganisation du rythme de la phrase. Ce phénomène, qui consiste à mettre en place une suite de

²³⁷ « Se dit d'un vers qui a six syllables ». <https://www.larousse.fr/dictionnaires/francais/hexasyllabe/186432>, consulté le 30 juillet, 2021.

circonstanciels/déclencheurs, peut être considéré comme une manifestation du style de Hugo, ici marqué par l'accumulation.

- **La combinaison de plusieurs figures de style dans une phrase.**

La deuxième caractéristique que nous repérons en analysant les inversions du sujet chez Hugo est la combinaison de plusieurs figures de style dans une seule phrase. C'est, par exemple, le cas des 17 occurrences d'inversions du sujet déclenchées par une construction clivée.

*31. C'est de cette étoffe-là, disait-il, **qu'était drapé** le lit de la duchesse d'Anville à La Roche-Guyon. (Les Misérables)*

Il y a souvent un double processus de mise en relief du déclencheur : une inversion qui met en relief le syntagme antéposé ; et un clivage pour renforcer la mise en relief. À part ces 17 occurrences, d'autres phénomènes de combinaison de figures de style sont aussi constatés dans les résultats, par exemple :

*32. « [...] et je vous le soutiens à tous, qui que vous soyez, **fussiez-vous publicistes, fussiez-vous économistes, fussiez-vous légistes, fussiez-vous plus connaisseurs en liberté, en égalité et en fraternité que le couperet de la guillotine !** » (Les Misérables)*

On a ici affaire à une inversion concessive, visant à renforcer l'idée que l'orateur de ce discours soutient, c'est-à-dire, tient tête, « à tous ». Cette inversion est combinée avec le parallélisme : les structures inversées dans l'exemple (« fussiez-vous COD »)

permettent d'énumérer les différentes catégories de personnes visées. Cette explication « énumérative », marquée par la répétition, crée un effet de redondance dans le texte.

L'exemple suivant montre la combinaison de l'inversion du sujet avec la métaphore :

33. Dans son impassibilité, peut-être seulement apparente, étaient empreintes les deux pétrifications, la pétrification du cœur, propre au bourreau, et la pétrification de l'esprit, propre au mandarin. (L'homme qui rit)

« La pétrification du cœur » et la « pétrification de l'esprit » sont deux exemples de métaphore hugolienne, qui relie souvent des notions abstraites avec des noms concrets. Nous avons discuté de cette caractéristique stylistique de Hugo dans la section 4.2. Ces métaphores permettent de « rendre sensible par une image des idées²³⁸ » (Hugo, 1985 [1827] p. 16). Les deux expressions qui se réfèrent à la pétrification dans cet exemple donnent une image sensible de « l'impassibilité ».

Le dernier exemple montre la combinaison de l'inversion avec la paronomase (procédé qui consiste à rapprocher, à l'intérieur d'une phrase ou d'un vers, des paronymes dans un but stylistique²³⁹) :

34. Au-dessous se tenait le greffier, griffonnant. (Notre-Dame de Paris)

Les deux mots « greffier » et « griffonnant » ont des prononciations similaires, tout en étant sémantiquement différents. L'ajout du gérondif « griffonnant » provoque non seulement une similitude sonore, qui donne un effet d'insistance sur le sujet de la phrase

²³⁸ Hugo, Victor. 1985 [1827]. « Préface de Cromwell » dans *Œuvres complètes*. Édité par Guy Rosa, Jacques Seebacher, et Groupe Hugo. 16 vol. Paris, France : R. Laffont, p. 16.

²³⁹ « Paronomase », Larousse, <https://www.larousse.fr/dictionnaires/francais/paronomase/58290>, consulté le 25 juin, 2021.

« greffier », mais il permet aussi d'équilibrer le rythme de la phrase, la phrase est coupée pour former quatre trisyllabes²⁴⁰ (3-3-3-3).

● Les inversions concessives

L'utilisation des inversions concessives est également une caractéristique stylistique de Hugo : nous relevons 16 occurrences dans les neuf romans de Hugo, dont 14 dans *Les Misérables* et *Notre-Dame de Paris*. C'est un phénomène caractéristique parce que nous ne trouvons par exemple aucune occurrence dans les romans de Chateaubriand, et une seule occurrence dans les romans de Flaubert. L'exemple suivant est une inversion concessive avec le verbe conjugué au subjonctif imparfait.

35. *Lorsqu'une religieuse est appelée au parloir, fût-ce la prieure, elle baisse son voile de façon, l'on s'en souvient, à ne laisser voir que sa bouche.* (Les Misérables)

L'emploi du subjonctif imparfait dans une proposition concessive est « très rare en français contemporain²⁴¹ et trahit toujours un niveau de langue archaïsant ou recherché²⁴² » (Morel, 1980, p. 111). Nous n'avons malheureusement pas trouvé d'étude systématique sur cette utilisation particulière du subjonctif imparfait par Hugo, ni sur les occurrences fréquentes de cette construction syntaxique dans les romans *Les*

²⁴⁰ « Se dit d'un mot, d'un vers composé de trois syllables ». <https://www.larousse.fr/dictionnaires/francais/trisyllabe/79821>, consulté le 30 juillet, 2021.

²⁴¹ Morel parle du côté archaïsant des inversions concessives par rapport aux écrivains contemporains, néanmoins, Hugo ne fait pas partie des écrivains contemporains, ainsi cette utilisation particulière d'inversion ne devrait pas être considérée comme surprenante pour Hugo. Cependant, dans notre corpus, nous constatons effectivement que ce phénomène, dit « archaïsant », n'apparaît pas souvent chez Chateaubriand et Flaubert, qui cherchent à innover le langage, alors qu'il apparaît relativement plus souvent chez Stendhal qui se réfère à la tradition littéraire et qui n'apprécie pas les innovations langagières (voir la section 5.2 du Chapitre 5 pour la discussion sur la notion d'innovation pour les quatre auteurs).

²⁴² Morel, Mary-Annick. 1980. « L' inversion du sujet dans certaines propositions concessives du français : Conditions d' emploi et valeur ». *DRLAV. Documentation et Recherche en Linguistique Allemande Vincennes* 22 (1): 105-12.

Misérables et *Notre-Dame de Paris*. Notre méthode supervisée facilite la détection de ce patron, négligé par les critiques littéraires sur le style de Hugo.

Les analyses quantitatives et qualitatives des inversions stylistiques dans les romans de Hugo révèlent plusieurs caractéristiques. Nous constatons que certaines caractéristiques sont conformes à ce que disent les critiques littéraires sur le style « hétérogène » de Hugo : le mélange des mots abstraits et concrets pour rendre une idée concrète ; les répétitions abondantes qui rendent le texte « épais²⁴³ » (Wulf, 2014, p. 171). Nous constatons également que notre méthode permet de trouver des caractéristiques qui ne sont pas discutées et étudiées par les critiques littéraires : l'emploi de plusieurs déclencheurs devant la construction inversée, la similitude entre *Les Misérables* et *Notre-Dame de Paris* en matière d'utilisation des inversions stylistiques ou encore, le nombre important d'inversions concessives dans ces deux romans, etc.

Cette étude sur les inversions stylistiques hugoliennes nous ramène à la discussion sur les propriétés de la stylométrie. Les outils informatiques apportent de nouvelles possibilités pour détecter des phénomènes stylistiques qui étaient relativement difficiles à repérer manuellement, ces phénomènes stylistiques servent non seulement à confirmer (ou, éventuellement, contredire) les critiques littéraires, mais aussi à les compléter par la mise au jour de nouveaux exemples caractéristiques.

En conclusion, on peut dire que l'étude des inversions du sujet dans la prose de Hugo révèle d'une part des tendances statistiques concernant l'utilisation de différents types d'inversions du sujet dans les romans hugoliens ; d'autre part, des caractéristiques qualitatives, telles que la cooccurrence de l'inversion avec la métaphore hugolienne, la répétition ou l'accumulation divergente, qui confirment ce que les critiques littéraires ont appelé le style « hétérogène » de Hugo. Nous admettons en même temps que l'inversion stylistique du sujet, hors de son contexte, est peu informative et peu

²⁴³ Wulf, Judith. 2014. *Op.cit.*, p. 171.

caractéristique du style de Hugo, d'autant que les fréquences observées sont peu importantes.

4.4.2 Le clivage

Dans cette partie, nous appliquons notre méthode d'extraction supervisée de structures clivées (dites clivages) sur le corpus Victor Hugo. Nous commençons par étudier la typologie des clivages dans les neuf romans de Hugo ; nous essayons ensuite de montrer l'évolution de l'utilisation de cette figure de style à travers les différences statistiques dans le nombre d'occurrences constatées dans les romans d'avant et d'après l'exil. Nous examinons ensuite de manière détaillée les occurrences de cette figure de style dans le texte original afin de déterminer son apport sur le plan stylistique. Cette partie est structurée de la même façon que celle consacrée aux inversions du sujet, en combinant les analyses quantitatives et les analyses qualitatives afin de révéler le style « hétérogène » de Hugo.

4.4.2.1 La définition et les fonctions stylistiques du clivage

Le clivage est un phénomène stylistique qui consiste à « accorder à un certain constituant de la phrase plus d'importance qu'aux autres^{244,245} » (Dessaintes, 1960) (Müller-Hauser, 1943). Il est une des manières syntaxiques de mettre en relief une partie de la phrase afin de susciter l'attention du lecteur. Nous avons discuté dans la section 3.6.2 (Chapitre 3) les deux procédés syntaxiques de mise en relief ainsi que les difficultés que nous avons rencontrées à extraire les phénomènes de clivage dans notre corpus. Nous présentons dans cette partie les résultats après vérification manuelle.

²⁴⁴ Dessaintes, Maurice. 1960. *Éléments de linguistique descriptive en fonction de l'enseignement du français*. Namur : La Procure.

²⁴⁵ Müller-Hauser, M.-L. 1943. *La mise en relief d'une idée en français moderne*. Genève.

4.4.2.2 Analyses quantitatives des clivages

Nous avons trouvé 868 occurrences de clivages dans neuf romans de Victor Hugo, dont 267 phrases avec la construction « c'est X que Y²⁴⁶ » et 601 phrases avec la construction « C'est X qui Y ». La dernière construction met en relief le sujet de la proposition alors que la première met en relief tous les autres cas.

- « c'est X que Y »

La modélisation des clivages dans nos résultats se fonde sur la typologie des syntagmes clivés. Ainsi, pour la construction « c'est X que Y », six grandes catégories ont été repérées dans les neuf romans de Hugo :

1. C'est+Groupe Nominal+que

36. *Mais c'est ma chambre qu'il faut voir.* (Les Misérables)

2. C'est+Groupe Prépositionnel+que

37. *Tout en convenant que ce n'était point pour une fonction politique que Dieu avait créé monseigneur Bienvenu, nous eussions compris et admiré la protestation au nom du droit et de la liberté, l'opposition fière, la résistance périlleuse et juste à Napoléon tout-puissant.* (Les Misérables)

3. C'est +Groupe Pronominal+que

²⁴⁶ La construction « C'est X que Y » que nous présentons ici comprend également la forme négative (« ce n'est pas X que Y »), la forme inversée (« est-ce X que Y ») et la forme négative inversée (« N'est-ce pas X que Y »). Pareil pour la construction « C'est X qui Y ».

38. *Ce n'est pas cela que je veux, tas de canailles !* (Les Misérables)

4. C'est +Groupe Adverbial+que

39. *C'est là que le piquet qui attendait s'est rallié à l'escorte.* (Le Dernier Jour d'un condamné)

5. C'est +Nom Propre+que

40. *Il est probable que c'est Gwynplaine qu'on a enterré, mais ce n'est pas certain.*
(L'Homme qui rit)

6. C'est+Locution Conjonctive+que

41. *C'est parce qu'ils ont le nid qu'ils ont le chant.* (Les Misérables)

Le tableau suivant montre les fréquences absolues pour la construction « C'est X que Y » dans les romans de Victor Hugo :

romans	Nominal	Prépositionnel	Pronominal	Adverbial	Nom Propre	Locution Conjonctive	TOTAL
Bug-Jargal		6	1			1	8
Han d'Islande	1	32		3	1	1	38
Le Dernier Jour d'un condamné	1	2		3			6
Notre-Dame de Paris	6	19	1	3			29
Claude Gueux	1						1
Les Misérables	22	56	2	13		4	97
Les Travailleurs de la mer	4	22	1				27
L'Homme qui rit	7	30		3	2		42
Quatrevingt-treize	5	12	1	1			19

Tableau 4.5 Fréquences absolues de « C'est X que Y » dans les romans de Hugo.

Pour éliminer l'effet de la longueur variable des textes sur les fréquences absolues, nous divisons ensuite chaque fréquence d'une catégorie par le nombre total de

fréquences dans un roman. Nous présentons dans le tableau suivant les fréquences relatives d'occurrences :

romans	Nominal	Prépositionnel	Pronominal	Adverbial	Nom Propre	Locution Conjonctive
Bug-Jargal	0.000%	75.000%	12.500%	0.000%	0.000%	12.500%
Han d'Islande	2.632%	84.211%	0.000%	7.895%	2.632%	2.632%
Le Dernier Jour d'un condamné	16.667%	33.333%	0.000%	50.000%	0.000%	0.000%
Notre-Dame de Paris	20.690%	65.517%	3.448%	10.345%	0.000%	0.000%
Claude Gueux	100.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Les Misérables	22.680%	57.732%	2.062%	13.402%	0.000%	4.124%
Les Travailleurs de la mer	14.815%	81.481%	3.704%	0.000%	0.000%	0.000%
L'Homme qui rit	16.667%	71.429%	0.000%	7.143%	4.762%	0.000%
Quatrevingt-treize	26.316%	63.158%	5.263%	5.263%	0.000%	0.000%

Tableau 4.6 Fréquences relatives de « C'est X que Y » dans les romans de Hugo.

La figure suivante présente la visualisation des données statistiques du tableau ci-dessus :

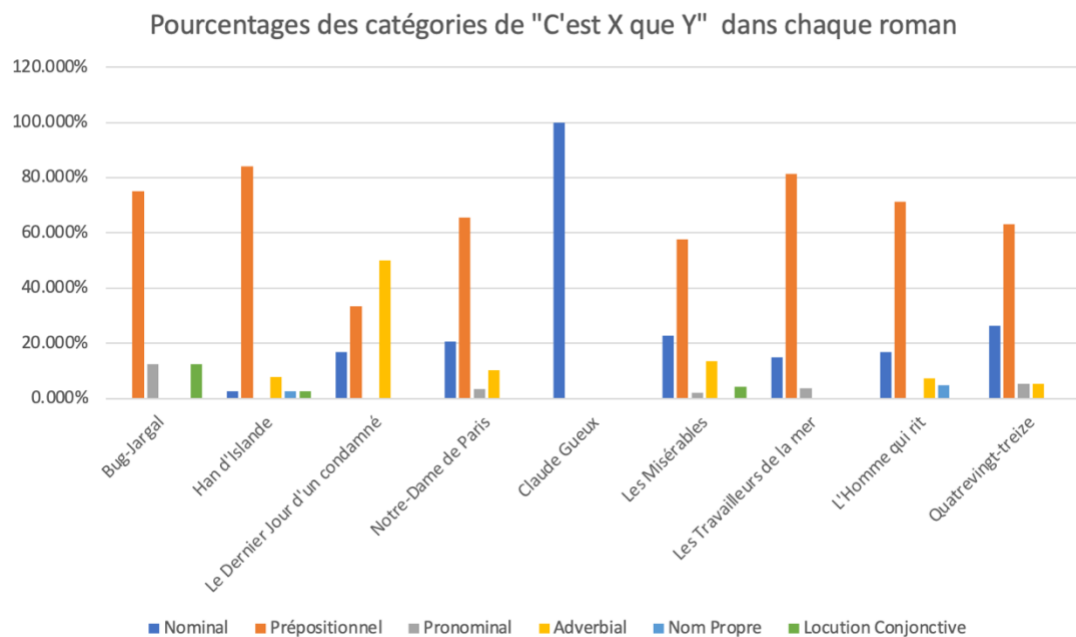


Figure 4.3 Pourcentages de différentes catégories de « C'est X que Y » dans les romans de Hugo.

En examinant les résultats des fréquences absolues et relatives de la construction « C'est X que Y » dans les romans de Hugo, nous pouvons constater les caractéristiques suivantes :

1. Le clivage n'est pas très développé dans le premier roman de Hugo : *Bug-Jargal*, dont la fréquence inférieure ne s'explique pas par la longueur du texte. Ce constat est également confirmé par la diversité relativement réduite des catégories de clivages dans ce roman.

2. La ressemblance que nous avons constatée dans notre étude sur les inversions se présente également dans les résultats de clivages. *Notre-Dame de Paris* et *Les Misérables* présentent les similitudes concernant les pourcentages d'occurrences des différentes catégories de clivages.

● « c'est X qui Y »

La modélisation de la construction « c'est X qui Y » est relativement simple parce que le X dans cette construction représente le sujet de la proposition (ainsi, la typologie du X n'est pas variée). Trois catégories de structures morphosyntaxiques sont présentes dans nos résultats pour X :

1. C'est+Groupe Nominal+qui

42. *C'était cette clarté qui* avait d'abord ébloui Gwynplaine. (L'Homme qui rit)

2. C'est+Groupe Pronominal+qui

43. Il dit à l'enfant : Est-ce que *c'est toi qui* m'apportes la lettre que j'attends ?
(Les Misérables)

3. C'est+Non Propre+qui

44. C'est *Homo* qui va être étonné ! (L'Homme qui rit)

Le tableau suivant présente les fréquences absolues de chaque catégorie de « c'est X qui Y » dans les romans de Hugo.

romans	Nominal	Pronominal	Nom Propre	TOTAL
Bug-Jargal	13	22	1	36
Han d'Islande	28	39	6	73
Le Dernier Jour d'un condamné	4	5		9
Notre-Dame de Paris	54	33	9	96
Claude Gueux	1	1		2
Les Misérables	87	84	32	203
Les Travailleurs de la mer	22	19	4	45
L'Homme qui rit	33	36	7	76
Quatrevingt-treize	28	18	15	61

Tableau 4.7 Fréquences absolues de « C'est X qui Y » dans les romans de Hugo.

Les valeurs relatives sont calculées et visualisées par la figure suivante :

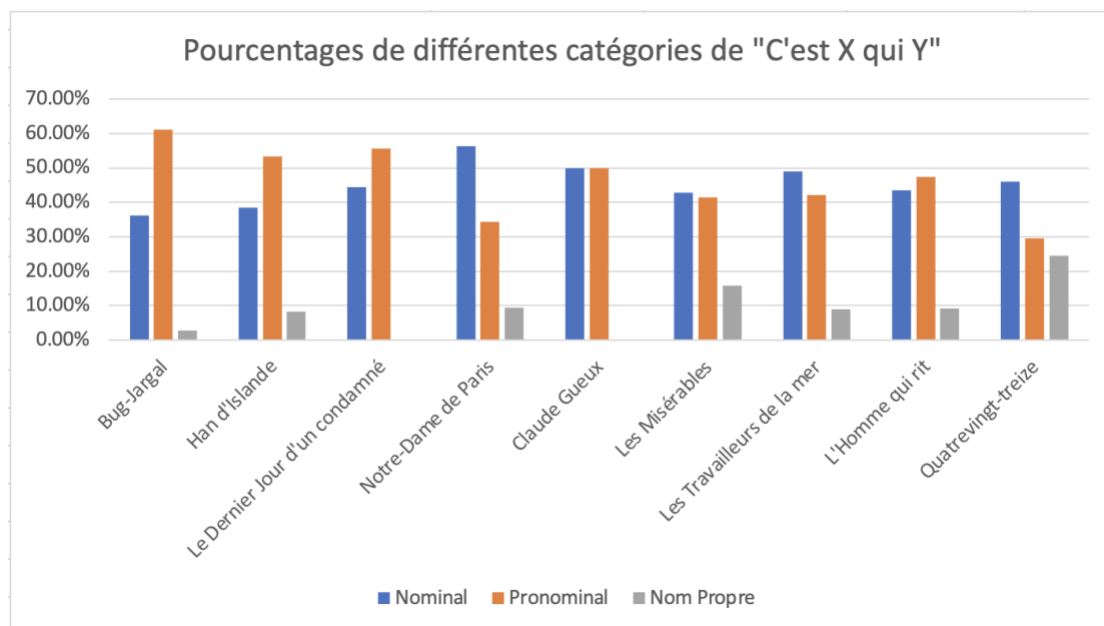


Figure 4.4 Pourcentages des différentes catégories de « C'est X qui Y » dans les romans de Hugo.

Pour la construction « C'est X qui Y », nous constatons d'abord que dans les trois premiers romans, c'est la catégorie « C'est+Groupe Pronominal+qui » qui est la plus fréquente dans les textes alors que dans les derniers romans de Hugo, l'importance statistique de cette catégorie est relativement réduite et remplacée par les deux autres catégories. De plus, l'utilisation de la catégorie « C'est+Nom Propre+qui » devient systématique dans les romans publiés après *Claude Gueux*.

En raison du manque de spécificité statistique pour la construction « C'est X qui Y », mais aussi du manque de diversité du sujet clivé, nous allons nous concentrer sur la construction « C'est X que Y » pour les analyses qualitatives dans la partie suivante.

4.4.2.3 Analyses qualitatives des clivages

Les analyses qualitatives des clivages commencent par la modélisation en détail des syntagmes clivés dans nos résultats. Nous avons modélisé l'élément X dans toutes les occurrences de la catégorie « C'est X que Y ». Le tableau suivant montre les fréquences respectives de chaque patron dans les cinq premiers romans et les quatre derniers romans de Victor Hugo. Notons que, dans le tableau, les groupes prépositionnels clivés se présentent uniquement par la préposition elle-même, par exemple, le « à » représente tous les patrons avec la structure « C'est à + X que Y » :

Patrons	fréquences dans les 5 premiers romans	fréquences dans les 4 derniers romans	Total
à	18	27	45
NOUN	9	32	41
de	9	27	36
dans	7	15	22
pour	6	13	19
par	2	16	18
là	3	9	12
avec	2	8	10
sur	3	5	8
ainsi	1	5	6
sans	6	0	6
vers	3	2	5
en	2	2	4
devant	1	2	3
parce que	2	1	3
sous	1	2	3
cela	1	1	2
celui/celle	1	1	2
maintenant	2	0	2
comme	0	2	2
PROP	0	2	2
aujourd'hui	1	0	1
autour	1	0	1
sur	1	0	1
assez	0	1	1
avant	0	1	1
d'après	0	1	1
quand	0	1	1
entre	0	1	1
PRON	0	1	1
pendant	0	1	1
vrai	0	1	1
rien	0	1	1

Tableau 4.8 Modélisation et fréquences des patrons pour la construction « C'est X que Y ».

Nous sélectionnons ensuite les patrons qui n'apparaissent que dans une des périodes de publication. Nous les définissons comme les patrons spécifiques pour cette période :

Patrons spécifiques pour les 5 premiers romans	Patrons spécifiques pour les 4 derniers romans
sans	comme
maintenant	PROPN
aujourd'hui	assez
autour	avant
sur	d'après
	quand
	entre
	PRON
	pendant
	vrai
	rien

Tableau 4.9 Patrons spécifiques de la construction « C'est X que Y » pour les deux périodes de publication.

Le retour au texte nous révèle les caractéristiques suivantes, d'après notre analyse stylistique de ces patrons :

-Le jeu du suspense

Le jeu du suspense est une des caractéristiques les plus fréquentes dans nos résultats. Cette technique vise à augmenter l'effet de suspense au niveau de la phrase par la mise en attente du prédicat au sein de Y (le verbe et son sujet), par l'introduction de longs adverbiaux avant le verbe principal (on parle ici de suspense car le lecteur attend le prédicat, qui tarde à se dévoiler du fait du nombre d'éléments le précédant). C'est le cas dans l'exemple suivant avec le patron « c'est sans X que Y » :

45. Ce ne fut pas sans avoir plus d'une fois trébuché sur des pierres roulantes, que l'imagination alarmée du vieillard prenait pour des os humains épars sur les degrés, qu'ils arrivèrent au premier étage de l'édifice, dans une salle ronde pareille à la salle inférieure. (Han d'Islande)

La partie clivée est déjà une construction longue, mais le sujet (« ils ») et le verbe (« arrivèrent ») n'apparaissent qu'après une proposition de subordination relative décrivant les qualités des « pierres roulantes » (« que l'imagination alarmée du vieillard prenait pour des os humains épars sur les degrés »). L'action (« arriver ») et sa manière (« sans avoir... ») sont éloignées par cette subordination relative pour créer un effet de retard et de suspense, quand l'auteur présente enfin l'action elle-même, après la description très détaillée de la manière.

L'exemple suivant montre le même phénomène, mais avec un agencement syntaxique différent, le patron en question est « C'est ADV que Y » :

46. C'est ainsi qu'un jour, à Hampton-Court, dans le jardin, étant derrière la duchesse, qui était derrière la reine, il entendit Anne, se conformant lourdement à la mode, émettre des sentences. (L'Homme qui rit)

Dans cette phrase, au lieu de rallonger la partie clivée pour éloigner le sujet et la partie mise en relief par le clivage, Hugo choisit d'accumuler quatre circonstanciels de lieu (« à Hampton-Court, dans le jardin, étant derrière la duchesse, qui était derrière la reine ») entre la construction clivée et le sujet de la phrase. L'effet de suspense est d'autant plus renforcé que l'auteur insère un participe en apposition (« se conformant lourdement à la mode ») entre le verbe et le complément.

-L'accumulation

Les figures de répétition sont très présentes dans nos résultats :

47. C'est de cet être adorable, c'est de ce cœur, c'est de cette adoption, c'est de cette tendresse, c'est de ce divin regard aveugle, le seul qui le vît sur la terre, qu'il s'était éloigné ! (L'Homme qui rit)

La douleur de la séparation est amplifiée dans cette phrase par l'énumération des qualités que possède l'objet dont le personnage se sépare. Les propositions répétitives mises devant le verbe (« s'éloigner ») créent un effet d'insistance avec les constructions déjà en clivage. L'exemple suivant montre le même phénomène, avec cet effet d'insistance créé à la fois par la répétition et le clivage :

48. C'était à cause de cette déité, de cette fée des nacres, de cette reine des souffles, de cette grâce née des flots, c'était à cause d'elle, on se le figurait du moins, que le souterrain était religieusement muré, afin que rien ne pût jamais troubler, autour de ce divin fantôme, l'obscurité qui est un respect, et le silence qui est une majesté.
(Les Travailleurs de la mer)

Les répétitions peuvent aussi apparaître dans la proposition principale :

49. C'est pour les beaux yeux du papa que tu es allé aux Douvres, que tu as eu froid, que tu as eu chaud, que tu as crevé de faim et de soif, que tu as mangé de la vermine de rocher, que tu as eu le brouillard, la pluie et le vent pour chambre à coucher, et que tu as exécuté la chose de me rapporter ma machine, comme on rapporte à une jolie femme son serin qui s'est échappé ! (Les Travailleurs de la mer)

Dans cette phrase, la souffrance pour « rapporter [la] machine » a été mise en relief par une suite d'énumérations ; une comparaison est ajoutée à la fin de la phrase (« comme on rapporte à une jolie femme son serin qui s'est échappé ! ») pour renforcer ce geste grandiose, mais difficile.

Nos analyses des répétitions pour le patron « C'est X que Y » confirment ce que disent les critiques littéraires sur le style hétérogène de Hugo : les répétitions sont fréquentes et abondantes dans les phrases, il s'agit souvent de répétitions de plusieurs

syntagmes ou propositions, soit énumératives, soit accumulatives, pour créer l'effet d'insistance.

-La métaphore

Nous avons évoqué dans la partie 4.2 que la métaphore est une caractéristique stylistique hugolienne qui consiste à rapprocher des éléments hétérogènes. Dans nos résultats pour le clivage, plusieurs occurrences de ce phénomène ont été trouvées, voici trois exemples de cette figure de style avec clivage :

50. *C'est avec des gouttes que l'océan vous broie.* (L'Homme qui rit)

« Les gouttes » dans cette phrase sont mises en relief par le clivage afin de présenter la manière dont l'océan « broie ». Le rapprochement de « l'océan » et de « broyer » le concrétise, ce qui est encore renforcé par les « gouttes » d'eau, un nom concret.

51. *C'est de la physionomie des années que se compose la figure des siècles.* (Les Misérables)

Ce dernier exemple compare les qualités des « années » et des « siècles » à celles des personnes. Une comparaison employée pour rendre les qualités des années et des siècles plus sensibles et compréhensibles aux lecteurs.

52. *C'est de ces trois vapeurs, bière, eau-de-vie, absinthe, qu'est fait le plomb de l'âme.* (Les Misérables)

Dans cette phrase, « le plomb de l'âme » crée un effet de surprise et d'exagération en étant combiné avec les « trois vapeurs » alcoolisées de la construction clivée. La

notion d'« âme » est concrétisée par le « plomb », la métaphore est ensuite détaillée par le terme « les trois vapeurs ».

La combinaison de mots abstraits avec de mots concrets dans les métaphores est une caractéristique importante du style de Victor Hugo : elle vise à rendre les idées abstraites plus sensibles. Cette caractéristique stylistique de Hugo est bien repérée par notre méthode d'analyse.

-L'antithèse

L'antithèse est une figure de style qui rapproche deux termes ou deux idées opposées. D'après l'ouvrage de Wulf²⁴⁷ (2014), mais aussi celui de Brombert²⁴⁸ (1985), l'antithèse est une caractéristique importante du style de Hugo. Dans nos résultats de clivages, nous constatons également un nombre important d'antithèses en combinaison avec le clivage.

*53. C'est là qu'a été dit, par une petite abandonnée, enfant trouvé que le couvent élevait par charité, **ce mot doux et navrant.** (Les Misérables)*

L'apparition du sujet de cette phrase (« ce mot ») est retardée par l'insertion d'un circonstanciel de manière et une apposition nominale assez longue. Le suspense qui se maintient jusqu'à l'apparition du sujet est renforcé ensuite par une antithèse d'adjectif qui rend le « mot » encore plus mystérieux.

*54. Même quand ils avortent, ils sont vénérables, et **c'est peut-être dans l'insuccès qu'ils ont plus de majesté.** (Les Misérables)*

²⁴⁷ Wulf, Judith. 2014. *Op. cit.*.

²⁴⁸ Brombert, Victor, et Héloïse Neefs. 1985. *Op.cit.*.

« L'insuccès » dans la construction clivée est opposé à la « majesté » dans la proposition principale, le fait de mettre « dans l'insuccès » en relief par le clivage amplifie l'effet d'opposition créé par l'antithèse.

L'analyse qualitative nous révèle des caractéristiques stylistiques qui complètent les observations statistiques. La combinaison de ces caractéristiques nous donne une idée générale du style de Hugo, ici autour de la figure de clivage. Les résultats de l'analyse qualitative sont conformes dans l'ensemble avec ce que disent les critiques littéraires du style de Hugo. Les occurrences de l'antithèse, de la métaphore et de la répétition qu'on a analysées peuvent être considérées comme les manifestations textuelles et formelles du style « hétérogène » de l'auteur. Nous relevons aussi la mise en scène du suspense dans les clivages, un trait qui, à notre connaissance, n'avait pas encore attiré l'attention des stylisticiens. Nos résultats concernant l'analyse quantitative sont à mettre en avant car les stylisticiens procèdent rarement à un relevé exhaustif des représentations d'un phénomène formel. Comme l'étude sur l'inversion du sujet, le clivage, hors contexte, n'est pas assez informatif pour caractériser à lui seul un style littéraire, mais le retour au texte révèle des caractéristiques qualitatives (comme les phénomènes antithétiques et métaphoriques) intéressantes. C'est, de notre point de vue, ainsi que notre recherche peut apporter de nouvelles perspectives pour la caractérisation du style littéraire, en fournissant des résultats à la fois quantitatifs (sur le patron lui-même) et qualitatifs (par le retour au texte).

4.4.3 La répétition

Nous abordons dans cette partie la dernière application de notre méthode d'extraction supervisée sur le corpus Victor Hugo. L'extraction des unités syntaxiques répétitives, basée non seulement sur l'identification de structures morpho-syntaxiquement identiques, mais également sur des relations de dépendance identiques, nous permet de repérer d'une manière précise et efficace cette figure caractérisante dans les textes. Nous présentons dans cette partie l'analyse des occurrences des répétitions dans les neuf romans de Hugo, la typologie des patrons répétés ainsi que l'analyse qualitative des phrases comprenant des répétitions. L'objectif de cette étude consiste non seulement à étudier les caractéristiques statistiques de cette figure de style dans les romans de Victor Hugo, mais aussi à retrouver des caractéristiques stylistiques pouvant confirmer ou compléter les études précédentes sur le style de Hugo.

4.4.3.1 Analyses quantitatives des répétitions syntaxiques

Les répétitions sont extraites par notre méthode d'extraction supervisée, en appliquant les profondeurs d'extraction de 3 et de 4 (pour les discussions sur les profondeurs d'extraction, voir la section 3.7.2 au Chapitre 3). Les résultats issus des deux profondeurs d'extraction sont combinés dans un seul fichier pour chaque roman, parce qu'il n'y a pas lieu de faire de distinction ici : les patrons peuvent être analysés de la même manière, sans discrimination.

Nous avons trouvé 1204 phrases comprenant les répétitions sémantiquement interprétables dans le corpus Victor Hugo, les fréquences pour chaque roman sont présentées en valeurs absolues dans le tableau suivant :

Romans	Fréquences
Bug-Jargal	14
Han d'Islande	71
Le Dernier Jour d'un condamné	12
Notre-Dame de Paris	114
Claude Gueux	5
Les Misérables	486
Les Travailleurs de la mer	121
L'Homme qui rit	220
Quatrevingt-treize	161

Tableau 4.10 Fréquences absolues des répétitions dans les romans de Hugo.

Pour éliminer l'influence de la longueur variable des textes sur les fréquences absolues, nous divisons ces valeurs par le nombre de phrases total pour chaque roman. Les pourcentages de répétitions par rapport à la totalité des phrases dans chaque roman sont présentés dans le tableau suivant :

Romans	Fréquences
Bug-Jargal	0.3881%
Han d'Islande	0.8689%
Le Dernier Jour d'un condamné	0.5571%
Notre-Dame de Paris	0.8912%
Claude Gueux	0.8375%
Les Misérables	1.2824%
Les Travailleurs de la mer	1.1024%
L'Homme qui rit	1.3400%
Quatrevingt-treize	1.6439%

Tableau 4.11 Fréquences relatives des répétitions dans les romans de Hugo.

On voit une augmentation évidente de l'utilisation de cette figure de style dans les romans de Hugo au cours du temps. Pour mieux visualiser cette tendance, nous présentons les résultats statistiques dans la figure suivante :

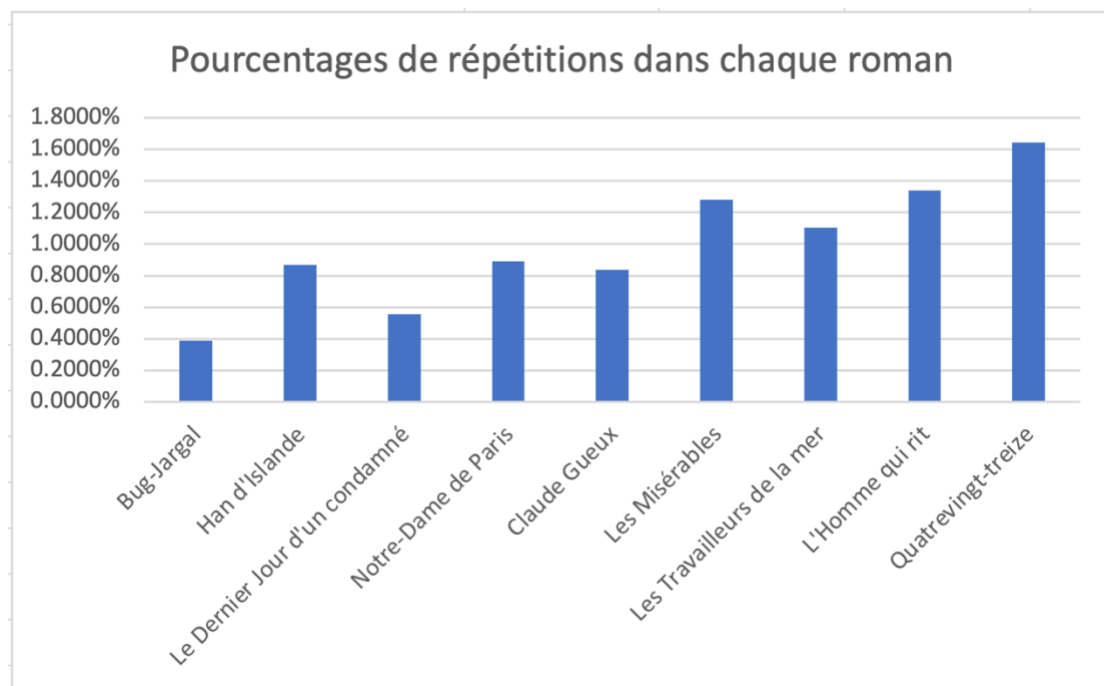


Figure 4.5 Pourcentages des répétitions dans les romans de Hugo.

En combinant les résultats des fréquences absolues et des fréquences relatives, nous constatons les caractéristiques suivantes pour la répétition :

1. Une augmentation chronologique des répétitions est constatée dans les romans de Hugo (d'après notre analyse, plus de 1% des phrases des quatre derniers romans, à savoir les romans publiés pendant et après l'exil de l'auteur, contiennent des répétitions, alors que le score est toujours inférieur à 1% pour les cinq premiers romans de Hugo).

2. Le premier roman de Hugo, *Bug-Jargal*, présente le moins d'occurrences de cette figure de style. Ce phénomène a été également constaté dans les analyses quantitatives des inversions et des clivages, présentées dans les deux parties précédentes. D'après les analyses statistiques des trois patrons caractérisants de notre corpus, *Bug-Jargal* est le roman qui a la densité de figures la plus faible (par rapport aux autres romans du corpus).

Les caractéristiques statistiques sont conformes à ce que disent les critiques littéraires sur le style de Hugo après l'exil : les figures de répétitions sont plus présentes dans les quatre derniers romans que les cinq publiés avant l'exil. Elles révèlent

également une autre caractéristique qui n'est pas discutée systématiquement par les critiques littéraires : la densité faible de figures de style dans *Bug-Jargal*.

4.4.3.2 Analyses qualitatives des répétitions

Dans cette partie de l'analyse, nous examinons les patrons en répétition en détail. Pour montrer la différence entre les romans d'avant et d'après l'exil, nous divisons le corpus en deux sous-corpus contenant respectivement les cinq premiers romans de Hugo et les quatre derniers. Le tableau suivant montre les patrons de répétition les plus fréquents dans les cinq premiers romans de Hugo :

Patrons	Fréquences
ADJ et ADJ	29
le NN de le NN	18
de NN et de NN	8
de le NN de le NN	7
de le NN de NP	7
le NN de NP	6
à le NN de le NN	6

Tableau 4.12 Patrons de répétition les plus fréquents dans les cinq premiers romans de Hugo.

Les patrons en répétition les plus fréquents pour les quatre derniers romans de Victor Hugo sont présentés dans le tableau suivant :

Patrons	Fréquences
le NN de le NN	74
, VB le NN	35
ADJ et ADJ	31
le NN de NP	26
, le NN de le NN	21
de le NN de NP	17
le NN de NN	16

Tableau 4.13 Patrons de répétition les plus fréquents dans les quatre derniers romans de Hugo.

La répétition est souvent reconnue pour ses effets stylistiques d'insistance et d'harmonie. Il ne nous semble pas nécessaire d'analyser exhaustivement toutes les phrases dans nos résultats afin de démontrer la fonction de mise en relief. Nous essayons dans cette partie d'étudier quelques patrons spécifiques pour chacune des deux périodes de publication, afin de trouver d'autres caractéristiques stylistiques liées à la répétition.

-Les répétitions en cascade

Les répétitions abondantes dans une phrase, engageant souvent plusieurs structures en parallèle sont très présentes dans nos résultats. Par exemple, pour le patron spécifique des quatre derniers romans de Hugo, « , VB le NN » :

55. La femme s'occupe alors, se passionne au maniement des espèces, s'y vert-de-grise les doigts, entreprend l'élève des métayers et le dressage des fermiers,

convoque les avoués, préside les notaires, harangue les tabellions, visite les robins, suit les procès, rédige les baux, dicte les contrats, se sent souveraine, vend, achète, règle, jordonne, promet et compromet, lie et résilie, cède, concède et rétrocède, arrange, dérange, thésaurise, prodigue, elle fait des sottises, bonheur magistral et personnel, et cela console. (Les Misérables)

Il s'agit d'une séquence d'actions, présentées par un patron assez court « , VB le NN » (« , convoque les avoués, préside les notaires, harangue les tabellions, visite les robins, suit les procès, rédige les baux, dicte les contrats »). L'agencement successif de ces structures répétitives et brèves de 4 ou 5 syllabes crée une intensité rythmique autour des tâches de « la femme ».

56. Il fondait des hôpitaux, ouvrait des écoles, visitait les malades, dotait les filles, soutenait les veuves, adoptait les orphelins ; (Les Misérables)

Dans cet exemple, l'auteur raconte d'une manière efficace et rythmique une succession d'actions grâce à plusieurs structures parallèles. L'exemple suivant montre le même phénomène avec le patron « , VB le NN », mais les répétitions dans cette phrase consistent à mettre en évidence une énumération des difficultés que le sujet « avait vaincues » :

57. Il avait vaincu l'isolement, vaincu la faim, vaincu la soif, vaincu le froid, vaincu la fièvre, vaincu le travail, vaincu le sommeil. (Les Travailleurs de la mer)

-La métaphore

La métaphore hugolienne, qui consiste souvent à relier des mots abstraits avec des

mots concrets, a été discutée en détail dans les deux précédentes sections. Ce phénomène est également présent avec la figure de répétition, par exemple, pour le même patron « , VB le NN » qui est statistiquement significatif pour les quatre derniers romans de Hugo :

*58. Voici donc le grand art : faire un peu rendre à un succès le son d'une catastrophe afin que ceux qui en profitent en tremblent aussi, assaisonner de peur un pas de fait, augmenter la courbe de la transition jusqu'au ralentissement du progrès, affadir cette aurore, dénoncer et retrancher les âpretés de l'enthousiasme, couper les angles et les ongles, **ouater le triomphe, emmitoufler le droit**, envelopper le géant peuple de flanelle et le coucher bien vite, imposer la diète à cet excès de santé, mettre Hercule en traitement de convalescence, délayer l'événement dans l'expédient, offrir aux esprits altérés d'idéal ce nectar étendu de tisane, prendre ses précautions contre le trop de réussite, garnir la révolution d'un abat-jour. (Les Misérables)*

Nous constatons ce même jeu consistant à rapprocher des mots abstraits avec des mots concrets dans les deux structures parallèles : « ouater le triomphe », « emmitoufler le droit ». Il s'agit ici de la combinaison d'un mot abstrait avec un verbe dont le sémantisme est lié au domaine du textile. L'objectif est de rendre « le grand art » plus compréhensible et sensible pour le lecteur.

-Le jeu sonore

Les figures de style opérant à partir de rapprochements sonores sont aussi présentes dans nos résultats. L'exemple suivant montre une phrase avec le patron « ADJ et ADJ » dans lequel le parallélisme se combine avec l'homéotéleute (figure de style qui consiste en la répétition d'une ou de plusieurs syllabes finales homophones de mots, de vers ou

de phrase²⁴⁹) :

59. *Le comte de Lamothé qui, en 1815, était un vieillard de soixante-quinze ans, n'avait de remarquable que son air **silencieux et sentencieux**, sa figure anguleuse et froide, ses manières parfaitement polies, son habit boutonné jusqu'à la cravate, et ses grandes jambes toujours croisées dans un long pantalon flasque couleur de terre de Sienne brûlée.* (Les Misérables)

Un effet d'insistance est créé par la combinaison de deux figures de style dans « silencieux et sentencieux » : le parallélisme qui crée un rythme répétitif ; et l'homéotéleute qui met en relief la consonne « s », également par le rapprochement et la répétition, ayant pour but de mettre en relief les qualités du comte de Lamothé.

Les analyses qualitatives et quantitatives des répétitions syntaxiques nous permettent d'aboutir à la même conclusion que pour les inversions et les clivages. Les résultats confirment souvent ce que disent les critiques littéraires sur le style « hétérogène » de Victor Hugo après l'exil (la répétition abondante, la métaphore), tout en relevant aussi des caractéristiques qui ne sont pas systématiquement étudiées par les critiques littéraires. Pour la répétition, nous constatons la plus grande fréquence de cette figure de style dans les derniers romans de Hugo ; nous repérons également la combinaison de cette figure de style avec les autres figures préférées et caractéristiques de Hugo. L'étude quantitative montre que Victor Hugo utilise de plus en plus cette figure de style au cours du temps, afin de créer des effets d'insistance dans ses derniers romans. Cependant, le fait que ces figures soient également significativement présentes dans les romans *Bug-Jargal* et *Le Dernier Jour d'un condamné*, ainsi que les jeux sur les sons dans les répétitions de Hugo, n'avaient pas été repérés jusqu'ici. Cela pourrait intéresser les études stylistiques ultérieures sur les œuvres de Hugo.

²⁴⁹ « Homéotéleute », <https://fr.wikipedia.org/wiki/Homéotéleute>, consulté le 24 juin 2021.

Nous concluons en soulignant que notre méthode permet d'extraire les patrons caractérisants dont les caractéristiques révélées par les analyses quantitatives et qualitatives permettent de caractériser le style de Victor Hugo de deux manières différentes : d'une part, par l'analyse statistique, qui déchiffre, d'une manière générale, la préférence de l'auteur à employer les répétitions dans ses écrits ; d'autre part, par les analyses qualitatives, qui permettent de mettre au jour les caractéristiques stylistiques de ces patrons caractérisants. La combinaison de ces deux parties d'analyses confirme et complète les critiques littéraires sur le style « hétérogène » de Hugo.

4.5 Application de la méthode d'extraction non-supervisée au corpus Victor Hugo

Nous avons jusqu'ici appliqué la méthode d'extraction supervisée au corpus de neuf romans de Victor Hugo. Nous avons montré les caractéristiques quantitatives et qualitatives du style de Hugo par le truchement des trois patrons caractérisants : l'inversion, le clivage et la répétition. Nous allons maintenant étudier les patrons syntaxiques identificatoires, extraits d'une manière non-supervisée par la méthode que nous avons présentée dans le Chapitre 3. Les patrons syntaxiques reposent sur des constructions repérées grâce à la grammaire de dépendance.

Pour qu'un patron soit identificatoire, il doit être statistiquement spécifique d'un texte par rapport aux autres. La première partie de cette étude porte sur les analyses quantitatives de tous les patrons afin de montrer les spécificités de chaque texte. La méthode statistique que nous avons choisie est l'analyse factorielle des correspondances (AFC), basée sur le test du χ^2 . Nous expliquerons le mécanisme de calcul ainsi que les résultats obtenus dans la partie dédiée à l'analyse statistique.

Les analyses qualitatives étudient les patrons spécifiques, d'après les résultats de l'analyse statistique. Les patrons sont examinés l'un après l'autre afin d'essayer de cerner leurs fonctions stylistiques dans le texte. Cette section inclut également une étude des caractéristiques stylistiques de l'ensemble des patrons parmi les quatre derniers romans de Hugo.

Nous avons évoqué dans les sections précédentes que les patrons identificatoires, bien que statistiquement spécifiques, ne sont pas forcément interprétables du point de vue stylistique. Ainsi, la plupart des patrons identificatoires dans nos résultats ne présentent pas de fonction stylistique facilement identifiable. Analyser les patrons dans leur contexte permet toutefois, dans certains cas, de repérer des phénomènes stylistiques pertinents et intéressants. Ce sont ces patrons interprétables qui nous intéressent en premier lieu.

4.5.1 Analyses quantitatives des patrons syntaxiques

4.5.1.1 La préparation des données

Les patrons extraits d'une manière non-supervisée pour chacun des romans de Victor Hugo sont traités par un script en Python, afin de calculer leur fréquence respective. Les résultats et les fréquences associées sont sauvegardés pour chaque roman avant d'être combinés dans un seul fichier (comprenant les fréquences de chaque patron pour chaque roman de Victor Hugo). Voici un extrait du tableau de fréquences :

Romans	le NN de le NN	ADJ et ADJ	à le NN de le NN	le NN de NP	de le NN de le NN	un NN de NN
BUG JARGAL	101	60	45	37	35	35
CLAUDE GUEUX	17	9	10	4	8	3
HAN DISLANDE	203	199	96	133	84	66
LE DERNIER JOUR DUN CONDAMNÉ	31	49	24	8	9	18
LES MISÉRABLES	780	635	364	350	296	334
LES TRAVAILLEURS DE LA MER	338	132	104	85	114	99
L'HOMME QUI RIT	379	236	192	171	119	149
NOTRE DAME DE PARIS	261	266	155	131	154	88
QUATRE VINGT TREIZE	210	90	86	122	99	95

Tableau 4.14 Extrait du tableau de fréquences absolues des patrons.

Deux profondeurs d'extraction (les profondeurs 3 et 4) ont été appliquées sur chaque roman (voir la section 3.7.2 au Chapitre 3 pour une discussion sur le choix des profondeurs d'extraction). Dans le fichier combiné de neuf romans, nous sélectionnons les 50 patrons les plus fréquents de profondeur 3 et les 50 patrons les plus fréquents de profondeur 4. La sélection du nombre de patrons est libre, elle devrait être déterminée en fonction de l'objectif d'étude. Dans notre étude, nous avons choisi les 100 patrons les plus fréquents parce que nous avons observé que les patrons les plus fréquents sont aussi les plus interprétables, ceux qui présentent les caractéristiques les plus généralisables grâce à leur fréquence élevée.

4.5.1.2 L'AFC et le test du χ^2

Nous avons choisi l'Analyse factorielle des correspondances pour faire les analyses statistiques en R. L'AFC est une méthode permettant de décrire et de hiérarchiser les relations statistiques qui peuvent exister entre des individus placés en ligne (les patrons) et des variables placées en colonne (les romans). Il existe d'autres moyens que la statistique multivariée pour mettre en corrélation des variables ou des individus : par exemple la R-value, la Q-value²⁵⁰ ou l'Analyse en composantes principales²⁵¹. Cependant, ces méthodes analysent les corrélations soit entre variables, soit entre individus, alors que l'AFC analyse les corrélations entre les individus et les variables. L'AFC est donc conforme à notre objectif de mettre en évidence les corrélations entre les œuvres et les patrons.

L'analyse factorielle des correspondances part des données d'entrée exprimées en valeur absolue. Ensuite, les valeurs relatives sont calculées en divisant chaque valeur absolue par le nombre total des fréquences de chaque ligne pour une variable. Le tableau ci-dessous présente le processus de transformation en valeur relative : la transformation a pour objectif d'éliminer l'effet de la longueur variable des textes sur les résultats. À la fin de la transformation, la moyenne est calculée pour chaque patron en valeur relative (« LA MOYENNE » en dernière ligne) :

²⁵⁰ [https://en.wikipedia.org/wiki/Q-value_\(statistics\)](https://en.wikipedia.org/wiki/Q-value_(statistics)), consulté le 2 août 2021.

²⁵¹ https://fr.wikipedia.org/wiki/Analyse_en_composantes_principales, consulté le 2 août 2021.

Romans	le NN de le NN	ADJ et ADJ	à le NN de le NN	le NN de NP	de le NN de le NN	un NN de NN
BUG_JARGAL	101	60	45	37	35	35
CLAUDE_GUEUX	17	9	10	4	8	3
HAN_DISLANDE	203	199	96	133	84	66
LE_DERNIER_JOUR_DUN_CONDAMNÉ	31	49	24	8	9	18
LES_MISÉRABLES	780	635	364	350	296	334
LES_TRAVAILLEURS_DE_LA_MER	338	132	104	85	114	99
LHOMME QUI RIT	379	236	192	171	119	149
NOTRE_DAME_DE_PARIS	261	266	155	131	154	88
QUATRE_VINGT_TREIZE	210	90	86	122	99	95

Romans	le NN de le NN	ADJ et ADJ	à le NN de le NN	le NN de NP	de le NN de le NN	un NN de NN
BUG_JARGAL	0.322683706	0.191693291	0.143769968	0.118210863	0.111821086	0.111821086
CLAUDE_GUEUX	0.333333333	0.176470588	0.196078431	0.078431373	0.156862745	0.058823529
HAN_DISLANDE	0.259923175	0.254801536	0.122919334	0.170294494	0.107554417	0.084507042
LE_DERNIER_JOUR_DUN_CONDAMNÉ	0.223021583	0.352517986	0.172661871	0.057553957	0.064748201	0.129496403
LES_MISÉRABLES	0.282711127	0.230155854	0.131931859	0.126857557	0.107285248	0.121058354
LES_TRAVAILLEURS_DE_LA_MER	0.387614679	0.151376147	0.119266055	0.097477064	0.130733945	0.11353211
LHOMME QUI RIT	0.304173355	0.1894061	0.154093098	0.137239165	0.095505618	0.119582665
NOTRE_DAME_DE_PARIS	0.247393365	0.252132701	0.146919431	0.124170616	0.145971564	0.083412322
QUATRE_VINGT_TREIZE	0.299145299	0.128205128	0.122507123	0.173789174	0.141025641	0.135327635
LA MOYENNE	0.295555514	0.21408437	0.145571908	0.12044714	0.117945385	0.106395683

Tableau 4.15 Extrait des tableaux de fréquences absolues et relatives des patrons.

Les données sont ensuite utilisées pour calculer la distance entre deux individus par une fonction en R, d'après le test du χ^2 (les X_a , X_b , Y_a , Y_b dans la formule représentent les valeurs relative d'un patron pour chaque roman, par exemple, la distance χ^2 entre *Bug Jargal* et *Claude Gueux* est calculée en remplaçant X_b par 0.333333333, X_a par 0.322683706, « Average profile x » par $(0.333333333+0.322683706)/2$, Y_b par 0.176470588, Y_a par 0.191693291, « Average profile y » par $(0.176470588+0.191693291)/2$, etc.) :

$$\text{Chi-squared distance}(A, B) = \sqrt{\frac{(x_B - x_A)^2}{\text{Average profile x}} + \frac{(y_B - y_A)^2}{\text{Average profile y}} + \dots}$$

Figure 4.6 Le test du χ^2 .

Les calculs sont réalisés en R. Nous constatons que le test du χ^2 est essentiellement une dérivation du calcul de l'écart type. Ainsi, la spécificité de chaque patron dans un certain roman est en fait la distance statistique de ce patron avec la moyenne.

4.5.1.3 Les résultats d'analyses quantitatives

Le tableau suivant présente la répartition des romans de Hugo sur un espace bidimensionnel, d'après le calcul des spécificités en fonction des patrons syntaxiques les plus fréquents :

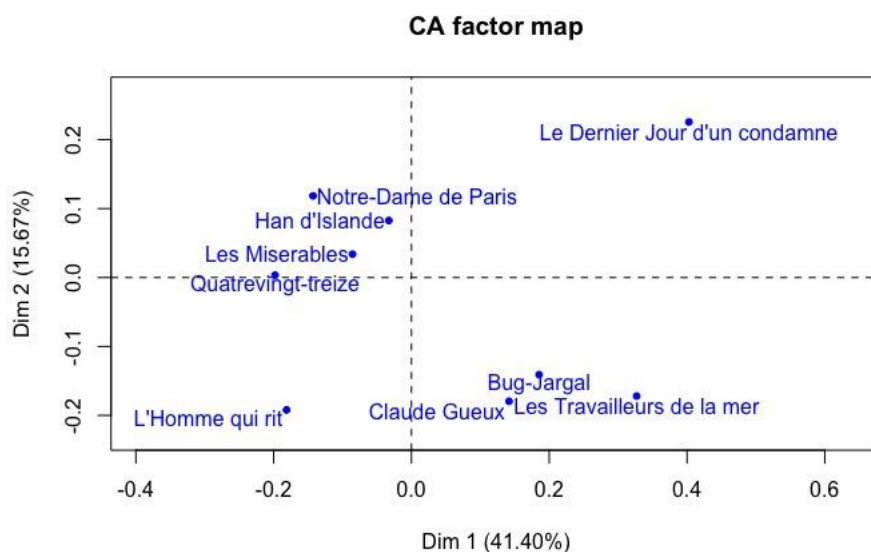


Figure 4.7 AFC (CA Factor Map) basée sur les 100 patrons les plus fréquents.

Plusieurs caractéristiques peuvent être constatées à partir de cette figure :

1. La ressemblance, déjà constatée avec les résultats des patrons caractérisants, entre *Les Misérables* et *Notre-Dame de Paris* est également visible ici avec les patrons identificatoires. Les patrons caractérisants et les patrons identificatoires montrent donc la même ressemblance entre ces deux romans de Hugo, publiés à deux époques différentes. De plus, selon les résultats sur les patrons syntaxiques, deux autres romans : *Quatrevingt-treize* et *Han d'Islande* possèdent aussi des caractéristiques syntaxiques analogues aux *Misérables* et à *Notre-Dame de Paris*.

2. Une forte ressemblance est également constatée entre *Bug-Jargal*, *Claude Gueux* et *Les Travailleurs de la mer*. Ceci n'était en revanche pas visible avec les patrons caractérisants, mais la prise en compte de tous les patrons les plus fréquents le permet ici.
3. *Le Dernier Jour d'un condamné* est toujours exceptionnellement écarté des autres romans.
4. *L'Homme qui rit* est également syntaxiquement spécifique par rapport aux autres romans de Victor Hugo.

La figure suivante présente la corrélation entre les patrons et les œuvres. Cette corrélation met en évidence la contribution de chaque patron à la différenciation des œuvres, et c'est ainsi que nous obtenons les patrons spécifiques pour chaque roman :

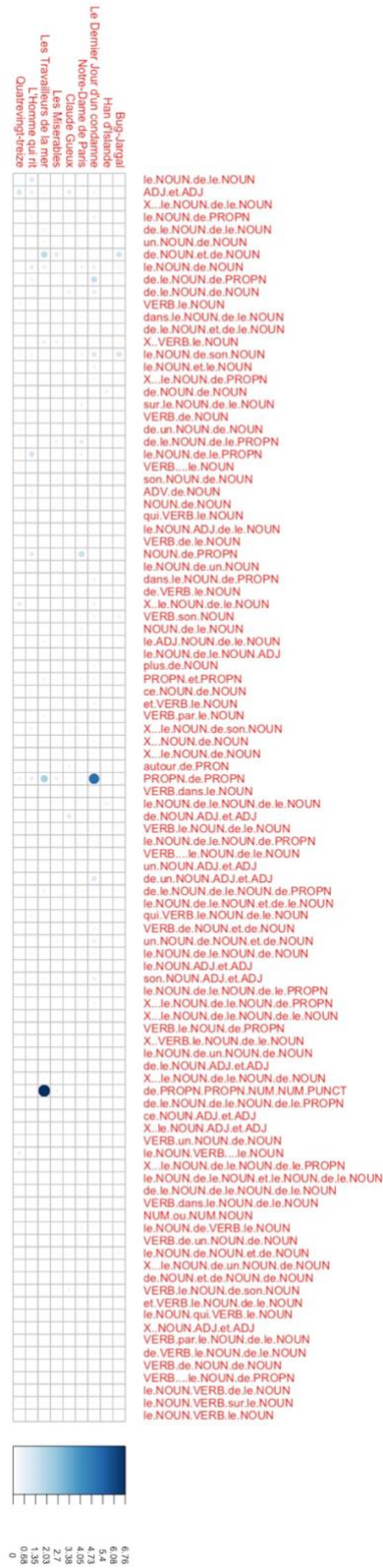


Figure 4.8 Corrélations entre les patrons et les œuvres.

Plus le cercle est foncé, plus le patron contribue à la différenciation de ce roman particulier²⁵². Nous considérons ainsi tous les patrons qui présentent un cercle foncé à l'entrecroisement d'un roman comme les patrons « spécifiques » pour ce roman.

Bug-Jargal	Han d'Islande	Le Dernier Jour d'un condamné	Notre-Dame de Paris	Claude Gueux	Les Misérables	Les Travailleurs de la mer	L'Homme qui rit	Quatrevingt-treize
de NN et de NN	de NN de NN	Le NN de NP	de le NN de le NP	ADJ et ADJ	de NN et de NN	de NN et de NN	le NN de le NN	ADJ et ADJ
le NN de son NN	le NN de le NN de le NN	de le NN de NN	NN de NP	de NN ADJ et ADJ	, VB le NN	, VB le NN	ADJ et ADJ	à le NN de le NN
VB son NN		le NN de son NN			NP de NP	NP de NP	le NN de NN	
le NN de NN et de NN		à le NN de le NN				de NP NP NUM NUM PUNCT	le NN de le NP	
		PORPN et NP				le NN de le NN	NN de NP	
		de un NN ADJ et ADJ					NP de NP	
		un NN de NN et de NN						

Tableau 4.16 Patrons spécifiques d'après l'AFC pour chaque roman de Hugo.

Le tableau ci-dessus présente les patrons spécifiques, d'après nos résultats statistiques, pour chaque roman de Victor Hugo. Nous procédons à l'analyse de ces patrons dans la partie suivante.

4.5.2 Analyses qualitatives des patrons identificateurs

4.5.2.1 Analyses des fonctions stylistiques des patrons spécifiques

Dans cette section, nous étudions l'un après l'autre les patrons spécifiques (d'après les résultats de l'AFC), pour chaque roman de Victor Hugo. Ces analyses consistent à montrer l'intérêt de ces patrons comme indicateurs stylométriques, à savoir leur capacité à déchiffrer et à généraliser les phénomènes stylistiques propres du texte. Nous nous concentrons dans cette partie que sur les patrons sur-employés pour chaque roman (ces patrons peuvent aussi être spécifiques en raison de leur fréquence moins importante

²⁵² En raison du nombre élevé de patrons dans notre calcul, la visualisation du résultat de corrélation présentée dans la Figure 4.8 n'est pas très claire. Les patrons spécifiques avec les cercles foncés sont listés dans le Tableau 4.16.

dans un texte, mais nous insistons sur le fait que les caractéristiques sont plus généralisables dans les patrons avec des fréquences relativement élevées). Nous avons décidé de ne pas inclure les patrons avec des noms propres et les patrons avec des chiffres (dates, etc.) parce qu'ils ne contiennent que très peu d'informations stylistiques. Plusieurs caractéristiques stylistiques coexistent souvent dans un seul patron.

- « **Le NN de le NN** »

Ce patron nominal est sur-employé dans *Les Travailleurs de la Mer* (261 occurrences) et *L'Homme qui rit* (338 occurrences), deux des romans publiés après l'exil de Victor Hugo. Il apparaît principalement avec la fonction objet (46.35%) et sujet (39.58%) dans les phrases.

Nous avons relevé trois caractéristiques stylistiques de ce patron dans le roman *Les Travailleurs de la Mer* et *L'Homme qui rit*, en étudiant les occurrences l'une après l'autre dans les phrases. Les exemples sont tous tirés de ces deux romans.

-La nominalisation d'un adjectif

Ce phénomène est très courant dans nos résultats. Au lieu d'utiliser le patron « le NN ADJ », Hugo nominalise l'adjectif de cette construction pour ensuite le placer devant le nom qu'il désigne. Ce phénomène stylistique apparaît souvent dans les phrases où le patron « le NN de le NN » joue le rôle du sujet de la proposition.

Dans l'exemple suivant, « l'abîme lointain » est remplacé par « les lointains de l'abîme », l'utilisation du pluriel renforce l'aspect concret de l'adjectif.

*60. Les choses sombres du monde ignoré deviennent voisines de l'homme, soit qu'il y ait communication véritable, soit que **les lointains de l'abîme** aient un grossissement visionnaire ; (Les Travailleurs de la Mer)*

Ou encore, dans *L'Homme qui rit*, l'exemple suivant montre le même phénomène pour transformer l'adjectif « intense » en nom « intensité » :

61. *L'intensité du froid* avait d'ailleurs fait de cette neige une poussière, assez incommode au marcheur. (L'Homme qui rit)

La nominalisation de l'adjectif donne un effet d'insistance sur le premier nom du patron « le NN de le NN ». Dans l'exemple suivant, l'accent est mis sur le mot « odieux » alors que si l'auteur avait préféré « l'hypocrisie odieuse », l'accent se serait trouvé sur « l'hypocrisie » :

62. *L'odieux de l'hypocrisie* commence obscurément dans l'hypocrite. (Les Travailleurs de la Mer)

Le dernier exemple montre que ce phénomène pourrait aussi se trouver dans les cas où le patron « le NN et le NN » est l'objet d'une proposition :

63. *Sur le fleuve Arinos, affluent du Tocantins, dans les forêts vierges au nord de Diamantina, il avait constaté l'effrayant peuple chauve-souris, les murcilagos, hommes qui naissent avec les cheveux blancs et les yeux rouges, habitent le sombre des bois, dorment le jour, s'éveillent la nuit, et pêchent et chassent dans les ténèbres, y voyant mieux quand il n'y a pas de lune.* (Les Travailleurs de la Mer)

Ce phénomène de nominalisation nous permet de confirmer les études littéraires sur le style « hétérogène » de Hugo, d'une part parce que le croisement entre les champs sémantiques abstrait et concret est une des caractéristiques proéminentes de son style, discutées dans la section 4.2, d'autre part parce que nous constatons une autre manière

de transformer le patron « le NN ADJ », un patron courant dans la langue française dont la spécificité n'a pas été démontrée d'après nos résultats statistiques.

-La mise en place d'une métaphore

Le patron « le NN de le NN » possède une fonction stylistique permettant de mettre en place les métaphores. *Les Travailleurs de la Mer* est une guerre affreuse entre l'homme et la mer, par conséquent, les métaphores que nous avons trouvées portent le plus souvent sur la personnification de la mer :

64. *Le monologue des flots ne sera point troublé.* (Les Travailleurs de la Mer)

65. *La palpitation de la mer se faisait sentir dans cette cave.* (Les Travailleurs de la Mer)

Mais il pourrait également s'agir des métaphores simples comme dans les trois exemples suivants : le premier présente le croisement du sémantisme médical avec le celui de la nature ; le second met en place une hypallage (qui consiste à attribuer à certains mots d'une phrase, comme « lieu » ce qui convient à d'autres mots de la même phrase, ici « solitude ») ; le troisième exemple concrétise « l'enfer » en le combinant avec « les soupiraux », comme si l'enfer était un endroit concret et perceptible :

66. *Il savait à fond la pathologie des saisons.* (Les Travailleurs de la Mer)

67. *La solitude du lieu avait on ne sait quoi de funèbre.* (Les Travailleurs de la Mer)

68. *Les soupiraux de l'enfer* doivent donner passage à des essaims pareils.
(L'Homme qui rit)

La métaphore est une des caractéristiques stylistiques les plus présentes dans les romans de Hugo. Rappelons que ce phénomène consiste à rendre les idées plus concrètes, en reliant par surprise le comparant et le comparé, issus de deux sémantismes différents.

-La mise en relation de plusieurs occurrences du même patron dans une phrase

Ce phénomène est essentiellement une figure de répétition. Nous avons constaté la reprise du patron « le NN de le NN » dans une même phrase :

69. *Il y avait le secret de la fraude* comme il y a *le secret de la confession*. (Les Travailleurs de la Mer)

70. *Voir le dedans de la mer*, c'est voir *l'imagination de l'Inconnu*. (Les Travailleurs de la Mer)

71. *La blancheur de l'horizon* augmentait *la noirceur de la mer*. (Les Travailleurs de la Mer)

72. *Et l'onde des effluves*, tantôt aide, tantôt contrarie *l'onde des airs* et *l'onde des eaux*. (L'Homme qui rit)

73. *Elles ont l'enchevêtrement du polypier*, *la sublimité de la cathédrale*, *l'extravagance de la pagode*, *l'amplitude du mont*, *la délicatesse du bijou*, *l'horreur du sépulcre*. (Les Travailleurs de la Mer)

Nous avons montré dans la section 4.2 que la répétition est une des manifestations linguistiques du style hétérogène de Hugo. Dans les trois premiers exemples, la reprise du patron « le NN de le NN » a été réalisée de manière symétrique. Les deux premières phrases mettent en place deux comparaisons entre « le secret de la fraude » et « le secret de la confession », et entre « le dedans de la mer » et « l'imagination de l'Inconnu ». Nous retrouvons toujours la volonté de l'auteur de rendre ses idées plus compréhensibles par l'analogie. La troisième phrase reprend la structure syntaxique du sujet dans les deux compléments d'objet direct, cette reprise créant un rythme répétitif. Le quatrième exemple montre une opposition entre « la blancheur de l'horizon » et « la noirceur de la mer », toujours avec un rythme bien symétrique pour cette phrase courte. Le dernier exemple présente le patron « le NN de le NN » en parallélisme, une des figures de répétitions très appréciées de Victor Hugo.

Les caractéristiques stylistiques que nous avons repérées pour le patron spécifique « le NN de le NN » dans *Les Travailleurs de la Mer* et *L'Homme qui rit* sont représentatives du style de Hugo. Nous n'avons pas remarqué de différences concernant les caractéristiques stylistiques de ce patron entre les deux romans, probablement en raison de leurs dates de publication très proches. La nominalisation des adjectifs modifie l'accent de phrase, elle met en évidence également le mélange de sémantismes éloignés, cher à Hugo, afin de rendre les idées plus concrètes. La reprise d'un même patron pour mettre en scène une comparaison ou un parallélisme est aussi emblématique, il s'agit d'un exemple typique de la répétition hugolienne. Nous concluons que l'étude des fonctions stylistiques du patron « le NN de le NN » apporte des éléments concrets et précis alimentant l'idée d'un style « hétérogène » de Victor Hugo.

- **« de NN et de NN »**

Nous explorons ensuite un autre patron nominal : « de NN et de NN ». Ce patron

est sur-employé statistiquement (187 occurrences) dans le roman *Les Misérables*, un des romans publiés après l'exil de l'auteur. Ce patron est également sur-employé (61 occurrences) dans *Bug-Jargal*, le premier roman de Victor Hugo. Nous allons montrer dans cette section les caractéristiques stylistiques de ce patron dans les deux romans, afin de vérifier s'il existe de différences d'utilisation de ce patron entre les deux périodes de publication de Hugo (avant et après l'exil).

-Expansion sémantique homogène

Ce patron est construit autour de deux noms coordonnés, ainsi il nous semble pertinent d'étudier ce patron à travers le sémantisme des deux noms en question. Quand les deux noms sont synonymes, le deuxième sert principalement à renforcer le premier en donnant un effet d'insistance :

74. *Oeil plein de soupçon et de conjectures.* (Les Misérables)

75. *Le bruit de ce gond sonna dans son oreille avec quelque chose d'éclatant et de formidable comme le clairon du jugement dernier.* (Les Misérables)

76. *Pour un observateur qui l'eût étudiée attentivement, ce qui se dégageait d'elle, à travers toute cette ivresse de l'âge, de la saison et de l'amourette, c'était une invincible expression de retenue et de modestie.* (Les Misérables)

77. *Une des idées fausses de la bourgeoisie de la Restauration en fait d'aristocratie et de noblesse, c'était de croire à la particule.* (Les Misérables)

78. *Elle reconnut ma voix, vit briller un poignard, et poussa un cri d'angoisse et de terreur.* (Bug-Jargal)

Les cinq exemples montrent ce même phénomène d'expansion sémantique : le deuxième nom est considéré comme une expansion de la même idée, déjà exprimée par le premier nom. Le patron « de NN et de NN » est un procédé efficace mettant en place deux noms différents pour désigner les qualités de l'objet qu'ils qualifient. Ce phénomène se trouve aussi bien dans *Les Misérables* que dans *Bug-Jargal*, le premier roman de Victor Hugo. Ce patron peut donc représenter le style de Hugo, dont la diversité se manifeste également au niveau lexical.

-Expansion sémantique hétérogène

Ce phénomène est très présent dans nos résultats et il se manifeste sous des formes différentes. Les deux noms en conjonction peuvent être antithétiques. Dans les deux exemples suivants, « rudesse » et « majesté » expriment deux qualités qui ne sont pas homogènes, l'association des deux mots crée un effet de surprise aux yeux du lecteur :

*79. Et pour que la nature, qui déconcerte les arrangements mesquins de l'homme et qui se répand toujours tout entière là où elle se répand, aussi bien dans la fourmi que dans l'aigle, en vînt à s'épanouir dans un méchant petit jardin parisien avec autant **de rudesse et de majesté** que dans une forêt vierge du Nouveau Monde.*
(Les Misérables)

*80. Il s'était déclaré roi, et celui-ci n'était qu'un esclave, mais je me rappelais, non sans étonnement, l'air de **rudesse et de majesté** empreint sur son visage au milieu des signes caractéristiques de la race africaine, l'éclat de ses yeux, la blancheur de ses dents sur le noir éclatant de sa peau, la largeur de son front, surprenante surtout chez un nègre, le gonflement dédaigneux qui donnait à l'épaisseur de ses lèvres et de ses narines quelque chose de si fier et de si puissant,*

la noblesse de son port, la beauté de ses formes, qui, quoique maigries et dégradées par la fatigue d'un travail journalier, avaient encore un développement pour ainsi dire herculéen ; (Bug-Jargal)

Ou encore, dans l'exemple suivant, l'accent mis à la fois sur la « mansuétude » et sur l'« autorité » se rapporte à deux qualités différentes : la douceur et la sévérité.

81. *Il l'interrompt d'un accent plein de mansuétude et d'autorité.* (Les Misérables)

82. *Un violent besoin d'en finir avec l'être inconnu qui osait ainsi associer le nom de Marie à des chants d'amour et de menace s'empara de moi.* (Bug-Jargal)

L'exemple suivant, tiré de *Bug-Jargal*, met en conjonction un nom mélioratif, « finesse », avec un mot péjoratif, « cruauté » :

83. *Sa figure ignoble offrait un rare mélange de finesse et de cruauté.* (Bug-Jargal)

La « gaîté » dans ce patron hugolien se mélange avec la « froideur » et la « crainte » : ces termes antithétiques rendent incertaines les qualités de « l'âme des jeunes filles » et de « la surprise » qu'éprouve le personnage :

84. *La blancheur de l'âme des jeunes filles, qui se compose de froideur et de gaîté, ressemble à la neige.* (Les Misérables)

85. *Il venait d'y regarder, et là où il s'attendait à ne trouver qu'un chaos de ténèbres, il avait vu, avec une sorte de surprise inouïe mêlée de crainte et de joie, étinceler des astres, Mirabeau, Vergniaud, Saint-Just, Robespierre, Camille*

Desmoulins, Danton, et se lever un soleil, Napoléon. (Les Misérables)

L'expansion sémantique hétérogène peut aussi se manifester par l'emploi de deux noms issus de deux sémantismes différents. Les deux exemples suivants montrent ce phénomène :

*86. Il était là tout à l'heure, il était de l'équipage, il allait et venait sur le pont avec les autres, il avait sa part **de respiration et de soleil**, il était un vivant. (Les Misérables)*

*87. Ainsi, pendant ces dix-neuf ans **de torture et d'esclavage**, cette âme monta et tomba en même temps. (Les Misérables)*

*88. En effet, une voix dont l'harmonie avait quelque chose **de mâle et de plaintif** à la fois sortit un moment après du fond du bois, et mêla aux notes graves de la guitare une romance espagnole, dont chaque parole retentit assez profondément dans mon oreille pour que ma mémoire puisse encore aujourd'hui en retrouver presque toutes les expressions. (Bug-Jargal)*

La première phrase s'ouvre par la mention de la « respiration » : le lecteur s'attend à trouver ensuite d'autres noms du même champ lexical (« palpitations », etc.), mais l'auteur change brusquement le champ sémantique et met en place le nom « soleil ». Il s'agit toujours d'une expansion sémantique pour désigner le personnage vivant, mais les deux noms ont un sémantisme radicalement différent. Le deuxième exemple montre la combinaison d'un mot concret avec un mot général et abstrait dans le patron « de NN et de NN » : « torture » est un nom concret, alors que « l'esclavage » exprime une idée plus générale. Le dernier exemple dans *Bug-Jargal* présente deux noms de sémantismes différents, « mâle » et « plaintif » pour désigner la « voix ».

L'expansion sémantique que nous avons constatée dans le patron « de NN et de NN » nous permet de confirmer le style « hétérogène » de Victor Hugo : les antithèses sont abondantes, le mélange de mots issus des sémantismes différents est également très courant.

-Le rapprochement sonore

Ce phénomène est également très caractéristique dans nos résultats, les deux noms en conjonction du patron « de NN et de NN » sont séparés par un seul mot « et », ainsi, les effets stylistiques sur le rapprochement sonore sont souvent faciles à créer. Les quatre phrases ci-dessous montrent ce phénomène :

89. *Une raison, à jeun **de science et de sagesse**, maigrit.* (Les Misérables)

90. *Les trois autres, moins timides, nous l'avons dit, étaient décolletées tout net, ce qui, l'été, sous des chapeaux couverts de fleurs, a beaucoup **de grâce et d'agacerie** ; mais, à côté de ces ajustements hardis, le canezou de la blonde Fantine, avec ses transparences, ses indiscretions et ses réticences, cachant et montrant à la fois, semblait une trouvaille provocante de la décence, et la fameuse cour d'amour, présidée par la vicomtesse de Cette aux yeux vert de mer, eût peut-être donné le prix de la coquetterie à ce canezou qui concourait pour la chasteté.*
(Les Misérables)

91. *Monsieur, dit le petit savoyard, avec cette confiance de l'enfance qui se compose **d'ignorance et d'innocence**, ma pièce ?* (Les Misérables)

92. *C'étaient tantôt des troupes de nègres absolument nus, munis de massues, de tomahawks, de casse-têtes, marchant au son de la corne à bouquin, comme les*

sauvages ; tantôt des bataillons de mulâtres, équipés à l'espagnole ou à l'anglaise, bien armés et bien disciplinés, réglant leurs pas sur le roulement d'un tambour ; puis des cohues de négresses, de négrellons, chargés de fourches et de broches ; des fatras courbés sous de vieux fusils sans chien et sans canon ; des griotes avec leurs parures bariolées ; des griots, effroyables de grimaces et de contorsions, chantant des airs incohérents sur la guitare, le tam-tam et le balafo. (Bug-Jargal)

Pour conclure, le patron « de NN et de NN » donne accès à caractéristiques stylistiques intéressantes dans *Les Misérables* et *Bug-Jargal*. L'expansion sémantique permet à l'auteur d'employer deux mots différents pour qualifier le même objet, ce que confirme partiellement le style de Hugo par la diversité lexicale ; l'expansion sémantique hétérogène comprend souvent des antithèses, c'est-à-dire des mots issus des sémantismes différents. Ce sont deux des caractéristiques du style « hétérogène » de Hugo, selon les critiques littéraires ; enfin, les jeux sonores au sein de ce patron n'ont pas été étudiés systématiquement dans les études précédentes, il s'agit d'un apport de la thèse permettant de contribuer aux recherches littéraires sur le style de Hugo. Nos résultats ne montrent cependant pas de différences radicales entre les deux romans *Les Misérables* et *Bug-Jargal*, pourtant publiés dans deux périodes différentes. Ce patron n'est donc pas spécifique d'une période donnée.

- « , VB le NN »

Ce patron verbal, « , VB le NN », est spécifique *des Misérables* (174 occurrences). En examinant les occurrences de ce patron en corpus, nous voyons qu'il correspond en fait à une grande partie d'incises :

93. Allons, *dit l'évêque*, faites une bonne nuit.

À l'exception des incises, nous avons constaté deux caractéristiques stylistiques qui se trouvent dans ce patron.

-L'enchaînement d'une séquence d'actions de l'actant sujet.

Le patron « , VB le NN » est employé très souvent pour enchaîner une séquence d'actions de l'actant sujet, probablement en raison de la virgule qui marque le commencement d'une autre proposition. Les trois phrases suivantes montrent ce phénomène : ce patron avec un rythme bref présente avec une intensité considérable l'énumération des actions exercées par le sujet de la phrase, le rythme haletant est créé par l'enchaînement des syntagmes verbaux courts et brefs :

*94. Il le prit, traversa la chambre à grands pas sans précaution et sans s'occuper du bruit, **gagna la porte**, rentra dans l'oratoire, **ouvrit la fenêtre**, saisit un bâton, enjamba l'appui du rez-de-chaussée, mit l'argenterie dans son sac, **jeta le panier**, **franchit le jardin**, sauta par-dessus le mur comme un tigre, et s'enfuit.*

*95. Il rouvrit les yeux, se dressa brusquement sur son séant, **étendit le bras** et tâta son havresac qu'il avait jeté dans le coin de l'alcôve, puis il laissa pendre ses jambes et poser ses pieds à terre, et se trouva, sans savoir comment, assis sur son lit.*

*96. Enfoncer les carrés, **pulvériser les régiments**, **rompre les lignes**, broyer et disperser les masses, tout pour lui était là, frapper, frapper, frapper sans cesse, et il confiait cette besogne au boulet.*

-La mise en place du parallélisme

Parmi les résultats, nous constatons également le parallélisme, construit sur la base du patron « , VB le NN ». Le parallélisme est une des figures de répétition dont l'importance est souvent remarquée par les critiques littéraires comme caractéristique du style de Hugo :

*97. Ce sont des êtres de la grande espèce neutre ; hongres, parasites, nuls, qui ont un peu de terre, un peu de sottise et un peu d'esprit, qui seraient des rustres dans un salon et se croient des gentilshommes au cabaret, qui disent : mes prés, mes bois, mes paysans, sifflent les actrices du théâtre pour prouver que ce sont des gens de goût, querellent les officiers de la garnison pour montrer qu'ils sont gens de guerre, chassent, fument, bâillent, boivent, sentent le tabac, jouent au billard, regardent les voyageurs descendre de diligence, vivent au café, dînent à l'auberge, ont un chien qui mange les os sous la table et une maîtresse qui pose les plats dessus, tiennent à un sou, **exagèrent les modes, admirent la tragédie, méprisent les femmes**, usent leurs vieilles bottes, copient Londres à travers Paris et Paris à travers Pont-à-Mousson, vieillissent hébétés, ne travaillent pas, ne servent à rien, et ne nuisent pas à grand-chose.*

*98. Il se pencha de nouveau, **fouilla le tas**, écarta ce qui faisait obstacle, **saisit la main, empoigna le bras, dégagea la tête, tira le corps**, et quelques instants après il traînait dans l'ombre du chemin creux un homme inanimé, au moins évanoui.*

*Rêver la prolongation indéfinie des choses défuntes et le gouvernement des hommes par embaumement, restaurer les dogmes en mauvais état, **redorer les châsses, recrépir les cloîtres, rebénir les reliquaires, remeubler les superstitions, ravitailler les fanatismes**, remmancher les goupillons et les sabres, reconstituer*

le monachisme et le militarisme, croire au salut de la société par la multiplication des parasites, imposer le passé au présent, cela semble étrange.

Les fonctions stylistiques qui se trouvent au sein de ce patron, « , VB le NN », sont emblématiques du style de Hugo. Les répétitions en cascade, l'énumération et l'accumulation sont les manifestations du style « hétérogène » de l'auteur. Nous concluons que ce patron verbal est capable de caractériser partiellement le style de Hugo par la détection des phénomènes de répétition dans le texte. Pourtant, nous insistons également sur le fait qu'un seul patron n'est jamais suffisant pour donner une conclusion générale sur un style littéraire. C'est pour cette raison que nous examinons tous les patrons sur-employés dans les romans de Hugo, c'est également l'intérêt de la méthode d'extraction non-supervisée, qui permet de couvrir un grand nombre de patrons syntaxiques différents.

- « ADJ et ADJ »

Le patron adjectival, « ADJ et ADJ », est sur-employé dans trois des romans de Victor Hugo : 49 occurrences dans *Claude Gueux* (un roman relativement court), 132 occurrences dans *L'Homme qui rit* et 90 occurrences dans *Quatrevingt-treize*. Après avoir étudié toutes les occurrences de ce patron dans les trois romans, nous avons relevé quatre caractéristiques stylistiques de cette construction.

-Expansion sémantique homogène

Ce phénomène est présent dans nos résultats parce que la construction syntaxique de ce patron, « ADJ et ADJ », met en conjonction deux adjectifs pour désigner un même objet. L'expansion homogène consiste à mettre deux adjectifs différents pour exprimer la même qualité de l'objet. Il s'agit souvent de deux adjectifs synonymes. Dans la

phrase suivante : « redoutable et mystérieux » exprime l'incertitude des qualités des « révolutions ».

*99. Il était né de ce besoin public de violence qui est le côté **redoutable et mystérieux** des révolutions. (Quatrevingt-treize)*

L'exemple suivant montre le même phénomène, « glorieux et triomphale » désigne les mêmes qualités de la « palinodie », mais d'une manière ironique. L'ironie est créée par la combinaison du mot péjoratif, « palinodie », avec les adjectifs mélioratifs qui le désignent :

*100. Pendant que la nation couvrait d'acclamations le roi, reprenant possession de l'Angleterre, pendant que l'unanimité prononçait son verdict, pendant que s'accomplissait la salutation du peuple à la monarchie, pendant que la dynastie se relevait au milieu d'une palinodie **glorieuse et triomphale**, à l'instant où le passé devenait l'avenir et où l'avenir devenait le passé, ce lord était resté réfractaire. (L'Homme qui rit)*

Dans cet exemple, « encombrant » doit être interprété comme « détestable », l'adjectif en conjonction, « réfractaire », complémente sémantiquement « encombrant » en donnant une explication sur la raison pour laquelle les familles sont « encombrantes » :

*101. C'était l'époque où l'on tronquait les familles **encombrantes et réfractaires**, où l'on coupait court aux filiations, où l'on supprimait brusquement les héritiers. (L'Homme qui rit)*

Ce phénomène d'expansion sémantique homogène permet à l'auteur de préciser et d'enrichir le vocabulaire de son texte. Nous l'avons déjà rencontré pour les analyses du

patron nominal « de NN et de NN », il s'agit d'une caractéristique hugolienne visant à créer de la diversité lexicale au sein du roman, un des éléments clés du style « hétérogène ».

-Expansion sémantique hétérogène

Ce phénomène est également courant dans la construction du patron « ADJ et ADJ », il consiste à mettre deux mots antithétiques, ou simplement deux antonymes en conjonction.

*102. Ainsi grondait le camp **victorieux et mécontent**. (Quatrevingt-treize)*

*103. L'ignorance faisant à la vérité, à la justice, au droit, à la raison, à la délivrance, une longue résistance **bête et superbe** ; (Quatrevingt-treize)*

*104. Il le pressentait, il l'entrevoyait d'avance, il le devinait **effrayant et magnifique** ; (Quatrevingt-treize)*

*105. Le flocon **inexorable et doux** fait son oeuvre en silence. (L'Homme qui rit)*

Les quatre exemples que nous citons montrent une caractéristique typique du style de Hugo : l'antithèse. Les adjectifs employés, de quasi-antonymes, créent un effet de surprise et leur interprétation devient, de fait, incertaine.

-La mise en place d'une métaphore

Quand les adjectifs avec le nom précédent forme un tout peu courant, il y a souvent métaphore. Dans l'exemple suivant, les deux adjectifs, « sournoise et colère » sont

employés pour désigner le « noroit », qui s'en trouve ainsi personnifié :

*106. C'était le noroit, qu'on nommait jadis vent de galerne, bise **sournoise et colère**.*
(L'Homme qui rit)

Dans la phrase suivante, le même procédé est employé. La « tempête » devient « aveugle et muette » :

*107. La tempête de neige, c'est la tempête **aveugle et muette**.* (L'Homme qui rit)

Le dernier exemple montre encore le même phénomène, autour du terme « solitude » cette fois.

*108. La large solitude **nue et livide** était devant lui.* (L'Homme qui rit)

Ce phénomène étant uniquement présent dans le roman *L'Homme qui rit*, nous pouvons conclure que la métaphore créée par le patron « ADJ et ADJ » est une caractéristique stylistique importante et spécifique du roman *L'Homme qui rit*.

-Le rapprochement sonore

Vu la présence de deux adjectifs sémantiquement et formellement proches dans ce patron, les jeux sur les sons sont souvent présents. Les quatre phrases suivantes montrent ce phénomène stylistique :

*109. Le marquis constata dans le hallier comme l'éparpillement d'une troupe **furieuse et joyeuse**.* (Quatrevingt-treize)

110. *En présence de la création fleurie, embaumée, **aimante et charmante**, le ciel splendide inondait d'aurore la Tourgue et la guillotine, et semblait dire aux hommes : Regardez ce que je fais et ce que vous faites.* (Quatrevingt-treize)

111. *Dans l'ombre **infinie et indéfinie**, il y a quelque chose, ou quelqu'un, de vivant ; mais ce qui est vivant là fait partie de notre mort.* (L'Homme qui rit)

112. *Ces singes montés en grade faisaient contrepoids aux hommes **brutalisés et bestialisés**.* (L'Homme qui rit)

Les analyses des fonctions stylistiques du patron « ADJ et ADJ » nous permettent de mettre au jour des caractéristiques stylistiques de Victor Hugo. On remarquera cependant qu'aucune phrase de *Claude Gueux* ne correspond aux catégories identifiées ici. Les phrases avec le patron « ADJ et ADJ » dans ce roman possèdent rarement d'effets stylistiques.

113. *L'affaire était terrible, **droite et simple**.* (Claude Gueux)

114. *En général, quand une catastrophe privée ou publique s'est écroulée sur nous, si nous examinons, d'après les décombres qui en gisent à terre, de quelle façon elle s'est échafaudée, nous trouvons presque toujours qu'elle a été aveuglément construite par un homme **médiocre et obstiné** qui avait foi en lui et qui s'admirait.* (Claude Gueux)

Dans les deux exemples d'au-dessus, le patron « ADJ et ADJ » a toujours la fonction sémantique d'expansion, mais les adjectifs en question ne sont ni des synonymes ni des antonymes, ils ne sont pas employés pour créer une figure de style, et aucun jeu sur les sons n'est présent.

Ceci nous amène à conclure que le patron « ADJ et ADJ » permet de caractériser le style de Hugo, et que le style « hétérogène » de l'auteur se manifeste plus fréquemment dans les deux derniers romans (*L'Homme qui rit* et *Quatrevingt-treize*) que dans *Claude Gueux*, publié avant l'exil.

- « à le NN de le NN »

Ce patron prépositionnel remplit principalement deux fonctions grammaticales dans les romans de Victor Hugo. Il est sur-employé dans deux des romans de l'auteur : 96 occurrences dans *Le Dernier Jour d'un condamné* et 86 occurrences dans *Quatrevingt-treize*.

-Le complément d'objet indirect

115. Ce patron peut jouer le rôle du complément d'objet indirect :

116. *Quand l'homme est ignorant, quand le désert est visionnaire, l'obscurité de la solitude s'ajoute à l'obscurité de l'intelligence ;*

117. *Du dehors, le flot battant le navire répondait **aux chocs du canon** par des coups de mer.*

118. *Quand ces apprêts furent terminés, un monsieur brodé en argent, qu'on appelait monsieur l'inspecteur, donna un ordre **au directeur de la prison** ;*

-La construction d'une locution prépositive

119. Il peut aussi s'agir d'une locution prépositive :

120. *Cimourdain croyait, dans son ingénuité implacable, que tout est équité **au service du vrai** ;*

121. *À l'insu de la famille, a-t-on besoin de permission pour créer une intelligence, une volonté et une droiture ?*

122. *Et, cessant de sourire, il reprit d'un accent péremptoire : Danton, rendez compte des trente-trois mille écus, argent sonnante, que Montmaurin vous a payés **au nom du roi**, sous prétexte de vous indemniser de votre charge de procureur au Châtelet.*

123. *Au pied de l'éminence, le regard se perdait tout de suite sous les arbres.*

Malgré ses fonctions grammaticales dans les phrases, ce patron prépositionnel ne nous permet pas de relever de caractéristiques générales du style de Hugo, nous avons choisi d'en parler malgré pour insister sur le fait que tous les patrons spécifiques pour un roman ne sont pas interprétables stylistiquement. Le patron « à le NN de le NN » est effectivement interprétable d'un point de vue grammatical, mais possède rarement une fonction stylistique. Un autre exemple est le patron « VB son NN », apparu 18 fois dans *Bug-Jargal*. Il est spécifique d'après le résultat de l'AFC, mais ce patron donne en fait très peu d'informations stylistiques, d'une part en raison de sa fréquence moins élevée, et d'autre part parce sa construction syntaxique (verbe + complément d'objet direct) est moins susceptible d'intégrer des phénomènes stylistiques spécifiques.

Il existe également d'autres patrons sur-employés dans nos résultats qui ne possèdent pas de fonctions stylistiques généralisables, à cause de leur fréquence relativement peu élevée. Par exemple, le patron « le NN de NN et de NN » est apparu seulement quatre fois dans *Bug-Jargal*, il est donc difficile d'en tirer des conclusions

claires.

Cela ne veut cependant pas dire que ces patrons ne sont pas importants pour les recherches stylistiques : les patrons qui ne possèdent pas de fonctions stylistiques peuvent être statistiquement identificatoires. La caractérisation du style littéraire doit prendre en compte à la fois la tâche d'identification et celle de caractérisation.

L'analyse qualitative des patrons spécifiques telle que nous l'avons menée dans cette partie nous révèle les caractéristiques stylistiques de Victor Hugo. Nous constatons que les patrons sont capables de représenter le style « hétérogène » de l'auteur, mais un seul patron n'est jamais suffisant pour caractériser le style. C'est pour cette raison que nous présenterons une étude générale sur l'ensemble des patrons spécifiques de Hugo d'après les résultats de l'AFC, afin de relever et regrouper les caractéristiques stylistiques qui existent dans ces patrons.

4.5.2.2 Les caractéristiques stylistiques des patrons syntaxiques

La dernière partie de notre étude sur les patrons syntaxiques est consacrée à l'analyse de toutes les occurrences des patrons spécifiques dans les romans de Hugo afin de révéler leurs caractéristiques qualitatives. Nous regroupons dans cette section toutes les caractéristiques des patrons spécifiques, y compris ceux dont la fréquence est relativement faible et ainsi insuffisante pour donner une conclusion claire. Nous essayons de montrer la capacité des patrons identificatoires à caractériser un style littéraire. Pour vérifier les phénomènes « hétérogènes » principalement dans les romans d'après l'exil de Victor Hugo, nous nous concentrons sur l'étude des patrons spécifiques des quatre derniers romans de Hugo.

-le rapprochement sonore

Une des caractéristiques courantes des patrons spécifiques aux quatre derniers romans de Hugo est le jeu sur les sons. L'exemple suivant montre un rapprochement sonore par l'homéotéleute dans le patron « de NN et de NN » :

*124. Monsieur, dit le petit savoyard, avec cette confiance de l'enfance qui se compose **d'ignorance et d'innocence**, ma pièce ? (Les Misérables)*

« Ignorance » et « innocence » créent une redondance sonore qui attire l'attention du lecteur sur les qualités de « cette confiance ». La mise en relief est également renforcée par l'insertion de la proposition de subordination assez longue (« avec cette confiance de l'enfance qui se compose d'ignorance et d'innocence ») qui interrompt le discours rapporté direct.

*125. En présence de la création fleurie, embaumée, **aimante et charmante**, le ciel splendide inondait d'aurore la Tourgue et la guillotine, et semblait dire aux hommes : Regardez ce que je fais et ce que vous faites. (Quatrevingt-treize)*

L'exemple ci-dessus marque le même effet de mise en relief par rapprochement sonore. L'auteur joue sur la prononciation de « aimante et charmante ». Les deux mots en conjonction ont non seulement une prononciation similaire, mais aussi le même nombre de syllabes.

*126. C'était on ne sait quel étrange **habitant**, **l'habitant** de la nuit. (L'Homme qui rit)*

Le rapprochement sonore peut aussi se faire par la simple répétition du même mot. L'exemple juste cité montre une occurrence du patron « le NN de le NN » : « l'habitant de la nuit ». Ce patron fait partie d'une épanalepse : la reprise littérale du mot sert d'amplificateur au syntagme « étrange habitant », en donnant une précision sur la qualité de cet habitant : « habitant de la nuit ».

-la métaphore

La métaphore hugolienne, comme nous l'avons évoqué plusieurs fois dans les parties précédentes, est très caractéristique de son style. En étudiant les patrons identificatoires dans les romans de l'exil de Hugo, nous avons repéré de nombreuses occurrences de cette figure de style.

*127. Le prodige du regard d'en haut s'opéra, la douce aveugle lumineuse, sans autre effort que sa présence, dissipa toute l'ombre en lui, **le rideau de nuage** s'écarta de cet esprit comme tiré par une main invisible, et Gwynplaine, enchantement céleste, eut dans la conscience une rentrée d'azur. (L'Homme qui rit)*

L'exemple ci-dessus montre une métaphore au sein du patron « le NN de NN » : « le rideau de nuage ». Les deux mots concrets (« rideau » et « nuage ») sont employés pour désigner et concrétiser le changement d'émotion dramatique de Gwynplaine.

L'exemple ci-dessous montre une métaphore construite avec le patron « le NN de le NN ».

*128. La loi pratiquait **la cautérisation du vagabondage**. (L'Homme qui rit)*

L'exemple met en relation deux mots avec un sémantisme très particulier : « cautérisation » et « vagabondage ». La combinaison qui se trouve dans cette phrase est tellement étrange que l'interprétation en est incertaine. Cette incertitude est une des manifestations de l'hétérogénéité dans le style de Hugo.

129. La turgescence du flot indique un étranglement ; dans le brouillard, de certains bourrelets de l'eau signalent un détroit. (L'Homme qui rit)

« La turgescence du flot » rapproche le sémantisme médical au sémantisme marin. L'effet de cette métaphore est la personnification de la mer, comme si celle-ci était victime d'une certaine pathologie.

-l'antithèse

L'antithèse est présente dans toutes nos analyses qualitatives des patrons caractérisants, elle est également courante dans les patrons identificatoires des romans d'après l'exil.

130. Il faut qu'il descende, le cœur plein de charité et de sévérité à la fois, comme un frère et comme un juge, jusqu'à ces casemates impénétrables où rampent pêle-mêle ceux qui saignent et ceux qui frappent, ceux qui pleurent et ceux qui maudissent, ceux qui jeûnent et ceux qui dévorent, ceux qui endurent le mal et ceux qui le font. (Les Misérables)

Le patron « de NN et de NN » dans cette phrase met en relation deux mots contradictoires (« de charité et de sévérité »). Les deux mots correspondent aux rôles du « frère » et du « juge », deux rôles avec eux-mêmes des caractéristiques opposées.

131. *On croit entendre le bruit que fait le silence du sépulcre en se brisant.*

(L'Homme qui rit)

Pour cette phrase, l'auteur aurait pu écrire « le bruit que fait le sépulcre en se brisant », mais Hugo insère le mot « silence » au sein du patron « le NN de le NN » (« le silence du sépulcre ») afin de créer une antithèse entre le bruit et le silence. L'antithèse met en évidence le contraste entre le mouvement et le résultat (le sépulcre qui se brise).

Les deux phrases suivantes comprennent des antithèses au sein du patron « ADJ et ADJ ».

132. *Il le pressentait, il l'entrevoyait d'avance, il le devinait **effrayant et magnifique** ;* (Quatrevingt-treize)

133. *Ainsi grondait le camp **victorieux et mécontent**.* (Quatrevingt-treize)

Les patrons « ADJ et ADJ » dans les deux phrases comprennent chacun une paire d'adjectifs antithétiques : « effrayant et magnifique » et « victorieux et mécontent ». Les antithèses entre les adjectifs donnent un effet de surprise, tout en ouvrant la possibilité à diverses interprétations possibles.

134. *Nous avons un devoir : travailler à l'âme humaine, défendre le mystère contre le miracle, **adorer l'incompréhensible** et rejeter l'absurde, n'admettre, en fait d'inexplicable, que le nécessaire, assainir la croyance, ôter les superstitions de dessus la religion ; écheniller Dieu.* (Les Misérables)

Ce dernier exemple est une antithèse construite sous la forme d'un patron verbal, « , VB le NN » (« , adorer l'incompréhensible »). Nous voulons montrer que l'antithèse

ne se limite pas au rapprochement de deux noms ou adjectifs contradictoires, elle peut également se manifester par deux idées contradictoires. Dans notre exemple, « adorer » est un verbe souvent suivi d'un complément d'objet direct sémantiquement mélioratif : combiner « adorer » avec « l'incompréhensible » crée un effet de surprise qui semble prendre le parti de l'athéisme, vers la fin de la phrase.

-la répétition

Le style répétitif de Victor Hugo peut aussi être constaté par le truchement des patrons identificatoires. Par exemple, pour le patron « de NN et de NN » :

*135. On sentait sous la création l'énormité de la source ; dans tous ces souffles pénétrés d'amour, dans ce va-et-vient **de réverbérations et de reflets**, dans cette prodigieuse dépense de rayons, dans ce versement indéfini d'or fluide, on sentait la prodigalité de l'inépuisable ; et, derrière cette splendeur comme derrière un rideau de flamme, on entrevoyait Dieu, ce millionnaire d'étoiles. (Les Misérables)*

L'occurrence de ce patron (« de réverbérations et de reflets ») se trouve dans une longue énumération grandiose derrière laquelle apparaît le mot Dieu. L'accumulation et l'énumération renforcent l'idée de « splendeur ».

*136. Il le prit, traversa la chambre à grands pas sans précaution et sans s'occuper du bruit, **gagna la porte**, rentra dans l'oratoire, **ouvrit la fenêtre**, saisit un bâton, enjamba l'appui du rez-de-chaussée, mit l'argenterie dans son sac, **jeta le panier**, **franchit le jardin**, sauta par-dessus le mur comme un tigre, et s'enfuit. (Les Misérables)*

L'exemple ci-dessus montre le patron « , VB le NN » en énumération. L'agencement de plusieurs propositions courtes découpe la phrase selon un rythme court et saccadé, afin de créer une tension et une urgence (pendant la fuite du protagoniste).

*137. Allez, philosophes, enseignez, éclairez, allumez, pensez haut, parlez haut, courez joyeux au grand soleil, fraternisez avec les places publiques, **annoncez les bonnes nouvelles, prodiguez les alphabets, proclamez les droits, chantez les Marseillaises, semez les enthousiasmes, arrachez des branches vertes aux chênes.***

(Les Misérables)

Ce dernier exemple montre une phrase construite presque uniquement à base d'énumérations et d'accumulations. Les propositions verbales courtes listent les missions des philosophes. Les effets de redondance et de répétition sont éminents et représentatifs.

Nous concluons d'après nos analyses quantitatives et qualitatives sur les patrons identificatoires que :

1. Les analyses statistiques sur les patrons syntaxiques hiérarchiques permettent de mettre en avant le degré de ressemblance entre les œuvres, elles nous permettent aussi de relever les patrons spécifiques pour chaque roman.

2. Les analyses qualitatives de chaque patron et de l'ensemble des patrons spécifiques montrent que les caractéristiques stylistiques (la métaphore, la répétition, l'antithèse) qu'évoquent les critiques littéraires sur le style de Hugo sont aussi très présentes dans nos résultats, mettant en évidence la capacité des patrons identificatoires à caractériser le style littéraire.

La notion d'hétérogénéité du style hugolien se manifeste à travers plusieurs niveaux linguistiques : ainsi, comme nous l'avons dit à plusieurs reprises, un seul

patron n'est jamais suffisant pour identifier les phénomènes hétérogènes dans les textes. Nous avons essayé de montrer, par le truchement des analyses quantitatives et qualitatives, que les patrons identificatoires, malgré leur manque de fonction stylistique propre, sont un outil performant pour caractériser un style littéraire quand l'étude met parallèlement en œuvre une vérification minutieuse des usages de l'auteur en contexte, dans son œuvre. Cette capacité à caractériser un style littéraire ne se fonde pas sur un ou deux patrons spécifiques, mais sur l'ensemble des patrons identifiés. Il est ainsi possible d'adopter un point de vue global en se servant de cet outil linguistique.

4.6 Conclusion

L'étude du style de Victor Hugo à travers ses neuf romans confirme ce que disait Stanley Fish sur les démarches stylistiques : la caractérisation du style ne se fait que rarement par un seul paramètre linguistique. Nous avons démontré, par le truchement des analyses quantitatives et qualitatives, que la combinaison des patrons caractérisants et des patrons identificatoires donne des résultats satisfaisants qui, non seulement confirment ce que disent les critiques littéraires sur le style hétérogène de Hugo dans la prose, mais les complètent aussi, grâce à la mise au jour de caractéristiques statistiques et formelles qui n'avaient pas encore fait l'objet d'étude systématique.

La méthode que nous avons présentée dans la section 4.3, pour mesurer la complexité syntaxique, prend en compte la complexité des phrases (en termes de nombre de propositions de subordination) par l'identification de la profondeur de l'arbre syntaxique de dépendance ; la largeur, à son tour, mesure la complexité en termes de nombre de propositions coordonnées ou juxtaposées. La combinaison de ces deux mesures permet d'obtenir un indice indicatif de la complexité syntaxique (que nous avons appelé LPIndice) dans les neuf romans de Victor Hugo. Nous constatons que les phrases hugoliennes deviennent de moins en moins complexes au fur du temps, et en particulier dans ses derniers écrits en prose.

L'analyse des patrons caractérisants montre l'évolution de l'utilisation des figures de style (telles que l'inversion, le clivage et la répétition) dans les romans de Victor Hugo. Nous avons constaté plusieurs traits caractéristiques, comme le fait que *Les Misérables* et *Notre-Dame de Paris* possèdent une forte ressemblance en ce qui concerne l'utilisation de ces trois figures de style, ou que le premier roman de Victor Hugo, *Bug-Jargal*, est le roman qui comprend le moins de figures de style. Les résultats sur *Le Dernier Jour d'un condamné* montrent la capacité des patrons syntaxiques à différencier les genres littéraires. Bien que les patrons caractérisants eux-mêmes sont peu informatifs pour caractériser le style de Hugo, le retour au texte montre qu'on

retrouve, dans leur construction syntaxique même, des phénomènes linguistiques (l'antithèse, la métaphore et la répétition) qui confirment le style hétérogène de Hugo. De plus, d'autres caractéristiques qui n'ont pas été discutées dans les critiques littéraires jusqu'ici, à notre connaissance, comme la mise en place du suspense ou le rapprochement sonore, sont également détectables grâce aux patrons caractérisants.

Les patrons identificatoires complètent notre étude sur les patrons caractérisants. Ces patrons sont également performants pour caractériser le style de Hugo. Les calculs de spécificité par l'Analyse factorielle des correspondances (AFC) révèlent les ressemblances entre les romans de Hugo, ils mettent également en évidence les patrons dits « identificatoires » pour chaque roman. Le retour au texte pour examiner les contextes d'emploi des patrons spécifiques est fructueux : il permet de repérer les caractéristiques stylistiques qui peuvent exister dans les patrons identificatoires. Ainsi, nous avons réussi à repérer les structures antithétiques, métaphoriques et répétitives, c'est-à-dire les principales manifestations formelles du style hétérogène de Hugo. D'autres caractéristiques de nature sonore et syntaxique ont également été mises en avant.

Le style d'un auteur, et de Hugo en particulier, est par nature un sujet vaste et instable car peu formalisé, et des notions comme l'hétérogénéité du style sont intrinsèquement liées à plusieurs niveaux d'analyse linguistique. Les démarches stylistiques ne sont jamais exhaustives : ainsi, un seul patron ou une seule catégorie de patrons est rarement capable de saisir toutes les caractéristiques stylistiques d'un auteur. Notre étude montre que la combinaison des patrons caractérisants et des patrons identificatoires peut donner une vue générale sur le style de Hugo, et que la combinaison des analyses quantitatives et des analyses qualitatives aide à cerner l'évolution des phénomènes stylistiques, ainsi que les caractéristiques stylistiques des patrons, qui doivent donc être examinés et interprétés minutieusement. Nous concluons que les patrons syntaxiques, par le truchement des analyses automatiques et manuelles, constituent un outil performant pour caractériser le style littéraire.

Chapitre 5 Caractérisation du style des quatre auteurs du 19^{ème} siècle de notre corpus par la méthode des patrons syntaxiques

5.1 Introduction

Nous présentons dans ce chapitre l'application de nos méthodes d'extraction supervisée et non-supervisée de patrons syntaxiques aux romans des quatre auteurs de notre corpus : Chateaubriand, Flaubert, Hugo et Stendhal. L'objectif de cette étude consiste à démontrer la capacité des patrons syntaxiques à caractériser le style littéraire. La caractérisation du style de chacun des quatre auteurs combine l'analyse statistique et l'analyse littéraire : il s'agit, d'une part, d'identifier et de différencier les auteurs, et d'autre part d'interpréter la fonction stylistique de ces patrons syntaxiques. Nous insistons sur le fait qu'un seul patron n'est jamais suffisant pour caractériser le style littéraire, mais l'ensemble des patrons identificatoires et caractérisants donne une vue générale du style d'un auteur.

La section 5.2 présente un bref aperçu du style de Chateaubriand, de Flaubert et de Stendhal d'après ce qu'en ont dit les critiques littéraires. Le style de Victor Hugo a déjà été discuté dans la section 4.2 du Chapitre 4. Ce parcours vise à identifier les particularités essentielles du style de chacun des trois auteurs : le style fait d'archaïsmes et d'innovation de Chateaubriand, l'harmonie du style de Flaubert ainsi que l'opposition du style romanesque de Stendhal, par rapport à ses contemporains. Les matérialisations formelles de chaque style sont aussi présentées comme la soudure entre les études littéraires et nos démarches formelles.

Nous commençons toujours par mesurer la complexité syntaxique des phrases des

quatre auteurs (sur la base de leurs romans). La méthode comprend les paramètres tels que la « profondeur », la « largeur » et le « LPIndice », discutés dans la section 3.5 (Chapitre 3). Ces paramètres permettent de donner une caractérisation générale de chaque auteur. Si les résultats confirment ce qui a été constaté sur les phrases de Chateaubriand, Flaubert et Hugo d'après les critiques littéraires, les résultats sur l'œuvre de Stendhal, qui manifestent une complexité syntaxique très élevée, contredisent l'idée en vogue d'un auteur au style clair et naturel²⁵³ (Talbot, 1985).

Dans la section 5.4, nous explorons les trois figures de style (l'inversion du sujet, le clivage et la répétition) que notre méthode d'extraction supervisée nous permet de relever dans le corpus. Les analyses statistiques de ce chapitre concernent tous les romans de nos quatre auteurs, tandis que les analyses qualitatives ne portent que sur Chateaubriand, Flaubert et Stendhal. L'étude qualitative des romans de Victor Hugo se trouve dans le chapitre précédent.

La section 5.5 est consacrée à la présentation des patrons identificatoires. Les patrons identificatoires complètent l'étude des figures de style syntaxiques en fournissant la possibilité d'exploiter globalement les spécificités de chaque corpus sur la base de la syntaxe. La spécificité des patrons pour chaque auteur, extraits sans discrimination, est calculée par l'Analyse factorielle des correspondances (AFC). Une étude sur tous les patrons spécifiques est également présentée pour montrer la capacité de ces patrons à relever les caractéristiques stylistiques propres à chaque auteur. Comme en 5.4, l'analyse qualitative sur l'œuvre de Victor Hugo n'est pas présente dans cette partie pour ne pas répéter ce que nous avons déjà montré en détail dans le chapitre 4. La combinaison des patrons identificatoires et caractérisants, et l'étude générale de la complexité syntaxique nous permettent de caractériser, d'une manière globale, le style de chaque auteur étudié.

²⁵³ Talbot, Emile J. 1985. *Stendhal and Romantic Esthetics*. French Forum Monographs 61. Lexington, Ky: French forum.

5.2 Les styles romanesques de Chateaubriand, Flaubert et Stendhal

Dans cette section, nous essayons de résumer brièvement les études critiques sur les styles romanesques de Chateaubriand, Flaubert et Stendhal. Le style romanesque de Victor Hugo a été présenté dans la section 4.2 du Chapitre 4.

5.2.1 Aperçu sur le style de Chateaubriand

Comme tous les auteurs de notre corpus, le style de Chateaubriand ne peut pas se résumer en quelques mots. Le style est une notion holiste qui prend en compte toutes les créations d'un auteur et les particularités d'un style concernent tous les niveaux linguistiques. Ce que nous essayons de présenter dans cette section est une étude générale sur le style de Chateaubriand, à partir d'exemples concrets.

D'après Sainte-Beuve, « toute l'école moderne émane plus ou moins directement de lui [Chateaubriand]²⁵⁴ » (Sainte-Beuve, 1926, p. 170). En effet, l'influence de Chateaubriand sur la littérature française est remarquable. À quatorze ans, Victor Hugo exprimait son admiration pour cet auteur en écrivant : « je veux être Chateaubriand ou rien²⁵⁵ » (Hovasse, 2001, p. 142). Barrès constatait dans le style de Chateaubriand « une beauté la plus singulière du grand style voluptueux²⁵⁶ » (Barrès, 1931, t. III, p. 189). Madame de Beaumont disait : « Le style de Chateaubriand me fait éprouver une espèce de frémissement d'amour : il joue du clavecin sur toutes les fibres²⁵⁷ » (Sainte-Beuve, 1861 [1843], t. I, p. 216).

²⁵⁴ Sainte-Beuve. 1926. *Mes poisons: cahiers intimes inédits*. Édité par Victor Giraud. Paris, France: Plon-Nourrit et Cie, impr.1926, p. 170.

²⁵⁵ Cité par Hovasse, Jean-Marc. 2001. *Victor Hugo : 1802-1851 , I: Avant l'exil. Victor Hugo*. Paris: Fayard, p. 142.

²⁵⁶ Barrès, Maurice. 1931. *Mes cahiers*. Paris, France: Librairie Plon, t. III, p. 189.

²⁵⁷ Sainte-Beuve, Charles-Augustin. 1861 [1843]. *Chateaubriand et son groupe littéraire sous l'Empire: cours professé à Liège en 1848-1849*. 2 vol. Paris, France: Garnier frères, t. I, p. 216.

Pour donner une idée générale sur le style de Chateaubriand, nous considérons que son style est un mélange d'archaïsme et d'innovation. Selon Sainte-Beuve, « il y a innovation et rénovation du langage en même temps que tradition [dans le style de Chateaubriand]²⁵⁸ » (Sainte-Beuve, 1926). Il continue et explique que :

« Il [le style de Chateaubriand] est sobre dans son audace... M. de Chateaubriand apparaît donc littérairement comme un de ces écrivains qui maintiennent une langue en osant la remuer et la rajeunir²⁵⁹ » (Sainte-Beuve, 1926).

Ainsi, deux notions, l'archaïsme et l'innovation, sont-elles importantes pour comprendre le style de Chateaubriand. Nous explorons ensuite les manifestations formelles de ces deux caractéristiques stylistiques dans les textes de cet auteur. Brunot et al.²⁶⁰ (1948) ont étudié systématiquement les phénomènes formels du style de Chateaubriand : les éléments suivants sont les phénomènes formels relevés dans l'étude de Ferdinand Brunot et Charles Bruneau.

-L'archaïsme

Le style archaïque de Chateaubriand se manifeste aux niveaux lexical et syntaxique :

- ◆ L'utilisation de l'adjectif relatif « lequel » au lieu de « qui » dans les phrases en situation de subordination ;
- ◆ Le présent inattendu dans le récit qui est majoritairement au passé ;
- ◆ La substitution de « aux » par « dans les » ;
- ◆ Le verbe à l'infinitif substantivé

²⁵⁸ Sainte-Beuve. 1926. *Op. cit.*.

²⁵⁹ Sainte-Beuve. 1926. *Op. cit.*.

²⁶⁰ Brunot, Ferdinand, et Charles Bruneau. 1948. *L'époque romantique. 1 vol. Histoire de la langue française, des origines à 1900* 12. Paris: A. Colin.

Ce sont quatre des phénomènes emblématiques de Chateaubriand.

-L'innovation

L'innovation chez Chateaubriand se manifeste par :

- ◆ Les noms abstraits au pluriel : des « immortalités », des « oisivetés », des « mysticités », etc.
- ◆ Les adjectifs et participes peuvent être substantivés : « les dignes », « le souffrant », etc.
- ◆ Les antithèses, une figure de style que Victor Hugo a sans doute hérité de Chateaubriand. Sainte-Beuve constatait que « l'antithèse chez lui [Chateaubriand] était devenue un tic, une vraie manie²⁶¹ » (Sainte-Beuve, 1861, t. I, p. 397).
- ◆ L'image, la figure de style d'analogie, est également très présente dans ses écrits.

-La phrase de Chateaubriand

Nous nous concentrons ensuite sur les phrases de Chateaubriand, parce que notre méthode se base sur la syntaxe.

L'étude de Ferdinand Brunot et Charles Bruneau constate que la répétition dans les phrases de Chateaubriand est « discrète et nuancée²⁶² » (Brunot et al., p. 324) : « l'ensemble [des répétitions] est homogène, avec une gradation bien marquée²⁶³ » (Brunot et al., p. 324). Rappelons que les répétitions hugoliennes sont abondantes.

Jean Mourot examine systématiquement les phrases de Chateaubriand et constate qu'il existe une structure syntaxique privilégiée de l'auteur :

²⁶¹ Sainte-Beuve, Charles-Augustin. 1861. *Op.cit.*, p. 397.

²⁶² Brunot, Ferdinand, et Charles Bruneau. 1948. *Op. cit.*, p. 324.

²⁶³ *Idem.*

« Lorsque Chateaubriand se laisse aller à son rythme, il construit des phrases dont l'apodose paraît d'autant plus étirée que la protase est plus réduite, des phrases à « traîne »²⁶⁴ » (Mourot, 1960, p. 286).

Dans cette citation, l'« apodose » renvoie à une proposition principale placée après une proposition conditionnelle appelée protase. La phrase possède donc une protase courte et une apodose souvent longue. Mourot explique ensuite les caractéristiques de la protase et de l'apodose dans cette structure syntaxique privilégiée de Chateaubriand :

« Les éléments brefs antéposés mettent en valeur, par le contraste des volumes et de l'intonation, la retombée et l'étalement progressifs de l'apodose, à laquelle est réservé d'autre part l'essentiel de la communication ; ils apparaissent ainsi comme une préparation, un avertissement. De là vient, entre autres raisons, l'air de solennité qu'ont tant de phrases de Chateaubriand, fondées sur ce tour et caractérisées par la longueur de l'apodose²⁶⁵ » (Mourot, 1960, p. 273-274).

Les tours favorables de Chateaubriand pour rallonger son apodose sont : « la relative en position finale, la proposition comparative en position finale (presque toujours introduite par *comme*, la proposition temporelle en positions finales (particulièrement avec *tandis que*)²⁶⁶ » (Mourot, 1960, p. 286).

Le style de Chateaubriand se situe donc entre l'archaïsme et l'innovation. Comme Victor Hugo, cet auteur est représentatif du point de vue romantique souhaitant mêler tradition littéraire et volonté de rénover la langue²⁶⁷ (Talbot, 1985). Évidemment, notre

²⁶⁴ Mourot, Jean. 1960. « Le Génie d'un style, Chateaubriand, rythme et sonorité dans les «Mémoires d'Outre-tombe» : Thèse pour le doctorat ès lettres présentée à la Faculté des Lettres de l'université de Paris ». S. l., p. 286.

²⁶⁵ Mourot, Jean. 1960. *Op. cit.*, p. 273-274.

²⁶⁶ *Ibid.*, p. 286.

²⁶⁷ Talbot, Emile J. 1985. *Op. cit.*.

étude n'est pas exhaustive, elle se concentre sur ces quelques éléments pour donner une idée générale du style de Chateaubriand.

5.2.2 Aperçu sur le style de Flaubert

Gustave Flaubert est un des grands écrivains français du 19^{ème} siècle. Contrairement à Victor Hugo, les principales créations littéraires de Flaubert sont en prose. Le rôle important de Flaubert dans la littérature française est indéniable, Albert Thibaudet le considère comme un auteur entre le romantisme et le réalisme :

« Il [Flaubert] a été de ce siècle par tout son être et tout son art, ne l'a débordé en rien. Il était fait pour en donner le tableau et la synthèse, pour en unir intelligemment toutes les puissances romantiques et réalistes²⁶⁸ » (Thibaudet, 1935).

Nous avons évoqué dans le Chapitre 2 la recherche ultime par Flaubert du « style parfait ». En effet, la prose de Flaubert est souvent le produit d'un perfectionnement laborieux, elle est considérée comme « harmonieuse » et « équilibrée²⁶⁹ » (Thibaudet, 1935).

« Il est possible que l'art de la prose tel que l'entendait Flaubert, un art de la prose attentif à des lois musicales aussi rigoureuses que celles des vers, soit propre, dans les temps modernes, à la littérature française, qui a, depuis Guez de Balzac, ses peseurs de mots, ses chercheurs de sonorités, ses créateurs de

²⁶⁸ Thibaudet, Albert. 1935. *Gustave Flaubert*. 4e éd. 1 vol. Paris: Gallimard. [https://fr.m.wikisource.org/wiki/Gustave_Flaubert_\(Thibaudet\)](https://fr.m.wikisource.org/wiki/Gustave_Flaubert_(Thibaudet)), consulté le 25 juin, 2021.

²⁶⁹ *Idem*.

coupes, et qui implique des exigences complexes de perfection, auxquelles les grands écrivains sacrifient plus ou moins péniblement, soit qu'ils les réalisent au courant de la plume, comme Bossuet, soit qu'ils n'y satisfassent que par un long labeur de polissage et de ciselure comme Flaubert²⁷⁰ » (Thibaudet, 1935).

Nous résumons, d'une manière brève, les manifestations formelles du style harmonieux de Flaubert dans cette section. Nous insistons sur le fait que la matérialisation linguistique du style joue le rôle de soudure entre les études littéraires et nos démarches formelles.

-Le « gueuloir » de Flaubert

Flaubert a écrit lui-même les caractéristiques stylistiques qu'il cherche à mettre en évidence dans ses œuvres :

« J'en conçois pourtant un, moi, un style, un style qui serait beau, que quelqu'un fera quelque jour, dans dix ans ou dans dix siècles, et qui serait rythmé comme le vers, précis comme le langage des sciences, et avec des ondulations, des renflements de violoncelle, des aigrettes de feu. Un style qui nous entrerait dans l'idée comme un coup de stylet et où notre pensée enfin voyagerait sur des surfaces lisses comme lorsqu'on file sur un canot avec un bon vent arrière²⁷¹ » (Flaubert, 1852).

Nous constatons que la prose idéale pour Flaubert est une prose à la limite de la poétique, pleine de nuances et de subtilités. Les phrases doivent être « rythmées » et

²⁷⁰ *Idem.*

²⁷¹ Flaubert, Gustave. 1852. « Flaubert à Louise Colet, Croisset, 24 avril 1852 ». <https://flaubert.univ-rouen.fr/jet/public/correspondance/trans.php?corpus=correspondance&id=9920>, consulté, le 20 juin, 2021.

« précises », articulées pour faire passer un message clair et riche de sens. Afin d'atteindre ce style harmonieux et poétique, Flaubert passait beaucoup de temps à lire ses phrases à haute voix, d'où vient la notion de « gueuloir ». Thibaudet constate ainsi le « fond oratoire²⁷² » (Thibaudet, 1935) du style de Flaubert, parce que ses phrases ont toutes subi une lecture à haute voix afin de vérifier leur harmonie. Ainsi, la musicalité et le rythme équilibré peuvent être considérés comme deux des caractéristiques formelles du style de Flaubert.

-l'image

L'image (et l'analogie) est une des figures de style préférées de Flaubert. « Flaubert appartient incontestablement à la race des grands créateurs d'images, et les siennes sont presque toujours visuelles²⁷³ » (Thibaudet, 1935). Les phrases pittoresques lui ont permis de créer de véritables tableaux dans ses romans. Ferrère, l'auteur de *L'Esthétique de Gustave Flaubert*, l'a fort bien mis en lumière : « Lorsque Flaubert travaillait, il composait par tableaux²⁷⁴ » (Ferrère, 1913). L'image est un outil efficace pour Flaubert pour mettre en place un accord parfait entre la description matérielle et les sentiments des personnages, un équilibre entre les milieux et les âmes :

« La comparaison habituelle à Flaubert consiste à essayer de préciser et de faire saisir un sentiment un peu délicat et compliqué en lui donnant une expression et une correspondance physiques²⁷⁵ » (Thibaudet, 1935).

²⁷² Thibaudet, Albert. 1935. *Op. cit.*.

²⁷³ *Idem.*

²⁷⁴ Ferrère, Étienne-Louis. 1913. « L'esthétique de Gustave Flaubert ». Thèse, France: Université de Paris (1896-1968). Faculté des lettres.

²⁷⁵ Thibaudet, Albert. 1935. *Op. cit.*.

-la période ternaire

Si les phrases avec protase brève et apodose longue sont fréquentes dans les romans de Chateaubriand, les études littéraires considèrent souvent que la structure syntaxique privilégiée par Flaubert est la période ternaire (rythme ternaire lorsque des éléments se rapportant à la même idée se suivent par groupe de 3), dont les trois membres sont souvent rangés dans un ordre de grandeur, soit croissante, soit décroissante²⁷⁶ (Thibaudet, 1935).

-La coupe

Les phrases de Flaubert, en raison du « gueuloir » oratoire, sont souvent soigneusement coupées pour créer un rythme équilibré :

« L'oratoire donne à la phrase son mouvement, mais la coupe lui donne son arrêt. [...] L'harmonie, le nombre versent la phrase dans cette réalité collective que sont le tableau, le chapitre ou le livre, mais la coupe fait à la phrase son individualité²⁷⁷ » (Thibaudet, 1935).

-La dissonance de temps

Ce phénomène est d'après Thibaudet, « plus habituel à Flaubert qu'à n'importe quel écrivain. [...] Il en a fait un procédé, qui donne toujours des effets, beaux ou pittoresques²⁷⁸ » (Thibaudet, 1935). Le jeu littéraire sur le temps de narration comprend également l'utilisation des participes (présents ou passés). Nous explorerons les différents effets créés par la rupture et la dissonance de temps dans nos résultats

²⁷⁶ *Idem.*

²⁷⁷ *Idem.*

²⁷⁸ Thibaudet, Albert. 1935. *Op. cit.*

d'analyse des patrons syntaxiques.

5.2.3 Aperçu sur le style de Stendhal

Deux notions sont importantes pour comprendre le style, ainsi que le point de vue sur la stylistique de Stendhal : la clarté et le naturel. Le style de Stendhal se distingue de celui des autres auteurs de son époque par l'opposition, non seulement avec Balzac, mais aussi avec Chateaubriand et Flaubert. Nous présentons la recherche du style clair et naturel de Stendhal en le mettant en opposition avec les autres auteurs de son époque.

La notion de clarté est présente dès les premières œuvres de Stendhal. « Il me semble que la première loi que le dix-neuvième siècle impose à ceux qui se mêlent d'écrire, c'est la clarté²⁷⁹ » (Stendhal, 1914 [1814], p. 5). Pour atteindre cette clarté dans le style, Stendhal se concentre majoritairement sur « l'utilisation d'une langue consacrée et d'un vocabulaire précis²⁸⁰ » (Talbot, 1985, p. 75). La recherche du style clair est, d'après la plupart des critiques littéraires, contradictoire avec le style des autres auteurs de son époque, et surtout avec le romantisme dont l'influence est indéniable au début du 19^{ème}. Victor Brombert constate également la particularité du style de Stendhal par rapport aux romantiques :

« Le lecteur moderne, avide de complexité, se méfie des œuvres dont la lucidité ou la cohérence trop rigoureuse semble trahir une absence de tension et de densité. Stendhal, en dépit de la faveur croissante dont il jouit auprès d'une certaine élite, continue cependant à être loué pour sa clarté et sa limpidité, à être opposé aux romantiques exubérants dont il fut le contemporain, comme si la marque de son génie consistait à avoir su donner

²⁷⁹ Stendhal, et Romain Préfacier Rolland. 1914 [1814]. *Vies de Haydn de Mozart et de Métaïtase*. Édité par Daniel Muller. Paris, France: H. Champion, p. 5.

²⁸⁰ Talbot, Emile J. 1985. *Op. cit.*, p. 75.

au pittoresque alors à la mode une forme cristalline, digne d'un disciple des idéologues²⁸¹ » (Brombert, 1954, p. 149).

Nous avons évoqué dans la première partie de cette section le fait que le style de Chateaubriand est marqué par une tension archaïsme et innovation. En effet, l'innovation des usages et la création de nouvelles tournures littéraires sont des caractéristiques importantes du romantisme. Stendhal se différencie de cette tendance en se référant uniquement à la tradition littéraire. L'innovation est presque interdite sous sa plume. Comme le souligne Émile Talbot dans *Stendhal and romantic esthetics* :

« De toute évidence, Stendhal a pu se dégager de cette position extrême, mais l'idéal d'une langue consacrée sur laquelle les écrivains et leur public seraient d'accord naturellement est celui auquel il a continué à adhérer et à servir de base à son affirmation selon laquelle les écrivains ne devraient pas innover dans le langage²⁸² » (Talbot, 1985, p. 75).

Ainsi, Stendhal se distingue des romantiques tels que Flaubert, Chateaubriand et Hugo dont les innovations (comme le rythme ternaire, l'image et la métaphore) sont toujours en vigueur de nos jours. Aussi, « l'art de Stendhal se trouve-t-il à l'opposé de celui d'un Flaubert et d'un James Joyce²⁸³ » (Brombert, 1954, p. 163). Pour mieux élucider la notion de clarté que cherche Stendhal dans ses écrits, nous pouvons citer une lettre de l'auteur, destinée à Balzac, dans laquelle il défend son point de vue sur la stylistique. En 1840, Balzac émet, dans *La Revue Parisienne*, des critiques sur le roman *La Chartreuse de Parme* de Stendhal, en soulignant précisément que « le côté faible de

²⁸¹ Brombert, Victor, et Institut d'études françaises. 1954. *Stendhal et la voie oblique, l'auteur devant son monde romanesque*, par Victor Brombert. New Haven Paris: Yale University press Presses universitaires de France (Mayenne, impr. de Floch), p. 149.

²⁸² Talbot, Emile J. 1985. *Op. cit.*, p. 75.

²⁸³ Brombert, Victor, et Institut d'études françaises. 1954. *Op. cit.*, p. 163.

cette œuvre est le style²⁸⁴ » (Talbot, 1979, p. 65). Stendhal, pour répondre à ses critiques sur son style, rédige ensuite une lettre dans laquelle il a écrit : « Enfin, ... je suis d'accord sur tout excepté sur le style. N'allez pas croire que ce soit excès d'orgueil. Je ne vois qu'une règle : le style ne saurait être trop clair, trop simple²⁸⁵ » (Stendhal, 1999 [1840], t. III, p. 394). Il ajoute ensuite dans la même lettre la définition d'un style clair : « En composant *la Chartreuse*, pour prendre le ton, je lisais de temps en temps quelques pages du Code civil²⁸⁶ » (Stendhal, 1999 [1840], t. III, p. 399). Il s'agit probablement d'une expression caricaturale, nous ne savons pas si Stendhal prenait vraiment le Code civil comme référence stylistique pour ses créations littéraires, mais nous retrouvons infailliblement la recherche de Stendhal d'une langue précise, efficace et claire.

La deuxième notion cruciale pour comprendre le style de Stendhal est « le naturel ». Le style, d'après Stendhal, doit non seulement être clair, mais aussi naturel. Francine Marill Albérès constate que « le naturel conçu d'abord comme une valeur purement psychologique, a dominé la pensée de Stendhal au point de s'y transformer en valeur morale²⁸⁷ » (Albérès, 1956, p. 15). Stendhal entend par naturel un style dépourvu de figures de style, parce que les figures lui semblent inutiles et « hypocrites ». Le style naturel se définit par l'opposition avec le style « pompeux » (d'après Stendhal) de Chateaubriand :

« On trouve un certain nombre de remarques élogieuses sur le style de Chateaubriand au cours des premières décennies des années 1800, mais une opposition se développe rapidement lorsqu'il [Stendhal] perçoit la grandiloquence de Chateaubriand. Pour lui, le langage pompeux est un signe certain d'hypocrisie (quand il masque une incompréhension du cœur humain).

²⁸⁴ Talbot, Emile J., éd. 1979. *La critique stendhalienne de Balzac à Zola*. York (S.C.), Etats-Unis d'Amérique: French Literature, p. 65.

²⁸⁵ Stendhal. 1999. *Correspondance générale*. Édité par Victor Del Litto. Correspondance générale. Paris, France: H. Champion., t. III, p. 394.

²⁸⁶ Stendhal. 1999. *Op.cit.*, p. 399.

²⁸⁷ Marill-Albérès, Francine. 1956. « Le Naturel chez Stendhal ». Paris, France: Nizet, p. 15.

Stendhal prend ses distances avec le langage poétique, ou plutôt avec la *poésie en prose* qui était devenue la favorite de ses contemporains²⁸⁸ » (Talbot, 1985, p. 87).

Ainsi, par toutes les oppositions entre Stendhal et les autres auteurs de son époque, nous concluons que le style de Stendhal, ou au moins le bon style que l'auteur recherche, est un style clair et naturel, manifesté d'une part par la précision et l'efficacité dans le langage, d'autre part par l'économie des figures de style, d'insistance et d'analogie. Nous allons maintenant procéder à l'application de notre méthode afin de vérifier si la clarté et le naturel peuvent être révélés par les patrons syntaxiques.

Ces résumés brefs des styles romanesques de Chateaubriand, Flaubert et Stendhal révèlent les points de vue des critiques littéraires sur le sujet. L'application de notre méthode consiste à prouver (ou éventuellement dénier) l'existence de ces phénomènes formels dans les romans des trois auteurs.

²⁸⁸ Talbot, Emile J. 1985. *Op. cit.*, p. 87.

5.3 Mesurer la complexité syntaxique dans les romans de Chateaubriand, Flaubert, Hugo et Stendhal

Dans cette partie, nous présentons l'application de notre méthode visant à mesurer la complexité syntaxique des romans de Chateaubriand, Flaubert, Hugo et Stendhal à partir de la structure de leurs phrases. Notre méthode, basée sur les relations de dépendance, a été discutée dans la partie 3.5 (Chapitre 3). La méthode mesure la largeur et la profondeur des arbres syntaxiques de dépendance et combine ces mesures afin de créer un indice numérique représentant la complexité syntaxique de chaque phrase. La largeur de l'arbre syntaxique est souvent liée au nombre de propositions en coordination (et en juxtaposition), alors que la profondeur de l'arbre syntaxique est souvent liée aux propositions en situation de subordination. Le LPIndice regroupe les deux paramètres, simplement en faisant le produit de la largeur et de la profondeur.

Le tableau suivant présente la complexité syntaxique des phrases des quatre auteurs, les valeurs absolues représentent les moyennes calculées à partir de l'ensemble des phrases d'un auteur :

Auteurs	Largeur	Profondeur	LPIndice
Chateaubriand	7.4590138	5.52094972	41.18084
Flaubert	6.7599979	4.6532916	31.456241
Hugo	6.0146864	4.40120889	26.471891
Stendhal	7.4737679	5.51584227	41.224125

Tableau 5.1 Complexité syntaxique moyenne pour les quatre auteurs dans leurs romans.

Les résultats montrent que les phrases de Hugo sont relativement moins complexes que celles des autres auteurs de son époque ; à l'inverse, Stendhal et Chateaubriand ont l'habitude d'utiliser les phrases syntaxiquement plus complexes dans leurs romans. La figure suivante présente la visualisation des différences entre ces quatre auteurs.

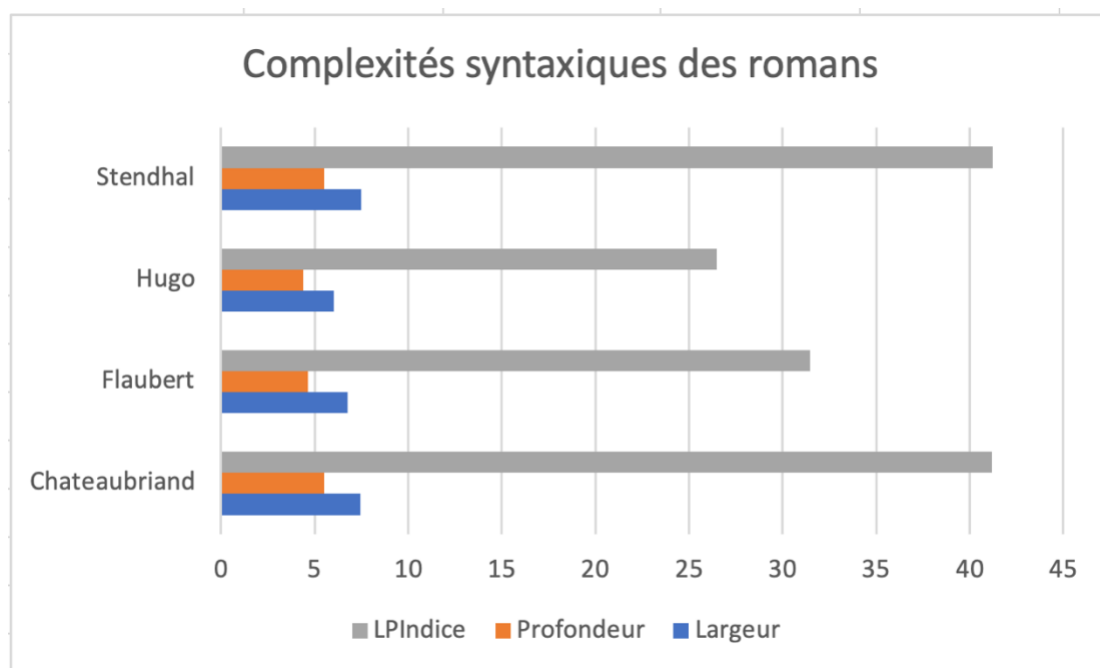


Figure 5.1 Complexité syntaxique moyenne pour les quatre auteurs dans leurs romans.

D'après nos résultats avec les paramètres « largeur », « profondeur » et « LPIndice », nous constatons que Stendhal est un auteur qui préfère les phrases complexes avec la « largeur » moyenne la plus importante parmi les auteurs considérés. On peut aussi imaginer que Chateaubriand utilise davantage de phrases en situation de subordination, vu la « profondeur » moyenne importante obtenue pour cet auteur. Victor Hugo est un auteur qui privilégie souvent les phrases simples.

La méthode que nous proposons permet de donner une vue générale de la syntaxe utilisée par chaque auteur de notre corpus. Les phrases de Hugo sont relativement moins complexes que celles des autres auteurs, mais cela n'empêche pas ses romans d'être « bavards²⁸⁹ » (Wulf, 2014, p. 20) et les lecteurs y trouvent énormément de digressions. Nous avons montré dans le chapitre précédent que le style « hétérogène » de Hugo se manifeste dans plusieurs niveaux linguistiques et la complexité syntaxique n'est qu'une de ses manifestations. Selon Jean Mourot, les phrases de Chateaubriand adoptent souvent le rythme ternaire croissant avec un rallongement de l'apodose afin de créer un

²⁸⁹ Wulf, Judith. 2014. *Op.cit.*, p. 20.

déséquilibre syntaxique ²⁹⁰ (Mourot, 1960). Ce rallongement se manifeste effectivement dans nos résultats par une « largeur » et une « profondeur » supérieures à la moyenne dans les romans de Chateaubriand. Notre méthode arrive à confirmer la tendance de Chateaubriand à rallonger l'apodose de ses phrases, mais la complexité syntaxique n'est qu'une des manifestations concrètes du style de Chateaubriand.

La discussion des résultats nous ramène au problème que nous avons évoqué dès le début de la thèse : les outils informatiques nous permettent de révéler des caractéristiques syntaxiques qui ne sont pas faciles à repérer manuellement par les critiques littéraires. Certaines caractéristiques conforment à ce que disent ces critiques littéraires sur le style d'un auteur (la complexité syntaxique importante des romans de Chateaubriand), d'autres contredisent les critiques littéraires (la complexité syntaxique élevée de Stendhal, en contradiction semble-t-il avec sa recherche d'un style clair et naturel). Nous essayons de montrer que les caractéristiques que nous avons remarquées grâce à notre méthode peuvent être un bon point de départ, avant une exploitation plus profonde de la syntaxe, présentée dans les parties suivantes.

²⁹⁰ Mourot, Jean. 1960. *Op. cit.*.

5.4 Application de la méthode d'extraction supervisée

5.4.1 L'inversion du sujet

Après avoir mesuré la complexité syntaxique dans les romans des quatre auteurs de notre corpus, nous procédons dans cette partie à l'analyse des figures de style fondées sur la syntaxe dans les œuvres de Chateaubriand, Flaubert, Hugo et Stendhal.

La première partie de cette application est consacrée à l'inversion du sujet. Dans la section 3.6.1 (Chapitre 3), nous avons présenté la méthode d'extraction des inversions du sujet ainsi que le nettoyage des résultats pour ne garder que les inversions stylistiques. Dans cette partie, nous suivons la même organisation textuelle qu'en 4.4.1 (Chapitre 4) pour présenter l'analyse quantitative des typologies et l'analyse qualitative des occurrences des inversions du sujet dans les romans des trois auteurs (les inversions du sujet de Victor Hugo ont été discutées en détail dans 4.4.1 du Chapitre 4). Les analyses combinées, à la fois statistiques et littéraires, nous permettent de montrer la capacité de l'inversion à détecter des phénomènes stylistiques propres à chaque auteur.

5.4.1.1 Analyses quantitatives des inversions du sujet

Nous avons trouvé 87 occurrences d'inversions du sujet stylistiques dans les romans de Chateaubriand, 95 occurrences pour Flaubert, 110 occurrences pour Hugo et 48 occurrences pour Stendhal. Les occurrences ne sont donc pas importantes sur le plan statistique, parce que la plupart des inversions ne sont pas stylistiques²⁹¹ (et ne sont donc pas prises en compte ici). À l'exception des incises, qui sont très présentes dans

²⁹¹ Pour les critères de nettoyage des résultats, voir 3.6.1.

nos résultats, la grande majorité des inversions non-stylistiques apparaissent dans les subordinations relatives. Nous avons fait une classification des inversions selon les déclencheurs qu'elles possèdent, placés devant la construction inversée verbe-sujet. La modélisation des inversions prend la forme suivante : Déclencheur+Verbe+Sujet. Dix grandes catégories se trouvent dans nos résultats :

1. Groupe prépositionnel+Verbe+Sujet

138. À l'extrémité opposée du pont sur le Houblon, c'est-à-dire du côté de Paris, se trouvait la jolie église gothique du pays ; (Lamiel, Stendhal)

2. Groupe adverbial+Verbe+Sujet

139. Là passaient les quadrilles superbement vêtus de brocart, là s'avançaient les galères chargées d'armes et de fleurs, les dragons qui lançaient des feux et qui recélaient dans leurs flancs d'illustres guerriers, ingénieuses inventions du plaisir et de la galanterie. (Les Aventures du dernier Abencerage, Chateaubriand)

3. Groupe mis en relief+Verbe+Sujet

140. Les dames se demandaient si c'était du maire tout seul que provenait cette haute inconvenance. (Le Rouge et le Noir, Stendhal)

4. Concessif avec inversion du sujet

141. Tu es un drôle avec qui j'ai démangeaison de ripailler, dût-il m'en coûter un douzain neuf de douze tournois. (Notre-Dame de Paris, Victor Hugo)

5. Verbe au début de la phrase+Sujet

142. Viennent à moi ceux qui sont couverts de vin, ceux qui sont couverts de boue, ceux qui sont couverts de sang ; (La Tentation de Saint-Antoine, Flaubert)

6. Auxiliaire+Verbe+Sujet

143. On voit sur leurs têtes se mouvoir leurs aigrettes de crin, qu'agitent les mouvements du coursier retenu avec peine dans le rang de ses compagnons. (Les Natchez, Chateaubriand)

7. Locution conjunctive+Verbe+Sujet

144. À mesure qu'augmentait leur ivresse, ils se rappelaient de plus en plus l'injustice de Carthage. (Salammbô, Flaubert)

8. « Que » exclamatif+Verbe+Sujet

145. Que ne suis-je un de ceux dont l'âme est toujours intrépide et l'esprit ferme, —comme le grand Athanase, par exemple. (La Tentation de Saint-Antoine, Flaubert)

9. Inversion complexe

146. Sans le vouloir, Armance lui laissait-elle voir, par un mot, qu'elle ne le regardait plus comme un ami intime, son cœur se serrait, il en perdait la parole pour un quart d'heure. (Armance, Stendhal)

10. Inversion sans déclencheur

147. *Se trouvait-elle seule un instant avec lui, elle le voyait visiblement embarrassé.* (Le Rouge et le Noir, Stendhal)

Le tableau suivant montre les occurrences en valeur absolue de chaque catégorie d'inversion du sujet dans les romans des quatre auteurs :

Catégories	Chateaubriand	Flaubert	Hugo	Stendhal
Groupe prépositionnel+Verbe+Sujet	61	62	65	12
Groupe adverbial+Verbe+Sujet	18	22	12	13
Groupe mis en relief+Verbe+Sujet	2	1	17	5
Concessif avec inversion du sujet		1	16	12
Verbe au début de la phrase+Sujet	5	5		
Auxiliaire+Verbe+Sujet	1			
Locution conjunctive+Verbe+Sujet		3		
« Que » exclamatif+Verbe+Sujet		1		
Inversion complexe				3
Inversion sans déclencheur				3

Tableau 5.2 Fréquences absolues de différentes catégories d'inversions du sujet pour les quatre auteurs.

Pour éliminer l'influence des nombres de phrases différents pour chaque auteur, nous transformons les valeurs absolues en valeurs relatives, en divisant les occurrences de chaque catégorie par le nombre total d'occurrences de chaque auteur. Nous obtenons le tableau suivant avec les valeurs relatives :

Catégories	Chateaubriand	Flaubert	Hugo	Stendhal
Groupe prépositionnel+Verbe+Sujet	70.1149%	65.2632%	59.0909%	25.0000%
Groupe adverbial+Verbe+Sujet	20.6897%	23.1579%	10.9091%	27.0833%
Groupe mis en relief+Verbe+Sujet	2.2989%	1.0526%	15.4545%	10.4167%
Concessif avec inversion du sujet	0.0000%	1.0526%	14.5455%	25.0000%
Verbe au début de la phrase+Sujet	5.7471%	5.2632%	0.0000%	0.0000%
Auxiliaire+Verbe+Sujet	1.1494%	0.0000%	0.0000%	0.0000%
Locution conjunctive+Verbe+Sujet	0.0000%	3.1579%	0.0000%	0.0000%
« Que » exclamatif+Verbe+Sujet	0.0000%	1.0526%	0.0000%	0.0000%
Inversion complexe	0.0000%	0.0000%	0.0000%	6.2500%
Inversion sans déclencheur	0.0000%	0.0000%	0.0000%	6.2500%

Tableau 5.3 Fréquences relatives de différentes catégories d'inversions du sujet pour les quatre auteurs.

En combinant les résultats des occurrences en valeur absolue et en valeur relative, nous relevons plusieurs caractéristiques concernant l'utilisation des inversions stylistiques du sujet dans les romans de Chateaubriand, Flaubert, Hugo et Stendhal :

1. Le patron le plus fréquent pour les inversions stylistiques du sujet est « Groupe Prépositionnel+Verbe+Sujet », sauf pour Stendhal qui préfère inverser le sujet par un déclencheur adverbial. Le groupe adverbial le plus fréquent, comme déclencheur d'inversion, est « là » :

148. Là, se réunissaient toutes les dévotes qui amenaient, le plus souvent qu'elles le pouvaient, leurs filles avec elles. (Lamiel, Stendhal)

Le tableau suivant montre la modélisation en détail des patrons pour cette catégorie d'inversions dans les romans de Stendhal :

patrons	fréquences
ici+verbe+sujet	3
là+verbe+sujet	7
après+verbe+sujet	1
à deux pas+verbe+sujet	1
ensuite+verbe+sujet	1

Tableau 5.4 Patrons d'inversion adverbiaux dans les romans de Stendhal.

2. Chateaubriand n'a jamais utilisé l'inversion du sujet pour exprimer le concessif, cette utilisation est également rare dans les romans de Flaubert avec une seule occurrence :

*149. Il remonta dans le nord brusquement, décidé à s'ouvrir une des villes tyriennes, **dût-il** en faire le siège. (Salammbô, Flaubert)*

À l'inverse, les inversions concessives sont relativement courantes dans les romans de Hugo et de Stendhal.

3. Plus de 15% des inversions hugoliennes sont combinées avec le clivage, un procédé stylistique pour mettre encore en relief le déclencheur de l'inversion. Le déclencheur en clivage est majoritairement un groupe prépositionnel. Ce phénomène est beaucoup moins fréquent dans les romans de Flaubert, avec une seule occurrence.
4. La catégorie « Verbe au début de la phrase+Sujet » n'apparaît que dans les romans de Chateaubriand et de Flaubert, le patron le plus fréquent est « Venir+Sujet » :

*150. **Vient** ensuite l'objection contre les fonctions des anges. (Les Martyrs, Chateaubriand)*

151. *Viennent à moi ceux qui sont couverts de vin, ceux qui sont couverts de boue, ceux qui sont couverts de sang ;* (La Tentation de Saint-Antoine, Flaubert)

5. La seule occurrence de la catégorie « Que exclamatif+Verbe+Sujet » se trouve dans *La Tentation de Saint-Antoine* de Flaubert.

152. *Que ne suis-je un de ceux dont l'âme est toujours intrépide et l'esprit ferme, —comme le grand Athanase, par exemple.* (La Tentation de Saint-Antoine, Flaubert)

Cette phrase au présent est un discours rapporté direct du personnage Antoine. Flaubert est également le seul auteur à inverser le sujet après la locution conjonctive « à mesure que », après laquelle l'inversion du sujet n'est pas obligatoire :

154. *Mais, à mesure que se serrait davantage l'intimité de leur vie, un détachement intérieur se faisait qui la déliait de lui.* (Madame Bovary, Flaubert)

6. Deux catégories d'inversions sont spécifiques à Stendhal, et n'apparaissent donc jamais chez les autres auteurs : les inversions complexes et les inversions sans déclencheur. Pour les premières, les inversions complexes « consistent à reprendre le sujet nominal par un pronom postverbal²⁹² » (Tellier et al., 2018). Cette construction est courante dans les phrases interrogatives, mais les trois occurrences dans les romans de Stendhal se trouvent dans les phrases assertives :

²⁹² Tellier, Christine, et Daniel Valois. 2018. *Op. cit.*.

155. *Trente imprudents se réunissent-ils pour lire un numéro du Constitutionnel, Rassi les déclare conspirateurs, et les envoie prisonniers dans cette fameuse citadelle de Parme, terreur de toute la Lombardie.* (La Chartreuse de Parme, Stendhal)

156. *Il y eut aussi une nuance de hauteur dans sa manière de répondre à cet intérêt subit ; au moins la vieille duchesse d'Ancre en fit-elle la remarque.* (Armance, Stendhal)

157. *Sans le vouloir, Armance lui laissait-elle voir, par un mot, qu'elle ne le regardait plus comme un ami intime, son cœur se serrait, il en perdait la parole pour un quart d'heure.* (Armance, Stendhal)

Dans les trois phrases ci-dessus, la reprise du sujet nominal par un pronom pour créer une inversion du sujet n'est pas obligatoire, ce phénomène n'a été constaté que chez Stendhal.

7. La deuxième catégorie d'inversions spécifique pour Stendhal comprend les inversions sans déclencheur. Il s'agit de la juxtaposition conditionnelle²⁹³ (Tellier et al., 2018).

158. *Se trouvait-elle seule un instant avec lui, elle le voyait visiblement embarrassé.* (Le Rouge et le Noir, Stendhal)

159. *Avait-elle l'air un peu souffrant, même les jours où la sagesse faisait entendre sa voix terrible, il ne se trouvait plus le courage de lui adresser un de ces mots cruels si indispensables, selon son expérience, à la durée de leur amour.* (Le Rouge et le Noir, Stendhal)

²⁹³ *Idem.*

160. *Allaient-ils rêver aux charmes de leurs belles au fond des forêts éclairées par le pâle rayon de la lune, elle pensait aux dangers qu'ils couraient d'être surpris par des voleurs armés de poignards, dont elle lisait les exploits détaillés, tous les jours, dans la Quotidienne.* (Lamiel, Stendhal)

Dans les trois phrases, la première proposition avec le sujet inversé remplit la même fonction qu'une proposition avec la construction « Si... » pour exprimer la condition. Ce phénomène d'inversion du sujet pour exprimer la condition n'a été remarqué que dans les écrits de Stendhal dans notre corpus.

Les caractéristiques statistiques que nous avons relevées dans cette partie sur l'ensemble des romans d'un certain auteur nous donnent une idée générale sur les régularités d'apparitions des inversions du sujet par auteur. Ces caractéristiques sont souvent difficiles à remarquer par un travail manuel. Même si les occurrences sont relativement peu importantes, elles nous permettent de découvrir les préférences de chaque auteur pour l'emploi de cette figure de style dans leurs écrits. Nous insistons toujours sur le fait que l'analyse quantitative complète les études qualitatives en leur fournissant des caractéristiques générales. Nous explorons dans la partie suivante chaque occurrence d'inversion stylistique afin de relever les caractéristiques qualitatives qui peuvent y exister.

5.4.1.2 Analyses qualitatives des inversions du sujet

- **Chateaubriand**

-L'emploi de deux déclencheurs d'inversion

Contrairement aux inversions du sujet de Hugo qui engage souvent plusieurs déclencheurs avec la construction inversée, les inversions de Chateaubriand sont plus « discrètes ». L'auteur limite souvent les déclencheurs à deux :

161. Dans la vaste enceinte du palais de Dioclétien, au milieu du jardin des thermes, s'élevait un cyprès qu'arrosait une fontaine. (Les Martyrs, Chateaubriand)

162. À distance égale, le long de cet axe, sont assis trois esprits sévères : le premier est l'ange du passé ; (Les Natchez, Chateaubriand)

163. En face du palais d'Hiéroclès, dans une vaste prairie que bordaient les eaux du Ladon, s'élevait le tribunal du gouverneur romain. (Les Martyrs, Chateaubriand)

-Le rallongement de l'apodose

La structure syntaxique privilégiée de Chateaubriand d'après Jean Mourot est omniprésente dans les phrases avec l'inversion du sujet.

Dans l'exemple suivant, deux propositions de subordinations relatives sont employées pour rallonger le sujet de la phrase : « les roches ». La protase assez courte joue le rôle de déclencheur circonstanciel, tandis que l'apodose très longue permet de modifier le sujet de la phrase (les marqueurs de rythme entre la protase et l'apodose dans les exemples suivants sont ajoutés par nous) :

164. *À l'extrémité de cette arène / s'élevait une de ces roches isolées **que les gaulois appellent Dolmin, et qui marquent le tombeau de quelque guerrier.*** (Les Martyrs, Chateaubriand)

Dans l'exemple suivant, l'apodose est rallongée par une proposition de subordination (« qui brisera... »), qui, à son tour, est encore rallongée par une proposition en coordination (« et mettra... ») :

165. *Déjà est né / le nouveau Cyrus **qui brisera les derniers simulacres des esprits de ténèbres, et mettra le trône des Césars à l'ombre des saints tabernacles.*** (Les Martyrs, Chateaubriand)

Ou encore dans la phrase suivante, l'apodose est rallongée par deux propositions relatives, la deuxième est tellement longue que l'auteur doit mettre « : » pour couper le rythme de la proposition :

166. *À quelque distance du grand village / s'étendait une vallée tout environnée de bois **qui croissaient en amphithéâtre sur les collines, et qui formaient les entours de cette belle salle bâtie des mains de la nature : là devaient se célébrer les jeux ; le jeu de la balle et ensuite celui des osselets.*** (Les Natchez, Chateaubriand)

Le rallongement pourrait aussi prendre la forme de la comparaison, ce phénomène est également caractéristique de Chateaubriand d'après l'étude de Jean Mourot :

167. *À l'extrémité opposée / s'élevait le groupe de Laocoon et de ses fils, **comme si le sage, au milieu de ses voluptés, n'avait pu se passer de l'image de l'humanité souffrante !*** (Les Martyrs, Chateaubriand)

-L'emploi de plusieurs figures de style

Nous retrouvons les figures de style privilégiées de Chateaubriand dans les phrases avec l'inversion du sujet. Dans les trois exemples suivants, l'image, la figure de style d'analogie est combinée avec l'inversion du sujet :

168. *À l'extrémité d'une côte dangereuse, sur une grève où croissent à peine quelques herbes dans un sable stérile, s'élève une longue suite de pierres druidiques, semblables à ce tombeau où j'avais jadis rencontré Velléda.* (Les Martyrs, Chateaubriand)

169. *À l'orient du grand village des Natchez, dans la même cyprière où s'élevait le temple d'Athaensic, s'ouvre perpendiculairement, comme le soupirail d'une mine, une caverne profonde.* (Les Natchez, Chateaubriand)

170. *Au-dessus des flots de l'armée se hérissaient les baïonnettes, telles que ces lances du roseau qui tremblent dans le courant d'un fleuve.* (Les Natchez, Chateaubriand)

Les deux exemples suivants montrent la combinaison de l'inversion du sujet avec le clivage, l'effet d'insistance est renforcé par l'existence de deux figures de style de mise en relief :

171. *À l'extrémité du palais, du côté de la montagne de Sion, s'ouvrait une porte cachée qui conduisait au Calvaire : c'était par là qu'Hélène se dérobait aux hommages des peuples lorsqu'elle allait prier au pied de la croix.* (Les Martyrs, Chateaubriand)

172. *C'était au coucher du soleil que devait commencer la délibération* ; (Les Natchez, Chateaubriand)

Le dernier exemple montre la combinaison de l'inversion du sujet avec l'accumulation, une figure de répétition. Nous avons trouvé une seule occurrence de ce phénomène dans les résultats. Par rapport aux répétitions abondantes de Hugo, les répétitions de Chateaubriand sont peu fréquentes.

173. *Là se trouvaient la colonne milliaire, la poutre percée des clous sacrés, la louve de bronze, et les armes de Romulus.* (Les Martyrs, Chateaubriand)

Nos analyses des inversions du sujet de Chateaubriand révèlent des caractéristiques stylistiques de l'auteur. Les phénomènes stylistiques que nous avons pu constater confirment ce que disent les études littéraires sur le style de Chateaubriand : les traits caractéristiques sont les répétitions discrètes, l'image comme figure de style privilégié et la construction syntaxique typique de phrases avec une courte protase et une longue apodose.

● **Flaubert**

-La période ternaire

Le rythme ternaire, une structure syntaxique privilégiée de Flaubert, est très présent dans les résultats d'inversion du sujet. Il peut se construire autour de deux déclencheurs en coordination, suivis d'une proposition principale :

174. *Derrière une montagne surgissaient d'autres montagnes, / et au milieu du lac immense, / se dressait une île toute noire et de forme pyramidale.* (Salammbô, Flaubert)

Le rythme ternaire peut aussi être créé par l'expansion sémantique du sujet inversé :

175. Venait ensuite une façon de sac qui se terminait par un polygone cartonné, / couvert d'une broderie en soutache compliquée, / et d'où pendait, au bout d'un long cordon trop mince, un petit croisillon de fils d'or, en manière de gland. (Madame Bovary, Flaubert)

Quelquefois, toute la phrase avec l'inversion du sujet peut se trouver dans une période ternaire en juxtaposition :

176. Le pavillon de Melkarth, en pourpre fine, abritait une flamme de pétrole ; / sur celui de Khamon, couleur d'hyacinthe, se dressait un phallus d'ivoire, bordé d'un cercle de pierreries ; / entre les rideaux d'Eschmoûn, bleus comme l'éther, un python endormi faisait un cercle avec sa queue ; (Salammbô, Flaubert)

-La coupe flaubertienne

La coupe flaubertienne est une des caractéristiques les plus distinctives du style de l'auteur. Dans la phrase suivante, plusieurs prépositions, adverbes, et pronoms relatifs sont employés pour couper une phrase très longue dont la plupart des propositions portent sur la description du « grand garçon ». Nous retrouvons également le « cauchemar » de Flaubert concernant les pronoms relatifs « qui » et « que » : au lieu d'employer un deuxième « qui », Flaubert l'a remplacé par « lequel » pour éviter la répétition. Dans cette phrase à l'imparfait, nous retrouvons les participes présents, qui, d'après Thibaudet sont emblématiques du style de Flaubert pour « exprimer une continuité sur laquelle trancheront les moments individuels et saillants des temps verbaux²⁹⁴ » (Thibaudet, 1935) :

²⁹⁴ Thibaudet, Albert. 1935. *Op. cit.*.

177. *Derrière son dos marchait un grand garçon, dans le costume classique du Dante, et qui était (elle ne s'en cachait plus, maintenant) l'ancien chanteur de l'Alhambra, lequel, s'appelant Auguste Delamare, s'était fait appeler primitivement Anténor Dellamarre, puis Delmas, puis Belmar, et enfin Delmar, modifiant ainsi et perfectionnant son nom, d'après sa gloire croissante ;* (L'Éducation sentimentale, Flaubert)

L'exemple suivant montre une phrase avec un rythme équilibré, parce que l'auteur la coupe en plusieurs syntagmes courts. L'accent de la phrase est mis sur la dernière proposition, assez brève, « qui souriait toujours » :

178. *Près de lui, / sur une natte et jambes croisées, / se tenait un enfant très beau, / qui souriait toujours.* (Trois Contes, Flaubert)

Le dernier exemple présente la coupe flaubertienne réalisée par une virgule et un participe. Les deux participes sont employés pour remplacer deux propositions de subordination qui auraient pu être mises en place par « qui », le pronom relatif que déteste Flaubert :

179. *Au milieu se hérissait la phalange, formée par des syntagmes ou carrés pleins, ayant seize hommes de chaque côté.* (Salammbô, Flaubert)

-L'image

L'image est une des figures de style typiques de Flaubert pour créer des tableaux dans ses romans, les trois exemples suivants contiennent chacun une occurrence de cette figure de style mêlé à une inversion du sujet. L'image est souvent employée pour désigner les qualités du sujet inversé :

180. *Derrière le château, s'étendait une forêt ayant le dessin d'un éventail.* (Trois Contes, Flaubert)

181. *Parfois, se dressait un énorme rocher, pareil à la proue d'un vaisseau ou au piédestal de quelque colosse disparu.* (Salammbô, Flaubert)

182. *Ils avaient été sur les rives de l'Orne choisir des granits, les avaient cassés, numérotés, rapportés eux-mêmes dans une charrette, puis avaient joint les morceaux avec du ciment, en les accumulant les uns par-dessus les autres ; et au milieu du gazon se dressait un rocher, pareil à une gigantesque pomme de terre.* (Bouvard et Pécuchet, Flaubert)

L'analyse des inversions du sujet de Flaubert nous permet de relever les caractéristiques propres à cet auteur. Nous retrouvons, à travers l'inversion du sujet, le rythme ternaire, la coupe flaubertienne et l'image comme figure de style. Ces caractéristiques nous permettent de différencier les phrases de Flaubert de celles des autres auteurs dans notre corpus.

● Stendhal

-L'emploi des adjectifs qualitatifs

Les caractéristiques stylistiques liées à l'inversion du sujet de Stendhal sont relativement moins nombreuses que chez les autres auteurs, ce qui confirme l'idée d'un style « clair et naturel » et le rejet des figures de style par Stendhal, qui les trouve « hypocrites ». Un des phénomènes que nous avons pu relever dans nos résultats pour l'inversion du sujet est l'emploi fréquent d'adjectifs qualitatifs : il s'agit souvent de

répétitions liées au parallélisme. Dans les deux exemples suivants, le sujet inversé est modifié par le patron adjectival « ADV ADJ », en parallélisme.

183. *Autour de la table, que les gens venaient d'apporter toute servie, se trouvaient sept à huit femmes **fort nobles, fort dévotes, fort affectées**, âgées de trente à trente-cinq ans.* (Le Rouge et le Noir, Stendhal)

184. *Là se trouvait un pont en fer **fort léger et fort étroit**, au milieu duquel était établi un portier.* (La Chartreuse de Parme, Stendhal)

-L'archaïsme

Nous avons évoqué dans la section 5.2.3 que l'innovation dans la langue est peu appréciée par Stendhal : il se réfère majoritairement à la tradition littéraire d'avant le 19^{ème} siècle. Nous retrouvons effectivement des phénomènes qui traduisent un aspect archaïque de la langue. Les trois phrases suivantes, où l'inversion du sujet exprime une condition, sont typiques dans nos résultats. La juxtaposition conditionnelle est un phénomène qui représente un usage archaïque dans la langue française. Elle n'est jamais apparue chez les autres auteurs de notre corpus.

185. ***Se trouvait-elle** seule un instant avec lui, elle le voyait visiblement embarrassé.*
(Le Rouge et le Noir, Stendhal)

186. ***Avait-elle** l'air un peu souffrant, même les jours où la sagesse faisait entendre sa voix terrible, il ne se trouvait plus le courage de lui adresser un de ces mots cruels si indispensables, selon son expérience, à la durée de leur amour.* (Le Rouge et le Noir, Stendhal)

187. ***Allaient-ils** rêver aux charmes de leurs belles au fond des forêts éclairées par le pâle rayon de la lune, elle pensait aux dangers qu'ils couraient d'être*

surpris par des voleurs armés de poignards, dont elle lisait les exploits détaillés, tous les jours, dans la Quotidienne. (Lamiel, Stendhal)

Le deuxième phénomène que l'on peut qualifier d'archaïque est l'inversion du sujet concessive, surtout avec le verbe conjugué au subjonctif imparfait. L'emploi du subjonctif imparfait dans une proposition concessive est « très rare en français contemporain et trahit toujours un niveau de langue archaïsant ou recherché²⁹⁵ » (Morel, 1980, p. 111). Nous avons retrouvé ce même phénomène dans les phrases de Hugo, une seule occurrence dans les phrases de Flaubert et 0 occurrence chez Chateaubriand.

188. *Dût-il s'y résoudre, eh bien !* (Le Rouge et le Noir, Stendhal)

189. *Dussé-je m'exposer à mille morts, j'emploierai les moyens même les plus dangereux pour introduire ce paquet de cordes dans la citadelle, au mépris, hélas !*
(La Chartreuse de Parme, Stendhal)

190. *Y eût-il un nouveau Pitt en Angleterre, on ne mystifie pas deux fois une nation par les mêmes moyens...* (Le Rouge et le Noir, Stendhal)

Le dernier exemple montre une phrase dont l'inversion du sujet ne possède pas de déclencheur évident. La construction de la phrase nous semble très particulière avec un ordre des mots surprenant.

191. *C'était sa façon d'exprimer les angoisses que **donnait** à sa vanité **un conseil aussi direct.*** (Féder, Stendhal)

²⁹⁵ Morel parle du côté archaïsant des inversions concessives par rapport aux écrivains contemporains, néanmoins, les écrivains dans notre corpus ne sont pas les contemporains. Cependant, dans notre corpus, nous constatons effectivement que ce phénomène, dit « archaïsant », n'apparaît pas souvent chez Chateaubriand et Flaubert, qui cherchent à innover le langage, alors qu'il apparaît relativement plus souvent chez Stendhal qui se réfère à la tradition littéraire et qui n'apprécie pas les innovations langagières (voir la section 5.2 pour la discussion sur la notion d'innovation pour les trois auteurs). Morel, Mary-Annick. 1980. « L' inversion du sujet dans certaines propositions concessives du français : Conditions d' emploi et valeur ». *DRLAV. Documentation et Recherche en Linguistique Allemande Vincennes* 22 (1): 105-12.

On peut remarquer chez Stendhal un usage archaïque de l'inversion du sujet, ce qui confirme son style moins « innovant » que les autres auteurs de son époque. Les figures de style courantes chez Flaubert, Chateaubriand et Hugo, comme l'image, la répétition, l'antithèse, sont rares chez Stendhal. Notre analyse des inversions du sujet dans les romans de Stendhal nous permet donc de confirmer le style archaïque de l'auteur avec une faible présence de figures de style.

Les analyses statistique et littéraire sur les inversions du sujet nous permettent d'adopter un point de vue à la fois général et détaillé concernant l'utilisation de cette figure de style chez nos trois auteurs. Les différences statistiques sont identificatoires alors que l'étude de chaque occurrence met en évidence les caractéristiques stylistiques de ces phrases. Les phénomènes formels, mais stylistiques, de chaque auteur, tels que la coupe flaubertienne, l'archaïsme pour Stendhal et l'apodose rallongée pour Chateaubriand servent à confirmer ce qui a été remarqué par les études littéraires antérieures, tandis que les résultats statistiques ouvrent de nouvelles possibilités pour étudier les spécificités de chaque auteur.

5.4.2 Le clivage

Nous continuons notre examen de la méthode d'extraction supervisée dans cette partie, pour relever les clivages dans les romans de Chateaubriand, Flaubert, Hugo et Stendhal. Comme précédemment pour l'inversion du sujet, l'analyse des clivages tient compte à la fois des caractéristiques statistiques et littéraires de nos résultats. La partie statistique présente les fréquences des différentes catégories de clivages, alors que la partie analytique littéraire vise à étudier toutes les occurrences de cette figure de style dans le texte original, afin de trouver les caractéristiques proéminentes pour chaque auteur. Rappelons que le clivage est une des figures de style syntaxiques d'insistance dont la fonction stylistique est de chercher à susciter l'attention du lecteur.

5.4.2.1 Analyses quantitatives des clivages

Après vérification manuelle des résultats d'extraction, nous avons relevé 190 phrases avec clivage dans les romans de Chateaubriand, 148 pour Flaubert, 868 pour Hugo et 252 pour Stendhal. Le nombre total pour chaque auteur comprend les deux constructions clivées dans la langue française : « C'est X que Y²⁹⁶ » et « C'est X qui Y », le dernier mettant en relief le sujet de la proposition et le premier tous les autres cas. Le tableau suivant présente les fréquences absolues de clivages pour chaque auteur, la dernière colonne du tableau montre le pourcentage de cette figure de style par rapport au nombre de phrases total pour chaque auteur.

²⁹⁶ La construction « C'est X que Y » que nous présentons ici comprend également la forme négative (« ce n'est pas X que Y »), la forme inversée (« est-ce X que Y ») et la forme négative inversée (« N'est-ce pas X que Y »). Idem pour la construction « C'est X qui Y ».

Auteurs	C'est X que Y	C'est X qui Y	Total	Pourcentage
Chateaubriand	78	112	190	1.0206%
Flaubert	42	106	148	0.3959%
Hugo	265	603	868	0.8476%
Stendhal	111	141	252	0.5643%

Tableau 5.5 Fréquences des clivages pour les quatre auteurs.

Les fréquences relatives dans le tableau ci-dessus montrent déjà des caractéristiques superficielles concernant l'utilisation du clivage dans les romans des quatre auteurs. Chateaubriand est l'auteur qui utilise le plus cette figure de style alors que, relativement, Flaubert s'en sert moins fréquemment. Pour les quatre auteurs, la mise en relief du sujet par clivage est toujours la tournure la plus fréquente. Nous essayons ensuite de dresser une typologie des structures clivées pour en relever les caractéristiques essentielles. Nous allons nous concentrer uniquement sur la construction « C'est X que Y », et ce pour deux raisons : premièrement, dans l'analyse quantitative pour les clivages dans les romans de Victor Hugo, présentée dans 4.4.2 du Chapitre 4, nous avons constaté que la construction « C'est X qui Y » ne présente pas beaucoup de variétés concernant la partie clivée. Le X dans cette construction est infailliblement le sujet de la proposition, qui se manifeste grammaticalement par la séquence « groupe nominal », « groupe pronominal » ou « nom propre ». Les caractéristiques statistiques ne sont pas très intéressantes d'un point de vue stylistique ; deuxièmement, les analyses qualitatives sur les clivages en forme de « C'est X qui Y » ne nous permettent pas de relever de caractéristiques stylistiques. Cela ne veut pas dire que les phrases elles-mêmes ne sont pas représentatives du style de l'auteur, nous constatons seulement que les phénomènes stylistiques se trouvent rarement dans cette construction clivée. Ainsi, nous jugeons que la construction « C'est X qui Y » contient relativement peu d'informations stylistiques tandis que la construction « C'est X que Y » est plus intéressante du fait de la grande diversité grammaticale de la partie clivée.

Notre modélisation se fonde sur la typologie des syntagmes clivés (voir 4.4.2.2 du

Chapitre 4). Ainsi, pour la construction « c'est X que Y », sept grandes catégories ont été repérées dans notre corpus, elles couvrent presque toutes les catégories grammaticales :

1. C'est+Groupe Nominal+que

192. Addison et Louis Racine ont fort bien démontré, au sujet du paradis perdu, que c'est l'action et non pas le héros qui fait l'épopée. (Les Martyrs, Chateaubriand)

2. C'est+Groupe Prépositionnel+que

193. Et c'était pour cet hypocrite, pour ce rebelle, qu'il avait, lui, fils du vice-roi de Norvège, élève du général Levin, compromis son avenir, exposé sa vie ! (Han d'Islande, Hugo)

3. C'est +Groupe Pronominal+que

194. Un noir (c'était celui-là même que je cherchais !) lui cria en français : Punis-moi, car je viens de t'offenser ; mais ne fais rien à mon frère, qui n'a touché qu'à ton rosier ! (Bug-Jargal, Hugo)

4. C'est +Groupe Adverbial+que

195. Mais c'était en vain que ces dames choisissaient les steppes les plus nues de la plage du Friesches-Haff (bras de mer voisin de Kœnigsberg), toujours, elles étaient contrepassées par de brillantes cavalcades de jeunes gens qui même avaient mis à la mode la couleur noire qui était celle du deuil de Mina.

(Le Rose et le Vert, Stendhal)

5. C'est +Nom Propre+que

196. *Et sans se soucier de leurs clameurs, il disait qu'en abandonnant le Suffète, c'était la République qu'on avait abandonnée.* (Salammbô, Flaubert)

6. C'est+Locution Conjonctive+que

197. *C'est parce qu'ils ont le nid qu'ils ont le chant.* (Les Misérables, Hugo)

7. C'est+Groupe Adjectival+que

198. *Ce fut, riche de cette expérience, que le surlendemain à midi, il se présenta à l'abbé Pirard, qui le regarda beaucoup.* (Le Rouge et le Noir, Stendhal)

Le tableau suivant présente les fréquences absolues de chaque catégorie de clivages pour nos quatre auteurs :

Auteurs	Nominal	Prépositionnel	Pronominal	Adverbial	Nom Propre	Locution Conjonctive	Adjectival	Total
Chateaubriand	13	40		23	2			78
Flaubert	7	22	3	8	2			42
Victor Hugo	47	177	6	26	3	6		265
Stendhal	17	73		16	2	2	1	111

Tableau 5.6 Fréquences absolues des différentes catégories de clivages pour les quatre auteurs.

Pour mieux visualiser la préférence de chaque auteur à utiliser chacune des

catégories de clivages, la figure ci-dessous montre les pourcentages de catégories pour chaque auteur :

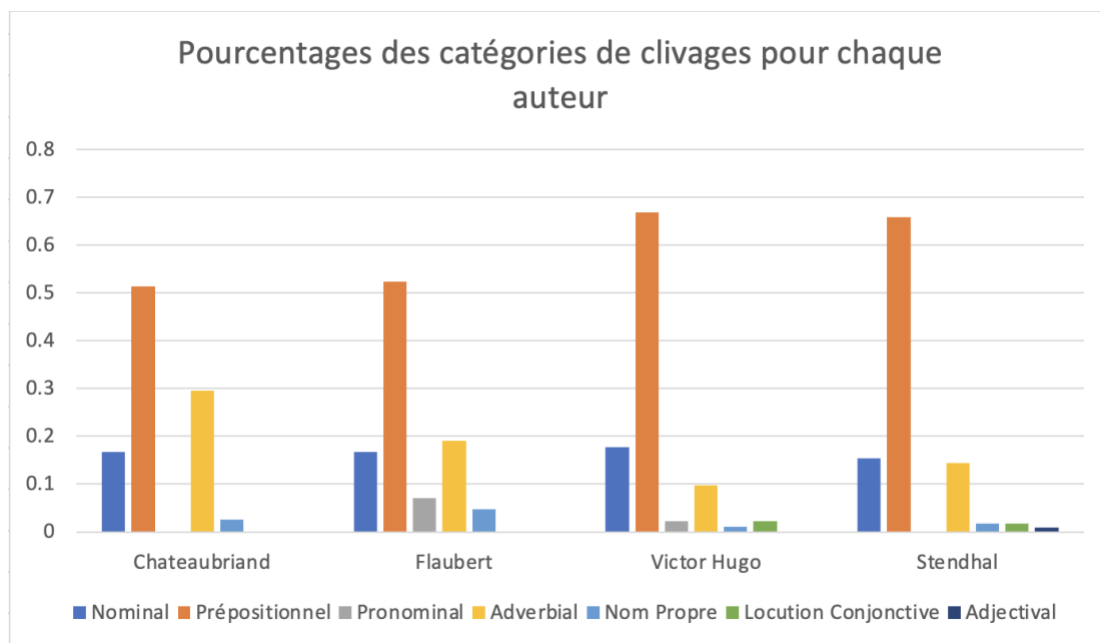


Figure 5.2 Pourcentages des catégories de clivages pour chaque auteur.

Les résultats statistiques de la typologie des clivages nous permettent de relever quelques caractéristiques concernant la préférence de chaque auteur pour mettre en relief ainsi une partie de la phrase :

1. La construction syntaxique « C'est+Groupe Prépositionnel+que » est toujours la plus fréquente chez les quatre auteurs. Cette construction met en relief un complément circonstanciel au sein de la phrase. Nous constatons que Victor Hugo est l'auteur qui met le plus souvent l'accent sur le complément circonstanciel dans ces phrases, alors que Chateaubriand utilise relativement moins cette technique.
2. Chez Chateaubriand, le pourcentage du patron « C'est+Groupe Adverbial+que » tient un rôle important, le tableau ci-dessous présente les adverbes les plus fréquents dans les parties clivées :

Adverbes	Fréquences
ainsi	7
alors	3
bien	1
demain	1
donc bien vainement	1
en vain	3
ici	5
là	2

Tableau 5.7 Patrons adverbiaux les plus fréquemment mis en relief par le clivage dans les romans de Chateaubriand.

Victor Hugo, en revanche, présente moins d'occurrences relatives de ce patron.

3. Victor Hugo et Stendhal sont les deux auteurs qui utilisent le patron « C'est+Locution Conjonctive+que ». La locution conjonctive qui apparaît le plus souvent est « parce que » :

199. C'est parce que je savais cette condamnation à la prison seulement, que j'étais effrayé des bruits d'exécution prochaine qui se répandent par la ville ; (La Charteruse de Parme, Stendhal)

Les deux auteurs sont également les seuls à mettre en relief un groupe pronominal par le clivage, un phénomène très courant dans la langue parlée :

200. Et si c'était lui en effet que la veille mess Lethierry avait déclaré son gendre, comment s'expliquer ce qu'il faisait là ? (Les Travailleurs de la Mer, Hugo)

Cette phrase de Hugo est un discours indirect libre du personnage Ebenezer, le

discours est à l'imparfait. La mise en relief du pronom « lui » dans cette phrase la rend proche de la langue parlée, d'un discours direct, mais aucun marqueur du discours direct n'est présent dans cette phrase. Nous remarquons les caractéristiques d'un discours indirect libre, conservant à la fois les marques du récit et les marques de la langue parlée.

4. Une seule phrase dans notre corpus met en relief un syntagme adjectival par le clivage :

201. Ce fut, riche de cette expérience, que le surlendemain à midi, il se présenta à l'abbé Pirard, qui le regarda beaucoup. (Le Rouge et le Noir, Stendhal)

La mise en relief d'un syntagme adjectival n'est pas impossible dans la langue française, nous avons rencontré, pendant notre vérification manuelle, des patrons syntaxiques comme : « c'est vrai que ... ». Cependant, nous avons décidé de ne pas les inclure dans nos résultats parce que cette construction est devenue une tournure courante dans la langue française, dont la fonction stylistique n'est souvent pas le résultat d'un choix subjectif de l'auteur. Mais cette phrase de Stendhal représente un phénomène qui n'a pas été constaté chez les autres auteurs : la mise en relief d'un syntagme adjectival, « riche de cette expérience », qui remplit la même fonction qu'une proposition adverbiale et participiale : « Ce fut, [en étant] riche de cette expérience, que le surlendemain à midi, il se présenta à l'abbé Pirard, qui le regarda beaucoup ». Néanmoins, le fait que ce phénomène n'apparaît qu'une seule fois dans les romans de Stendhal montre la difficulté de généraliser ce phénomène.

L'analyse statistique et la typologie de la construction « C'est X que Y » révèlent

les préférences propres à chaque auteur concernant l'emploi de cette figure de style marquant l'insistance. Comme tous les autres aspects statistiques, ces caractéristiques sont souvent négligées par les critiques littéraires, parce que leurs analyses portent rarement sur l'ensemble des écrits d'un auteur. Nous admettons en même temps que ces fréquences peu importantes sont difficiles à interpréter pour arriver à une conclusion générale sur la syntaxe d'un auteur.

Pourtant, les caractéristiques que nous avons pu constater autour du clivage contribuent à la caractérisation du style littéraire, en donnant accès aux préférences des auteurs quant à l'emploi de cette figure de style. Nous avons évoqué à plusieurs reprises dans la thèse que la caractérisation du style doit se fonder sur plusieurs paramètres, une seule figure de style ou un seul patron syntaxique n'est jamais suffisant pour donner une conclusion sur les qualités d'un style littéraire. Nous étudierons dans la partie suivante les occurrences des clivages dans les textes originaux, afin de relever les caractéristiques stylistiques qui peuvent exister dans cette construction syntaxique.

5.4.2.2 Analyses qualitatives des clivages.

- **Chateaubriand**

-La protase brève + l'apodose longue

Cette construction syntaxique privilégiée de Chateaubriand concerne la construction même du patron clivé. Dans l'exemple suivant, le Y dans « C'est de X que Y » est énormément rallongé par deux propositions de subordination avec le pronom relatif « qui ».

202. *C'est de ce moment, ô René, / que j'ai conçu une merveilleuse idée **de cette religion qui**, dans les forêts, au milieu de toutes les privations de la vie, peut remplir de mille dons les infortunés ; **de cette religion qui**, opposant sa puissance au torrent des passions, suffit seule pour les vaincre, lorsque tout les favorise, et le secret des bois, et l'absence des hommes, et la fidélité des ombres. (Atala, Chateaubriand)*

Ou encore, dans la phrase suivante, une proposition de subordination est mise en place pour désigner « [les] victimes de Verrès ».

203. *C'est ainsi qu'au rivage opposé de la Sicile, / je croyais voir ces victimes de Verrès, **qui** du haut de l'instrument de leur supplice, tournoient inutilement vers Rome leurs regards mourants. (Les Martyrs, Chateaubriand)*

Le dernier exemple présente une phrase remarquable avec sept propositions de subordination, ayant pour but de rallonger d'une manière exagérée l'apodose.

204. *Ce fut en vain que je cherchai cette sage et sérieuse Égypte, **qui** donna Cécrops et Inachus à la Grèce, **qui** fut visitée par Homère, Lycurgue et Pythagore, et par Jacob, Joseph et Moïse, **cette Égypte où** le peuple jugeait ses rois après leur mort, **où** l'on empruntait en livrant pour gage le corps d'un père, **où** le père qui avait tué son fils était obligé de tenir pendant trois jours le corps de ce fils embrassé, **où** l'on promenait un cercueil autour de la table du festin, **où** les maisons s'appelaient des hôtelleries, et les tombeaux des maisons. (Les Martyrs, Chateaubriand)*

-La répétition

La répétition chez Chateaubriand, comme nous l'avons évoqué dans la partie 5.2, est « discrète ». Elle se limite souvent à deux ou trois structures répétitives, avec les éléments homogènes. Le premier exemple montre une répétition dans le X de la construction clivée « C'est X que Y » ; nous remarquons également une antithèse dans cette répétition entre la « joie sublime » et la « tendre tristesse ».

205. C'est dans cette extase d'admiration et d'amour, dans ces transports d'une joie sublime, ou dans ces mouvements d'une tendre tristesse, que les élus répètent ce cri de trois fois saint, qui ravit éternellement les cieux. (Les Martyrs, Chateaubriand)

L'exemple suivant montre deux répétitions homogènes, la première donne une précision de lieu où « sont réunis » les sujets, la deuxième est une énumération des sujets de la phrase. Tous les éléments pour les deux répétitions sont homogènes, ils tombent dans la même catégorie sémantique respective.

206. C'est dans les parvis de la cité sainte, et dans les champs qui l'entourent, que sont à la fois réunis et partagés les chœurs des chérubins et des séraphins, des anges et des archanges, des trônes et des dominations : tous sont les ministres des ouvrages ou des volontés de l'éternel. (Les Martyrs, Chateaubriand)

-Syntagme de négation clivé + syntagme d'affirmation clivée

Il s'agit d'un phénomène spécifique, que nous avons repéré dans les phrases clivées de Chateaubriand : l'auteur utilise un syntagme de négation clivé suivi par un syntagme d'affirmation clivé ou non, pour mettre l'accent sur le syntagme d'affirmation. Étant

donné que le clivage est déjà une figure de style de mise en relief, le fait de présenter une affirmation après une négation met encore l'accent sur la partie affirmative :

207. *Il a fallu, pour m'attaquer, changer de poids et de mesures, et reprocher aux martyrs ce qu'on approuve partout ailleurs : **car ce n'est pas la manière, c'est le fond qu'on censure dans l'épisode de Velléda : et pourtant Velléda est-elle autre chose que Circé, Didon, Armide, Eucharis, Gabrielle ?*** (Les Martyrs, Chateaubriand)

208. *Si ce n'est pas sans quelques regrets, c'est du moins sans remords que j'ai jeté un regard sur les premiers jours de ma vie ;* (Les Martyrs, Chateaubriand)

209. *Ce ne fut plus ici par des sanglots que je troublai le récit d'Atala, ce fut par ces emportements qui ne sont connus que des Sauvages.* (Atala, Chateaubriand)

210. *Mais que craignez-vous aujourd'hui d'un vieillard aveugle ? — **Ce ne sont pas tes coups que nous craignons, s'écrièrent à la fois les colons, mais tes conseils.*** (Les Natchez, Chateaubriand)

211. *Car ce n'est pas un homme qu'il vient secourir, c'est le fils même de l'éternel !* (Les Martyrs, Chateaubriand)

Nos analyses sur le clivage de Chateaubriand montrent la capacité de ce patron à révéler les caractéristiques stylistiques de l'auteur. Au sein de la construction syntaxique du clivage, que ce soit sur le X ou le Y dans « C'est X que Y », nous retrouvons la structure syntaxique privilégiée de Chateaubriand, nous constatons également les répétitions discrètes et homogènes. La dernière caractéristique qui met en contraste deux syntagmes opposés dans la partie X n'a pas encore été

systématiquement mise en avant par les critiques littéraires, notre étude exhaustive permet de le repérer.

- **Flaubert**

-La période ternaire

Nous retrouvons, dans nos résultats autour des clivages les périodes ternaires, une des structures syntaxiques préférées de Flaubert. Nous constatons que la période ternaire de Flaubert est différente de celle de Chateaubriand, qui préfère souvent l'ordre croissant pour cette construction. La période ternaire de Flaubert se construit majoritairement avec la deuxième proposition plus courte que les autres. Rappelons-nous que le style de Flaubert possède souvent un aspect oratoire (le « gueuloir »), ainsi, le rythme dans les phrases de Flaubert est censé être équilibré. Dans les trois exemples suivants, la proposition courte au milieu de la phrase joue presque le rôle d'une division rythmique pour que le lecteur reprenne son souffle pour la proposition finale, plus longue. :

212. Ce fut donc avec joie qu'il accepta la proposition de l'hôtesse de dîner en la compagnie des nouveaux venus, / et l'on passa dans la grande salle, / où Mme Lefrançois, par pompe, avait fait dresser les quatre couverts. (Madame Bovary, Flaubert)

213. Ce fut de cette manière, à force de l'entendre, / qu'elle apprit le catéchisme, / son éducation religieuse ayant été négligée dans sa jeunesse ; (Trois contes, Flaubert)

214. *C'était par hasard qu'il se trouvait au festin, / son père le faisant vivre chez les Barca, / selon la coutume des rois qui envoyaient leurs enfants dans les grandes familles pour préparer des alliances.* (Salammbô, Flaubert)

-La coupe flaubertienne

Dans l'exemple suivant, la proposition clivée est mise à la fin de la phrase afin de créer un effet d'insistance :

215. *Si j'ai éprouvé des moments d'enthousiasme, c'est à l'art que je les dois ;*
(Mémoires d'un fou, Flaubert)

Deux propositions dans la phrase suivante, construites l'une autour d'un participe présent, l'autre autour d'un pronom relatif, sont employées pour modifier le Y (« elle s'était jointe à celui-là ») dans la construction « C'est X que Y ». Le participe présent remplace un pronom relatif, « qui », que Flaubert préfère ne pas répéter dans ses phrases. Les deux propositions sont soigneusement placées dans cette phrase courte pour créer un rythme équilibré (sans trop rallonger l'apodose) par la coupe syntaxique :

216. *C'était pour y atteindre que, **délaissant** son premier époux, elle s'était jointe à celui-là, **qui** l'avait dupée, pensait-elle.* (Trois contes, Flaubert)

Le dernier exemple représente parfaitement le style de Flaubert dans ses phrases, plusieurs propositions sont mises en place pour créer un « mouvement²⁹⁷ » (Thibaudet, 1935) continu du rythme de la phrase. Les prépositions, les pronoms relatifs divers enchaînent les différentes propositions. L'effet de fluidité est créé parce qu'aucune

²⁹⁷ Thibaudet, Albert. 1935. *Op. cit.*

proposition n'est trop longue pour rompre le mouvement rythmique :

217. M. Lheureux, en effet, prêtait sur gages, et c'est là qu'il avait mis la chaîne en or de Mme Bovary, avec les boucles d'oreilles du pauvre père Tellier, qui, enfin contraint de vendre, avait acheté à Quincampoix un maigre fonds d'épicerie, où il se mourait de son catarrhe, au milieu de ses chandelles moins jaunes que sa figure.

(Madame Bovary, Flaubert)

Les caractéristiques stylistiques dans les phrases clivées de Flaubert sont relativement limitées en nombre, cependant, les phrases clivées flaubertiennes se distinguent clairement de celles des autres auteurs par le rythme équilibré, la coupe soignée et l'harmonie générale qui donnent un effet oratoire de fluidité dans son style.

● Stendhal

-La coupe de Stendhal

La coupe phrastique de Stendhal est très différente de celle de Flaubert. Si ce dernier apporte un soin extrême au rythme oratoire de ses phrases, la coupe stendhalienne est relativement moins harmonieuse et moins rythmique. Par exemple, dans la phrase suivante, on pourrait imaginer un point au lieu d'une virgule devant le mot « jamais », parce que la proposition après « jamais » est sémantiquement indépendante de la proposition précédente. La coupe par la virgule donne l'impression que la partie après « jamais », rallongée par les propositions relatives, est une digression.

218. Ce fut en vain que Fabrice, dont le talent semblait plus brillant chaque jour,

prêcha encore plusieurs fois dans cette même petite église, voisine du palais Crescenzi, **jamais** il n'aperçut Clélia, qui même à la fin prit de l'humeur de cette affectation à venir troubler sa rue solitaire, après l'avoir déjà chassée de son jardin. (La Chartreuse de Parme, Stendhal)

Le deuxième exemple montre le même phénomène, la coupe par une virgule devant « et lui partit » ne relève pas d'une nécessité syntaxique. Le rallongement de la phrase ne vise pas à créer un rythme équilibré, l'auteur aurait pu mettre un point après « un cadeau » et initier une autre phrase avec « Il partit... ».

219. *Ce fut en 1821, qu'ayant subi passablement tous ses examens, son directeur d'études ou gouverneur eut une croix et un cadeau, et lui partit pour voir enfin cette ville de Parme, à laquelle il songeait souvent.* (La Chartreuse de Parme, Stendhal)

Dans l'exemple suivant, la coupe avec « : » marque aussi le changement de sujet. En effet, la phrase comprend plusieurs sujets sémantiques : le personnage parle d'abord de la « position » où il veut « recevoir [les] ordres », ensuite, il se moque de la présence de l'interlocuteur, et enfin, initié par « jadis », il désigne les qualités du « malheureux fou », et le pronom « je » est remplacé par « il ». L'agencement de plusieurs sujets sémantiques dans la même phrase crée un effet de digression et de désordre, le lecteur aurait suivi le développement des différentes propositions, mais il aurait sans doute du mal à revenir à la première proposition de la phrase où l'accent est mis sur « dans cette position » par le clivage. La coupe nous semble moins équilibrée et moins logique par rapport à celle des autres auteurs dans notre corpus.

220. *C'est dans cette position que je veux recevoir vos ordres, s'écria le chanoine,*

*car il est évident que vous avez quelque service à me demander; autrement vous n'auriez pas honoré de votre présence la pauvre maison d'un malheureux fou : **jadis transporté d'amour et de jalousie, il se conduisit avec vous comme un lâche, une fois qu'il vit qu'il ne pouvait vous plaire.*** (La Chartreuse de Parme, Stendhal)

Quelquefois, Stendhal ne coupe pas ses phrases, la phrase devient très longue et moins équilibrée, sans marqueur de pause. Les propositions dans la phrase suivante sont enchaînées par les pronoms relatifs et les prépositions, mais aucune ponctuation n'est employée entre les propositions pour couper la phrase.

*221. Il avait tellement envie de rire **que** ce ne fut qu'à grand-peine qu'il ne suivit pas l'idée folle **qui** lui vint de distribuer des pièces de cinq francs aux laquais de la maison **qu'il** trouva dans l'antichambre **rangés en haie sur** son passage. (Lucien Leuwen, Stendhal)*

La coupe stendhalienne peut aussi créer une rupture. Dans la phrase suivante, la construction de comparaison « beaucoup plus ... que » est coupée par le circonstanciel « avec dix laquais portant sa livrée ». Si nous avons évoqué le fait que la coupe flaubertienne crée souvent un effet de fluidité, cet effet est rarement présent dans la coupe stendhalienne.

*222. Ce fut une grande leçon de philosophie pour Fabrice que de se trouver parfaitement insensible à tous ces honneurs, et **beaucoup plus** malheureux dans cet appartement magnifique, avec dix laquais portant sa livrée, **qu'il** n'avait été dans sa chambre de bois de la tour Farnèse, environné de hideux geôliers, et craignant toujours pour sa vie. (La Chartreuse de Parme, Stendhal)*

-La dissonance de temps

La dissonance de temps est une des caractéristiques proéminentes des phrases clivées de Stendhal. Dans la phrase suivante, le subjonctif plus-que-parfait est employé, mais les autres verbes sont conjugués au présent et au futur simple.

*223. C'est bien pour le coup que la duchesse m'eût dit, si jamais il m'eût été donné de revoir ses beaux yeux, que mon âme **trouve** du plaisir à contempler ce qui **arrivera** dans dix ans, et **oublie** de regarder ce qui **se passe** actuellement à mes côtés. (La Chartreuse de Parme, Stendhal)*

Le dernier exemple présente une phrase où le présent est employé avec le passé simple. La première phrase est une simple description du « corridor » où Fabrice « remarqua trois portes de fer ». Le récit est majoritairement au passé simple, le présent dans la description est inattendu.

*224. C'est par un corridor obscur, placé au centre du bâtiment, que l'on **arrive** à ces chambres qui toutes **ont** deux fenêtres ; et dans ce corridor fort étroit, Fabrice **remarqua** trois portes de fer successive formées de barreaux énormes et s'élevant jusqu'à la voûte. (La Chartreuse de Parme, Stendhal)*

Les analyses des phrases clivées de Stendhal révèlent les caractéristiques stylistiques de ce patron. Ces caractéristiques rythmiques nous ont amenés à nous pencher sur la coupe syntaxique stendhalienne, moins équilibrée et moins harmonieuse que celle de Flaubert. Bien qu'il y ait effectivement moins de figures de style dans les phrases de Stendhal, celles-ci sont loin d'être claires et simples. Les spécificités qui existent dans la coupe stendhalienne créent souvent des problèmes de compréhension

liés à la rupture, au rallongement de plusieurs propositions sémantiquement indépendantes et au manque de marqueur de pause dans les phrases complexes. Nous constatons également que la dissonance de temps est présente dans les phrases clivées, mais leurs fonctions semblent souvent non justifiées dans le contexte. Nous concluons que, d'une part, nos analyses révèlent le style naturel de Stendhal (si nous acceptons, comme Stendhal lui-même, d'entendre par « naturel » la faible présence de figures de style d'insistance ou d'analogie), et d'autre part, les caractéristiques rythmiques dans les phrases de Stendhal ne semblent pas améliorer la clarté du propos.

L'analyse statistique et littéraire des phrases clivées montre les spécificités de cette figure de style chez Chateaubriand, Flaubert, Hugo et Stendhal. Les spécificités sont liées à la typologie d'usage, aux fréquences d'occurrences et aux phénomènes stylistiques liés à ce patron. La plupart de nos découvertes confirment les études littéraires sur le style de chaque auteur. Remarquons, de plus, que les phrases clivées de Stendhal sont loin d'être toujours claires et parfaitement interprétables, d'après nos constats.

5.4.3 La répétition

La dernière partie de l'application de notre méthode d'extraction supervisée est consacrée à l'analyse quantitative et qualitative des répétitions dans les romans de Chateaubriand, Flaubert, Hugo et Stendhal. La méthode d'extraction nous permet de relever, d'une manière automatique, les structures morphosyntaxiques répétitives qui possèdent également des relations syntaxiques de dépendance identiques. Les analyses quantitatives dans cette partie se concentrent sur les fréquences de cette figure de style dans les romans des quatre auteurs, elles présenteront également la typologie des patrons les plus fréquents. Nous étudierons ensuite les patrons dans les textes originaux afin de mettre en évidence leurs fonctions stylistiques.

5.4.3.1 Analyses quantitatives des répétitions

Les répétitions sont extraites par notre méthode d'extraction supervisée, en appliquant les profondeurs d'extraction de 3 et de 4. Les résultats issus des deux profondeurs d'extraction sont combinés dans un seul fichier pour chaque auteur. Les deux profondeurs d'extractions ont toutes pour objectif de relever les structures syntaxiques répétitives, ainsi, il n'existe pas de différences qualitatives entre les résultats.

Les profondeurs 3 et 4 ne nous permettent pas de relever toutes les occurrences de répétitions, mais nous jugeons que les autres profondeurs, soit trop petites et ainsi imprécises, soit trop grandes pour donner suffisamment de résultats, ne sont pas idéales pour révéler les caractéristiques stylistiques (pour les discussions sur les profondeurs d'extraction, voir la section 3.7.2 du Chapitre 3).

Le tableau suivant présente les fréquences absolues et relatives de cette figure de style dans les romans des quatre auteurs. Les fréquences relatives sont obtenues en

divisant les fréquences absolues par le nombre de phrases total pour chaque auteur dans notre corpus :

Auteurs	Fréquences absolues	Fréquences relatives
Chateaubriand	288	1.5471%
Flaubert	299	0.7999%
Hugo	1204	1.1757%
Stendhal	128	0.2866%

Tableau 5.8 Fréquences des répétitions pour les quatre auteurs.

Nous constatons un phénomène intéressant concernant l'emploi des figures de style dans les romans des quatre auteurs. En combinant les résultats statistiques sur l'inversion du sujet, le clivage et la répétition, nos résultats sont identiques, c'est-à-dire que Chateaubriand est alors toujours l'auteur dont les romans contiennent le plus de figures de style. Les romans de Chateaubriand présentent une densité de figures de style élevée par rapport aux autres auteurs dans notre corpus. Stendhal est l'auteur qui utilise le moins de figures de style dans ses romans. La détestation de Stendhal du style « pompeux » (avec les figures d'analogie et d'insistance) est prouvée par nos résultats statistiques sur les figures de style. Évidemment, ce constat se base seulement sur les trois figures de style que nous avons pu extraire, mais cela pourrait être, plus largement, représentatif des préférences de chaque auteur en matière d'emploi des figures de style dans ses textes.

Nous procédons ensuite à l'étude statistique de la typologie des répétitions, à savoir les patrons syntaxiques qui se répètent dans la prose de chaque auteur. Le tableau suivant montre les 10 patrons de répétition les plus fréquents pour les quatre auteurs :

Chateaubriand		Flaubert		Victor Hugo		Stendhal	
PATRONS	FREQUENCES	PATRONS	FREQUENCES	PATRONS	FREQUENCES	PATRONS	FREQUENCES
le NN de le NN	50	, VB le NN	26	le NN de le NN	90	le plus ADJ	17
à le NN de le NN	21	le NN de le NN	26	ADJ et ADJ	60	ADJ et ADJ	15
de le NN de NP	18	, de le NN de NN	20	, VB le NN	36	le NN de le NN	8
le NN de NP	18	, le NN de le NN	20	le NN de NP	32	NN de NP	6
de le NN de le NN	13	, le NN de NN	7	, le NN de le NN	25	NUM ou NUM	5
, le NN de le NN	12	, VB son NN	6	de le NN de NP	24	de NN de NP	5
ADJ et ADJ	7	, le NN de son NN	6	de le NN de le NN	23	ADV de NN	5
le plus ADJ	5	le NN de son NN	6	à le NN de le NN	20	de le NN de le NN	4
, le NN de NP	5	, NN de le NN	6	le NN de NN	18	de le NN de NP	3

Tableau 5.9 Patrons de répétition les plus fréquents pour les quatre auteurs.

Les résultats statistiques nous permettent de relever les caractéristiques stylistiques suivantes :

1. Pour Chateaubriand, les structures répétitives se construisent souvent avec les patrons nominaux. Par exemple,

225. *Il pousse un cri, comme un coupable frappé du glaive des bourreaux, comme un assassin percé de la pointe des remords.* (Les Martyrs, Chateaubriand)

2. Flaubert est l'auteur qui répète le plus les patrons verbaux :

226. *Pour cent francs par an, elle faisait la cuisine et le ménage, cousait, lavait, repassait, savait brider un cheval, engraisser les volailles, battre le beurre, et resta fidèle à sa maîtresse, — qui cependant n'était pas une personne agréable.* (Un Cœur simple, Flaubert)

227. *Julien accourut à son aide, détruisit l'armée des infidèles, assiégea la ville, tua le calife, coupa sa tête, et la jeta comme une boule par-dessus les remparts.* (La Légende de Saint-Julien l'Hospitalier, Flaubert)

La description séquentielle des « mouvements » est une des caractéristiques syntaxiques les plus connues de la phrase de Flaubert selon les critiques

littéraires²⁹⁸ (Thibaudet, 1935).

3. Les répétitions de Victor Hugo présentent une plus grande diversité : un mélange de patrons nominaux, verbaux et adjectivaux a été constaté parmi les patrons les plus fréquents. Cela confirme d'une certaine manière le style « divers » et « hétérogène » de Victor Hugo, selon les études littéraires.
4. Les répétitions de Stendhal sont avant tout marquées par la forte présence de patrons adjectivaux « le plus ADJ » et « ADJ et ADJ » :

*228. Mais pour s'amuser lui-même, tout en lui faisant la cour la plus assidue, il avait grand soin toutes les fois qu'il trouvait l'occasion de lui parler en particulier, de ne lui adresser que **les mots les plus indifférents et les propos les plus décolorés.*** (Mina de Vanghel, Stendhal)

*229. M. des Ramiers, **le plus moral, le plus fénelonien** des rédacteurs du journal ministériel par excellence, récemment nommé député à Escorbiac, dans le Midi, à une majorité de deux voix, faisait une cour assidue au ministre et à madame la comtesse de Vaize.* (Lucien Leuwen, Stendhal)

L'utilisation du patron adjectival « le plus ADJ » pour exprimer le superlatif et du patron « ADJ et ADJ », qui renforce également les qualités de l'objet nominal, est une des caractéristiques les plus proéminentes des phrases de Stendhal.

L'analyse statistique des fréquences et de la typologie des répétitions syntaxiques chez les quatre auteurs nous révèle des caractéristiques stylistiques qui peuvent aider à différencier les écrits de chaque auteur. Les uns préfèrent répéter les patrons nominaux alors que les autres préfèrent les adjectivaux et verbaux. Les répétitions syntaxiques de

²⁹⁸ Thibaudet, Albert. 1935. *Op. cit.*.

Victor Hugo sont aussi remarquables car elles sont marquées par une grande variété. Dans la partie suivante, nous examinons les patrons spécifiques à chaque auteur dans le texte original, afin d'étudier les caractéristiques stylistiques.

5.4.3.2 Analyses qualitatives des répétitions

- **Chateaubriand**

Nous commençons par relever les caractéristiques stylistiques dans les répétitions de Chateaubriand.

-Le superlatif en répétition

Il s'agit du patron « le plus ADJ » en répétition. Nous avons constaté dans le Tableau 5.9 qu'il est parmi les patrons les plus fréquents chez Chateaubriand. Effectivement, le superlatif en répétition est très présent dans nos résultats :

*230. Toi qui, parmi ces déserts, es pleine de grandeur; parce que tu es pleine d'innocence ! amitié sainte ! Prête-moi **tes paroles les plus fortes et les plus naïves, ta voix la plus mélodieuse et la plus touchante**, tes sentiments exaltés, tes feux immortels, et toutes les choses ineffables qui sortent de ton cœur, pour chanter les sacrifices que tu inspires ! (Les Natchez, Chateaubriand)*

*231. Quiconque est frappé de son trait divin ou brûlé de son flambeau céleste, embrasse avec transport **les dévouements les plus héroïques, les entreprises les plus périlleuses, les sacrifices les plus douloureux**. (Les Martyrs, Chateaubriand)*

232. *Il repousse dans la nuit des pôles, et retient sous des chaînes de glace les brouillards, les nuages et les tempêtes ; il connaît les écueils les plus cachés, les détroits les plus déserts, les terres les plus lointaines, et les découvre tour à tour au génie de l'homme ;* (Les Martyrs, Chateaubriand)

Ce phénomène n'est pas exclusif de Chateaubriand (Stendhal l'utilise également), mais il est spécifique par rapport aux répétitions de Hugo et de Flaubert, qui, à leur tour, ne l'apprécient pas autant.

-La symétrie rythmique.

Le rythme dans les phrases de Chateaubriand, d'après Charles Bruneau, est souvent harmonieux et soigneusement coupé. Dans la phrase suivante, les deux premières propositions en répétition sont organisées pour présenter la symétrie rythmique en 6 syllabes (6-6).

233. *On redoubla mes gardes, on multiplia mes chaînes, on écarta mon amante.*
(Atala, Chateaubriand)

Les deux occurrences répétitives du patron « un NN de NN et de NN » dans l'exemple suivant sont non seulement morphosyntaxiquement identiques, mais elles sont également octosyllabiques.

234. *Dans cette pensée, Aben-Hamet attendit avec impatience le lendemain pour faire connaître sa résolution à Blanca et changer **une vie de tristesse et de larmes en une vie de joie et de bonheur.*** (Les Aventures du dernier Abencerage, Chateaubriand)

-L'image

Considérées comme une des caractéristiques les plus proéminentes du style de Chateaubriand, les images sont également présentes dans nos résultats. La répétition consiste soit à mettre en comparaison deux occurrences d'un patron répétitif (la première phrase), soit à créer le parallélisme entre les deux comparants (la deuxième phrase) :

235. *La lune décroissante, **paraissait au milieu du ciel**, comme les lampes demi-circulaires que les premiers fidèles **allumaient aux tombeaux des martyrs**.* (Les Martyrs, Chateaubriand)

236. *Il pousse un cri, **comme un coupable frappé du glaive des bourreaux**, **comme un assassin percé de la pointe des remords**.* (Les Martyrs, Chateaubriand)

-Antithèse

La répétition de Chateaubriand peut aussi mettre en place l'antithèse dans les patrons répétés. Les trois exemples d'au-dessous présentent ce phénomène :

237. *C'est donc, repartit la fille d'Homère, comme le beau jeune homme de l'Arabie, pleuré des femmes de Byblos, et rendu à la lumière des cieux par la volonté de Jupiter ? — Cymodocée, répliqua Eudore avec une douce sévérité, vous connaîtrez quelque jour combien cette comparaison est impie et sacrilège : au lieu **des mystères de honte et de plaisir**, vous voyez ici **des miracles de modestie et de douleur** ;* (Les Martyrs, Chateaubriand)

238. *Quelquefois elle s'élançe comme enivrée de plaisir et se retire comme accablée de douleur.* (Les Aventures du dernier Abencerage, Chateaubriand)

239. *Après avoir conversé avec ce qu'il y a de plus sublime dans la société, je vais entendre les paroles de ce qu'il y a de plus simple dans la nature : mais quels que soient les lieux où le Grand-Esprit conduise mes pas, sous l'arbre, au bord du fleuve, sur le rocher, je rappellerai tes leçons et je tâcherai de devenir sage de ta sagesse.* (Les Natchez, Chateaubriand)

Notre étude sur la répétition dans les romans de Chateaubriand nous permet de découvrir les phénomènes stylistiques autour de ce patron. Comme on l'a vu pour les autres motifs, la répétition elle-même, en dehors de son contexte, est souvent peu informative pour caractériser le style de l'auteur, mais ses occurrences dans le texte original nous révèlent souvent des caractéristiques importantes du style de Chateaubriand.

● **Flaubert**

-L'Accumulation

Dans nos résultats des répétitions syntaxiques pour Flaubert, nous remarquons un phénomène d'accumulation similaire à celui de Victor Hugo. Les répétitions, contrairement à celles de Chateaubriand, sont abondantes :

240. *Ce sont des têtes d'alligators sur des pieds de chevreuil, des hiboux à queue de serpent, des pourceaux à muflle de tigre, des chèvres à croupe d'âne, des grenouilles velues comme des ours, des caméléons grands comme des*

hippopotames, des veaux à deux têtes dont l'une pleure et l'autre beugle, des fœtus quadruples se tenant par le nombril et valsant comme des toupies, des ventres ailés qui voltigent comme des moucheron. (La Tentation de Saint-Antoine, Flaubert)

Nous avons évoqué dans les analyses quantitatives que les patrons les plus fréquents dans les répétitions flaubertiennes sont les patrons verbaux, voici deux exemples de ces patrons incluant de multiples répétitions :

*241. Chaque jour, en effet, **tu bouleverses la terre, tu creuses des canaux, tu bâtis des palais, tu enfermes les fleuves entre des pierres, tu cueilles l'herbe, tu la pétris et tu la manges** ;* (Mémoires d'un fou, Flaubert)

*242. Je voudrais avoir des ailes, une carapace, une écorce, souffler de la fumée, **porter une trompe, tordre mon corps, me diviser partout, être en tout, m'émaner avec les odeurs, me développer comme les plantes, couler comme l'eau, vibrer comme le son, briller comme la lumière, me blottir sur toutes les formes, pénétrer chaque atome, descendre jusqu'au fond de la matière, — être la matière !*** (La Tentation de Saint-Antoine, Flaubert)

Nous remarquons que les répétitions de Flaubert sont plus proches de celles de Victor Hugo, il s'agit souvent de plusieurs reprises d'un même patron, alors que les répétitions de Chateaubriand sont plus discrètes, avec seulement deux ou trois reprises.

-La période ternaire

Pour les phrases de Flaubert, il est impossible d'ignorer la période ternaire, présente également dans les phrases avec répétitions :

243. *Viennent à moi ceux qui sont couverts de vin, / ceux qui sont couverts de boue, / ceux qui sont couverts de sang ;* (La Tentation de Saint-Antoine, Flaubert)

Les exemples suivants montrent deux phrases au rythme ternaire, avec les deux dernières propositions comprenant la figure de style préférée de Flaubert²⁹⁹ (Thibaudet, 1935), l'ellipse (omission d'un ou plusieurs mots dans un énoncé dont le sens reste clair³⁰⁰) :

244. *Les poutrelles du plafond étaient vermoulues, / les murailles noires de fumée, / les carreaux gris de poussière.* (Trois contes, Flaubert)

245. *J'ai donné mes mains aux manchots, / mes jambes aux boiteux, / mes prunelles aux aveugles ;* (La Tentation de Saint-Antoine, Flaubert)

-L'image

La mise en place d'un tableau par analogie, l'image, est très fréquente dans les résultats. Cette figure se combine avec la répétition, pour donner une description plus précise.

246. *On entendait, à côté du lourd patois dorien, retentir les syllabes celtiques bruissantes comme des chars de bataille, et les terminaisons ioniennes se heurtaient aux consonnes du désert, âpres comme des cris de chacal.* (Salammbô, Flaubert)

247. *Ainsi, comme les échos d'une voix qui s'éloigne, comme les effluves d'un*

²⁹⁹ Thibaudet, Albert. 1935. *Op. cit.*

³⁰⁰ « Ellipse », <https://www.cnrtl.fr/definition/ellipse>, consulté le 1 juillet, 2021.

parfum qui s'évapore, comme les feux du soleil qui se couche, les Puissances émanées du Principe vont toujours s'affaiblissant. (La Tentation de Saint-Antoine, Flaubert)

248. *Retenus à la terre par nos chevelures, longues comme des lianes, nous végétons à l'abri de nos pieds, larges comme des parasols.* (La Tentation de Saint-Antoine, Flaubert)

Nos analyses sur la répétition dans les romans de Flaubert révèlent les caractéristiques stylistiques propres à cet auteur. À l'exception de la période ternaire et de l'image, nous remarquons également la ressemblance entre Flaubert et Victor Hugo en matière des répétitions abondantes, proches de l'accumulation.

- **Stendhal**

-La répétition des patrons adjectivaux

Nous n'avons malheureusement trouvé qu'une seule caractéristique concernant la répétition dans les romans de Stendhal. Il s'agit de sa tendance (dont la spécificité a été démontrée dans les analyses quantitatives) à répéter des patrons adjectivaux dans ses phrases. Les cinq exemples d'au-dessous présentent cinq patrons adjectivaux différents, qui apparaissent souvent dans nos résultats des répétitions stendhaliennes : « à le NN le plus ADJ », « le NN le plus ADJ », « un NN ADJ et ADJ », « de NN si ADJ » et « ADJ et ADJ et ADJ ».

249. *Jamais Mathilde n'avait eu d'amant, et dans cette circonstance de la vie qui donne quelques illusions tendres même aux âmes les plus sèches, elle était en*

proie aux réflexions les plus amères. (Le Rouge et le Noir, Stendhal)

250. *Dans un cœur passionné, l'amour est sujet à exagérer les nuances les plus légères, il en tire les conséquences les plus ridicules.* (La Chartreuse de Parme, Stendhal)

251. *Féder prit sa main, tandis qu'elle le regardait d'un air indécis et surpris, et il lui dit d'un ton philosophique et froid : — Je suis plus honnête homme qu'amant, je n'ose vous dire que je vous aime avec passion, de peur que cela ne cesse d'être vrai un jour ; et, sur toutes choses, je ne voudrais pas tromper une amie qui a pour moi des sentiments si sincères.* (Féder, Stendhal)

252. *Délivré tout à coup et d'une façon si imprévue, par un vieux médecin de province, de l'effroyable ennui qui l'accablait depuis deux mois, Lucien n'eut pas le courage de se priver d'une vision si amusante.* (Lucien Leuwen, Stendhal)

253. *Mon sort est-il donc de passer ma vie entre des légitimistes fous, égoïstes et polis, adorant le passé, et des républicains fous, généreux et ennuyeux, adorant l'avenir ?* (Lucien Leuwen, Stendhal)

Nous remarquons que les adjectifs dans ces patrons fréquents ne sont ni antithétiques, ni synonymes. L'association des adjectives avec le nom qu'ils désignent ne crée pas d'effet surprenant. Les occurrences de ces patrons répétitifs ne suivent pas de rythme particulier (par rapport au rythme ternaire de Flaubert). Bref, la seule caractéristique que nous pouvons confirmer chez Stendhal est sa préférence pour la répétition de patrons adjectivaux, et ces occurrences sont généralement dépourvues d'autres phénomènes stylistiques (contrairement à ce que nous avons vu chez Chateaubriand, Flaubert et Hugo). Cela confirme la détestation de Stendhal du style

« pompeux », qui engage souvent beaucoup de figures de style d'analogie et d'insistance, mais il nous semble difficile de faire un lien entre la répétition stendhalienne et la clarté que l'auteur cherche à atteindre. Dans la section suivante, nous essayons de compléter nos analyses sur le style de Stendhal en recourant aux patrons identificateurs.

Comme tous les patrons caractérisants, la répétition, en elle-même, en dehors de son contexte, est peu informative pour caractériser le style (son rôle identificateur, néanmoins, est indéniable). Le retour au texte nous permet de révéler les caractéristiques stylistiques liées à cette figure de style. Nous remarquons que les phénomènes stylistiques dans la répétition confirment les études littéraires en leur fournissant également des preuves statistiques.

5.5 Application de la méthode d'extraction non-supervisée à notre corpus

Les analyses quantitatives et qualitatives des patrons caractérisants (l'inversion, le clivage et la répétition) nous ont permis de relever partiellement les caractéristiques stylistiques de Chateaubriand, Flaubert, Hugo et Stendhal. Nous avons également constaté que les patrons syntaxiques sont capables de caractériser le style par le truchement des analyses statistiques et littéraires. Pourtant, les trois figures de style présentées dans la section précédente ne sont pas suffisantes pour donner une vue générale du style des auteurs, d'une part parce que les fréquences observées ne sont pas très importantes, d'autre part parce que les caractéristiques stylistiques ne peuvent pas être réduites à seulement trois figures de style d'ordre syntaxique. C'est pour cette raison que nous avons introduit dans la thèse la méthode d'extraction non-supervisée, qui donne potentiellement accès à tous les patrons syntaxiques remarquables statistiquement, quelle que soit leur structure.

Afin de produire les analyses concernant ces patrons, les méthodes statistiques doivent être mises en place. Nous avons dit dans la section 4.5 du Chapitre 4 que notre méthode statistique est l'Analyse factorielle des correspondances (AFC), basée sur le test du χ^2 . La préparation des données d'entrée ainsi que le mécanisme de calcul ont également été discutés dans cette même section. La méthode d'analyse va maintenant être appliquée aux romans de Chateaubriand, Flaubert, Hugo et Stendhal, afin de montrer les différences stylistiques de ces quatre auteurs au niveau syntaxique.

Nous essayons de montrer dans cette section que les patrons syntaxiques identificatoires possèdent la capacité de différencier et de caractériser les styles littéraires. La combinaison de nos méthodes d'extraction supervisée et non-supervisée donne une vue globale des caractéristiques syntaxiques et stylistiques d'un auteur.

5.5.1 Analyses quantitatives des patrons syntaxiques

La préparation des données d'entrée pour les analyses statistiques des romans de nos quatre auteurs est identique à celle suivie pour les neuf romans de Victor Hugo. Deux profondeurs d'extraction sont appliquées sur l'analyse syntaxique en dépendance : les profondeurs 3 et 4 (pour la discussion sur les profondeurs d'extraction idéales, voir la section 3.7.2 du Chapitre 3). Tous les romans d'un certain auteur sont regroupés dans un seul fichier vu que le but de nos analyses est de montrer les différences stylistiques entre auteurs, et non pas entre œuvres individuelles. Nous sélectionnons les 50 patrons les plus fréquents de profondeur 3 et les 50 patrons les plus fréquents de profondeur 4 dans un fichier combiné des patrons syntaxiques pour les quatre auteurs. Nous avons choisi d'étudier les patrons les plus fréquents pour chaque auteur parce que les fréquences élevées de ces patrons sont plus à même de contenir des caractéristiques stylistiques remarquables (par opposition aux patrons apparus seulement une ou deux fois dans le texte). Ainsi, nos données d'entrée comprennent tous les romans des quatre auteurs et les 100 patrons les plus fréquents à l'échelle du corpus, issus de deux profondeurs d'extraction différentes.

Ces données sont ensuite traitées par l'application de la méthode d'Analyse factorielle des correspondances (AFC), grâce à un programme écrit en R. La figure suivante présente la répartition des quatre auteurs sur un espace bidimensionnel d'après le calcul de spécificité basé sur les 100 patrons syntaxiques les plus fréquents :

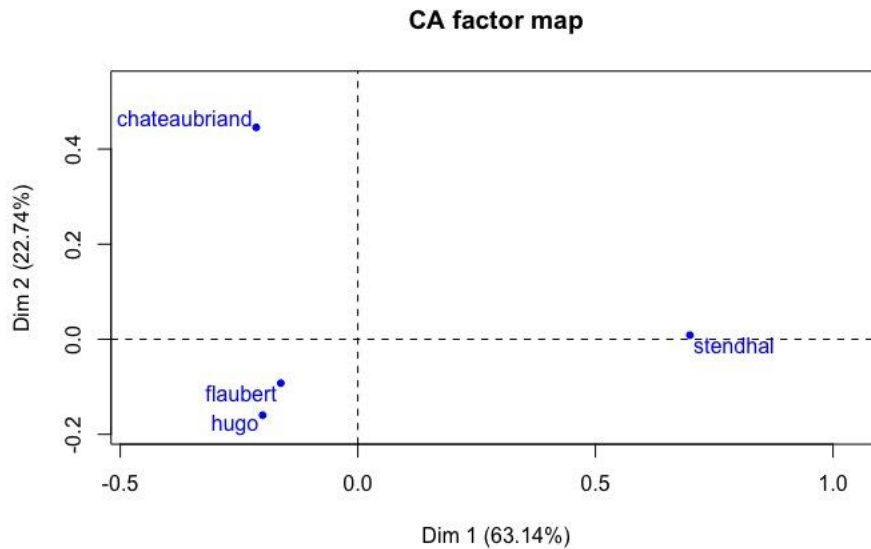


Figure 5.3 CA Factor Map pour les quatre auteurs.

Les résultats de l'AFC pour les patrons relevés de manière non-supervisée sont conformes aux résultats sur les patrons caractérisants (l'inversion, le clivage et la répétition) : rappelons-nous que les résultats statistiques sur les figures de style montrent également la ressemblance entre Hugo et Flaubert, tout en mettant en évidence la densité élevée de figures de style dans les écrits de Chateaubriand. Les écrits en prose de Stendhal sont ceux qui contiennent, d'une manière générale, le moins de figures stylistiques. Sur la carte ci-dessus, le même phénomène est constaté : les romans de Chateaubriand sont éloignés (sur l'axe horizontal) des trois autres auteurs alors que les romans de Stendhal sont éloignés (sur l'axe vertical). Hugo et Flaubert présentent une ressemblance au niveau de l'utilisation des patrons syntaxiques.

Afin de connaître quels sont les patrons spécifiques, et ainsi identificatoires, pour chaque auteur, la figure suivante présente la corrélation entre patrons et auteurs. La corrélation met en évidence la contribution de chaque patron à la différenciation des auteurs, et c'est par ce calcul que nous repérons les patrons spécifiques pour chaque auteur.

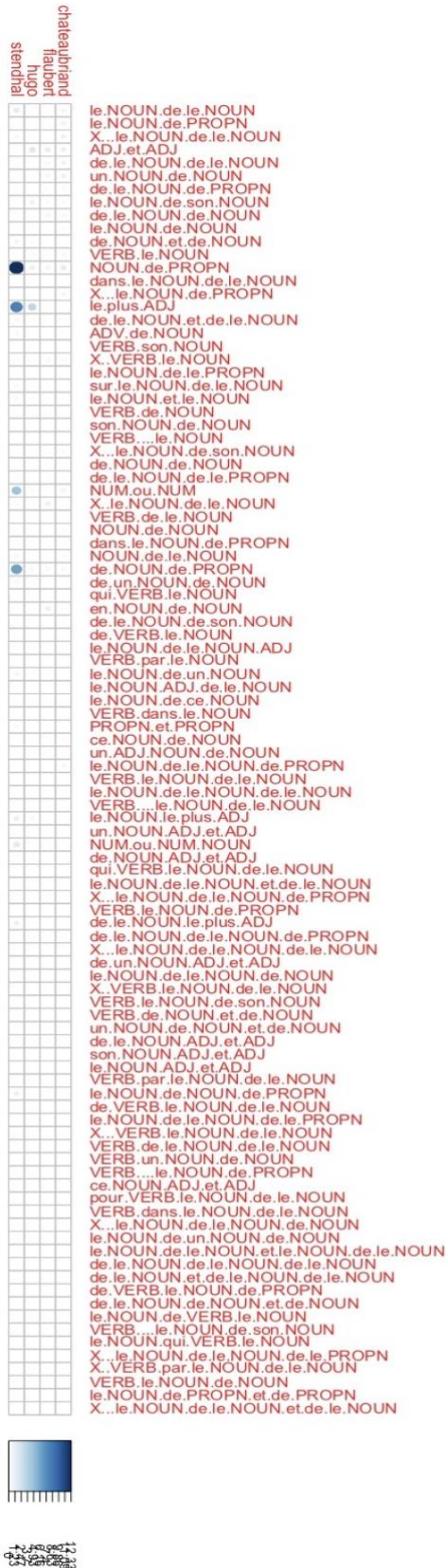


Figure 5.4 Corrélations entre les patrons et les auteurs.

Plus le cercle est foncé, plus le patron est spécifique d'un auteur donné, c'est-à-dire, plus il contribue à la différenciation de cet auteur par rapport aux autres écrivains du corpus³⁰¹. Le tableau ci-dessous présente les listes des patrons spécifiques pour chaque auteur de notre corpus. Nous avons décidé de ne pas inclure les patrons avec des noms propres ou des chiffres, parce qu'ils présentent souvent peu d'informations stylistiques :

Chateaubriand	Flaubert	Hugo	Stendhal
à le NN de le NN	, à le NN de le NN	, à le NN de le NN	à le NN de le NN
à le NN de son NN	, de le NN de NN	ADJ et ADJ	ADV de NN
ADJ et ADJ	, le NN de le NN	ADP de le NN	de le plus ADJ NN
de le NN de le NN	, un NN de NN	de le NN de le NN	le NN de un NN
le NN de son NN	, VBE de le NN	de le NN de NN	le plus ADJ
le NN de le NN	, VBE de NN	le plus ADJ	le plus ADJ NN
un NN de NN	à le NN de le NN ,	un NN de NN	le NN de le NN
	ADJ et ADJ		NN de le NN
	de le NN de le NN		NN de son NN
	en NN de NN		un ADV ADJ
			un NN ADV ADJ
			de le NN le plus ADJ
			le NN le plus ADJ

Tableau 5.10 Patrons spécifiques d'après l'AFC pour chaque auteur.

Avant de procéder au retour au texte, nous pouvons déjà constater quelques caractéristiques superficielles :

1. Presque tous les patrons spécifiques les plus fréquents chez Chateaubriand sont des patrons nominaux.
2. Flaubert est le seul auteur à avoir un nombre considérable de patrons spécifiques commençant ou se terminant par une virgule. Cela indique l'existence de deux ou plusieurs propositions dans les phrases de Flaubert. Ces propositions servent sans doute à découper les phrases longues et à créer un rythme équilibré. Les analyses

³⁰¹ En raison du nombre élevé de patrons dans notre calcul, la visualisation du résultat de corrélation présentée dans la Figure 5.4 n'est pas très claire. Les patrons spécifiques avec les cercles foncés sont listés dans le Tableau 5.10.

qualitatives, présentées dans la section suivante, permettront de vérifier cette hypothèse.

3. Les patrons les plus spécifiques de Stendhal sont des patrons adjectivaux : « le plus ADJ », « le plus ADJ NN », « un ADV ADJ », « un NN ADV ADJ », « de le NN le plus ADJ », « le NN le plus ADJ ». Le superlatif est également très présent dans ces patrons spécifiques. Par exemple, il est l'auteur qui utilise le plus le patron « le plus ADJ ».

Les caractéristiques statistiques nous permettent d'avoir une idée générale des composants syntaxiques des phrases en prose de chaque auteur. Pour mieux déchiffrer les caractéristiques syntaxiques et stylistiques des phrases, nous allons maintenant étudier ces patrons identificatoires de manière plus précise, à travers des exemples en contexte.

5.5.2 Analyses qualitatives des patrons syntaxiques

Nous procédons dans cette section à l'examen des patrons spécifiques pour chaque auteur. Seuls les patrons suremployés par un auteur donné sont pris en considération dans nos analyses, parce que leur fréquence élevée est plus susceptible de donner accès à des caractéristiques stylistiques généralisables. Nous laissons de côté les patrons construits autour d'un nom propre et de chiffres du fait de leur manque d'intérêt stylistique.

A) Chateaubriand

- « à le NN de le NN », « à le NN de son NN »

Les deux patrons sont suremployés uniquement dans les romans de Chateaubriand (744 occurrences pour le premier et 179 pour le deuxième). Nous les regroupons parce que leurs fonctions grammaticales sont identiques, il s'agit soit d'un circonstanciel, soit d'un complément d'objet indirect. Deux phénomènes stylistiques sont directement liés à ces deux patrons dans les œuvres de Chateaubriand.

-L'image

Quand ces deux patrons sont précédés d'un verbe de comparaison (« ressembler »), ou un adjectif de comparaison (« semblable »), Chateaubriand a fréquemment recours à une « image » :

254. *Nous aperçûmes à travers les arbres un jeune homme, qui, tenant à la main un flambeau, ressemblait **au Génie du printemps**, parcourant les forêts pour ranimer la nature.* (Atala, Chateaubriand)

255. *À quelques pas plus loin, dans la montagne, vous n'entendez plus qu'un langage grossier, semblable **au croassement des corbeaux**.* (Les Martyrs, Chateaubriand)

256. *On entendait un bruit confus semblable **aux échos des vagues** dans les rochers.* (Les Natchez, Chateaubriand)

257. *Les mystères couvraient son front d'un bandeau, et les amours, nouant autour d'elle les crêpes de son écharpe, ne laissaient paraître qu'une de ses mamelles, semblable **au globe de la lune**.* (Les Natchez, Chateaubriand)

258. *Des cris étouffés, qui ressemblaient **au rôle de la mort**, s'échappèrent peu à peu du sein d'Akansie.* (Les Natchez, Chateaubriand)

-La nominalisation d'un adjectif

La nominalisation d'un adjectif présente des effets stylistiques divers dans les phrases de Chateaubriand. Dans l'exemple suivant, au lieu d'écrire « les rochers pâles », l'auteur préfère « la pâleur des rochers » afin d'éviter la répétition d'une structure syntaxique « un NN ADJ et ADJ » (« une teinte rouge et ardente ») dans la phrase assez courte.

259. *Une teinte rouge et ardente succède à **la pâleur des rochers**.* (Les Martyrs, Chateaubriand)

En revanche, les deux exemples suivants montrent la nominalisation d'un adjectif visant à créer la symétrie rythmique avec un autre patron syntaxique à proximité « le NN de le NN » :

260. *Combien votre aimable fille, que vous comparez si justement à une colombe, trouverait des soupirs plus touchants encore, si **la pudeur du sujet** répondait à **l'innocence de la voix** !* (Les Martyrs, Chateaubriand)

261. *De crainte qu'on ne dérobe quelque partie de sa fortune à **l'avidité de l'empereur**, on force, par **la violence des supplices**, les enfants à déposer contre*

leurs pères, les esclaves contre leurs maîtres, les femmes contre leurs maris. (Les Martyrs, Chateaubriand)

Bien que les deux patrons soient statistiquement spécifiques pour Chateaubriand, nous remarquons que leurs fonctions stylistiques sont limitées : dans la plupart des occurrences, ces deux patrons jouent simplement le rôle d'un circonstanciel (« au fond de ... », « au milieu de ... »).

- **« de le NN de le NN »**

Ce patron apparaît 583 fois dans les phrases de Chateaubriand, il est considéré comme suremployé par rapport à ses occurrences chez les autres auteurs, en particulier Flaubert (164 occurrences). Nous soulignons trois caractéristiques stylistiques liées à ce patron.

-L'antithèse

Nous avons évoqué dans la section 5.2 que l'antithèse est « une vraie manie³⁰² » (Sainte-Beuve, 1861, t. I, p. 397) pour Chateaubriand. Les analyses du patron « de le NN de le NN » correspondent à cette caractéristique particulière de l'auteur. L'exemple suivant met en place deux mots antithétiques « vertu » et « misère » :

262. Nous bénissons la Providence qui, sur la faible tige d'une fleur, avait placé cette source limpide au milieu des marais corrompus, comme elle a mis l'espérance

³⁰² Sainte-Beuve, Charles-Augustin. 1861. *Op. cit.*, t. I, p. 397.

*au fond des cœurs ulcérés par le chagrin, comme elle a fait jaillir la vertu **du sein des misères de la vie.*** (Atala, Chateaubriand)

D'une manière similaire, la phrase suivante présente la naissance des « arts » du sein de la « barbarie » :

*263. Les arts renaissant **du sein de la barbarie**, l'honneur, la loyauté, la chevalerie des anciens temps, unis à la politesse des siècles civilisés, les tourelles gothiques ornées des ordres de la Grèce, et les dames gauloises rehaussant la richesse de leurs atours par l'élégance athénienne.* (Les Aventures du dernier Abencerage, Chateaubriand)

Le dernier exemple présente le personnage né « pour l'amour », mais qui habite éternellement dans la « haine » :

*264. Né pour l'amour, éternel habitant **du séjour de la haine**, il supporte impatiemment son malheur ; trop délicat pour pousser des cris de rage, il pleure seulement, et prononce ces paroles avec de profonds soupirs : « Dieux de l'Olympe, et vous que je connais moins, divinités du Brachmane et du Druide, je n'essaierai point de le cacher ; oui, l'enfer me pèse !* (Les Martyrs, Chateaubriand)

-La métaphore

Ce patron particulier, « de le NN de le NN », pourrait aussi être employé pour mettre en place une métaphore. Cependant, nous remarquons que la métaphore de Chateaubriand est différente de celle de Victor Hugo, elle est plus discrète et moins surprenante (rappelons-nous que la métaphore de Hugo est souvent étonnante en

combinant des mots issus des sémantismes différents). Dans la phrase suivante, la rosée est comparée aux « larmes du matin » :

265. *J'allai cueillir une rose de magnolia, et je la déposai humectée **des larmes du matin**, sur la tête d'Atala endormie.* (Atala, Chateaubriand)

Ou encore, dans la phrase suivante :

266. *De temps en temps, le religieux plongeait un rameau fleuri dans une eau consacrée, puis secouant la branche humide, il parfumait la nuit **des baumes du ciel**.* (Atala, Chateaubriand)

Le troisième exemple montre une métaphore proche de l'hypallage, figure de style qui consiste à attribuer, à un ou plusieurs mots d'une phrase, ce qui convient à d'autres³⁰³. Le mot « sommeil » est employé pour désigner la « terre », mais en réalité, c'est Atala qui est morte et ainsi succombe au « sommeil ».

267. *Et j'achevai de couvrir Atala **de la terre du sommeil**.* (Atala, Chateaubriand)

Le dernier exemple montre une métaphore proche de celle de Hugo, en combinant mots abstraits et mots concrets. L'emploi du pluriel pour les « ombres » donne l'impression que la « mort » est une existence physique :

268. *Une grande et dernière faute, en le jetant dans un grand malheur, l'a fait sortir **des ombres de la mort**.* (Les Martyrs, Chateaubriand)

³⁰³ « Hypallage », <https://www.cnrtl.fr/definition/hypallage>, consulté le 1 juillet, 2021.

-La répétition

La répétition chez Chateaubriand, comme nous l'avons constaté dans la section 5.4.3, est discrète. Nous trouvons rarement de phrases avec des répétitions abondantes, engageant plus de deux syntagmes répétés. Les deux phrases suivantes montrent ce phénomène avec le patron « de le NN de le NN » en répétition :

269. Heureux, si j'ai prouvé que notre religion peut lutter sans crainte avec les plus grandes beautés d'Homère, et qu'elle donne, dans l'infortune, un courage au-dessus de la rage des persécuteurs, et de la cruauté des bourreaux ! (Les Martyrs, Chateaubriand)

270. Il pousse un cri, comme un coupable frappé du glaive des bourreaux, comme un assassin percé de la pointe des remords. (Les Martyrs, Chateaubriand)

- « le NN de le NN », « le NN de son NN »

Nous regroupons dans cette partie les deux patrons nominaux, suremployés dans les romans de Chateaubriand (1254 occurrences pour « le NN de le NN » et 391 occurrences pour « le NN de son NN »). Nous les analysons de la même manière parce qu'il n'y a pas de différence radicale entre les deux constructions syntaxiques pour les analyses stylistiques.

-L'image

L'image, une figure d'analogie, est une des caractéristiques stylistiques propres à Chateaubriand. Nous retrouvons dans nos résultats ce phénomène, construit autour des patrons « le NN de le NN » et « le NN de son NN » :

271. *Il faut qu'il coupe **le fil de ses jours**, comme un tisserand coupe **le fil de sa toile** à moitié tissue.* (Les Martyrs, Chateaubriand)

272. *L'indienne, en chantant ces paroles, tremblait, et des larmes coulaient comme des perles **le long de ses joues** : elle ne savait pourquoi, à la vue du frère d'Amélie, elle se souvenait des derniers conseils de sa mère.* (Les Natchez, Chateaubriand)

273. ***Le cœur de l'homme** est comme **l'éponge du fleuve**, qui tantôt boit une onde pure dans les temps de sérénité, tantôt s'enfle d'une eau bourbeuse, quand le ciel a troublé les eaux.* (Atala, Chateaubriand)

274. *Malgré le poids des neiges, ma mémoire a conservé fidèlement le souvenir de cette aventure, comme les colliers gardent **les actions des aïeux**.* (Les Natchez, Chateaubriand)

-La métaphore

La métaphore peut également être présente au sein des deux patrons nominaux considérés. Remarquons que la métaphore chez Chateaubriand est relativement moins audacieuse et moins surprenante que celle de Hugo. Par exemple, la phrase suivante se sert des « plaintes » pour désigner métaphoriquement les qualités sonores du « vent », une métaphore plutôt classique, sans effet de surprise :

275. *À mesure que le bruit des hommes s'affaiblit, celui du désert augmente, et au tumulte des voix succèdent **les plaintes du vent** dans la forêt.* (Atala, Chateaubriand)

Les trois exemples suivants montrent le même phénomène :

276. *J'entendais tes plaintes, je voyais **les flammes de l'enfer** te consumer.* (Atala, Chateaubriand)

277. *Et les guerriers, sillonnés du feu de l'éclair, écoutent en silence **la voix des orages**.* (Les Martyrs, Chateaubriand)

278. *Les voiles blanches, teintes de pourpre par les rayons de l'aurore, s'enflaient aux haleines des zéphyr, et les rames dorées fendaient **le cristal des mers**.* (Les Martyrs, Chateaubriand)

Nous avons étudié dans cette section les patrons spécifiques de la prose de Chateaubriand. La plupart des patrons suremployés dans les romans de Chateaubriand sont nominaux. Les patrons « ADJ et ADJ » ainsi que « un NN de NN » n'ont pas fait l'objet d'une analyse parce qu'ils sont sous-employés par Chateaubriand.

Les caractéristiques que nous soulignons par le truchement des patrons spécifiques correspondent aux traits marquants identifiés par les critiques littéraires pour le style de Chateaubriand : l'antithèse, l'image, la répétition et la métaphore sont présentes dans nos résultats. Nous remarquons également que les répétitions et les métaphores, sous la plume de Chateaubriand, sont différentes de celles de Victor Hugo. Elles sont souvent plus discrètes et moins surprenantes : les répétitions se limitent souvent à moins de trois syntagmes répétitifs, les métaphores sont moins étonnantes parce que le comparé et le comparant proviennent souvent du même sémantisme. Nous concluons que les patrons

syntaxiques sont capables de caractériser le style de Chateaubriand, mais notre méthode présente également des limites. Par exemple, la structure syntaxique privilégiée de Chateaubriand (phrases avec la protase assez courte et l'apodose longue) est difficile à reconnaître par le truchement des patrons syntaxiques. Cette structure syntaxique est omniprésente dans nos résultats, mais elle n'est pas liée à un patron particulier.

B) Flaubert

Nous analysons dans cette section les patrons suremployés dans les romans de Flaubert.

- « **en NN de NN** »

Ce patron est apparu 160 fois dans les romans de Flaubert, il est donc relativement suremployé sous la plume de l'auteur. Nous avons trouvé une seule fonction stylistique particulière à ce patron.

-Le modifieur nominal

Il s'agit d'un groupe nominal qui modifie le substantif qui lui précède :

279. *Près de l'auberge, une fille **en chapeau de paille** tirait des seaux d'un puits ;*
(L'Éducation sentimentale, Flaubert)

280. *Il lui semblait étrange de voir cette belle dame **en robe de nankin** tout au milieu de cette misère.*

281. *Le lit était un grand lit d'acajou en forme de nacelle.* (Madame Bovary, Flaubert)

Nous remarquons que ce patron possède rarement une fonction stylistique, mais sa fréquence élevée dans les romans de Flaubert indique la préférence de ce dernier pour les modifieurs nominaux (au lieu des modifieurs adjectivaux ou adverbiaux).

- « , VB de NN », « , VB de le NN »

Les patrons verbaux avec une virgule au début de la construction sont très spécifiques du style de Flaubert, d'après les calculs statistiques. Le patron « , VB de NN » est apparu 63 fois au total et le patron « , VB de le NN » est apparu 55 fois.

-La construction participiale

La construction participiale a pour but de remplacer « d'une manière brève et efficace » une proposition de subordination. Thibaudet la considère également comme un jeu littéraire sur le temps de narration, propre au style de Flaubert³⁰⁴ (Thibaudet, 1935). Les trois phrases suivantes montrent les occurrences de ce patron avec le participe passé. Le participe exprime une continuité avec laquelle tranchent les moments individuels et saillants des temps verbaux.

282. *Mme Aubain penchait son front, accablée de souvenirs.* (Trois contes, Flaubert)

283. *Antipas, ému de reconnaissance, lui demanda comment elle l'avait.* (Trois contes, Flaubert)

³⁰⁴ Thibaudet, Albert. 1935. *Op. cit.*.

284. *L'homme vêtu de jaune ne céda pas à cette envie, et, lui tendant une coupe d'or où fumait un bouillon de vipère : — Bois ! dit-il, pour que la force des serpents, nés du soleil, pénètre dans la moelle de tes os, et prends courage, ô reflet des Dieux !* (Salammbô, Flaubert)

Les trois exemples suivants présentent ce patron avec comme verbe un participe présent. Le participe oppose son mouvement ralenti ou faible au mouvement rapide et à la force du temps verbal pur.

285. *Britomartis, hurlant de peur, se cramponne aux mailles de son filet.* (La Tentation de Saint-Antoine, Flaubert)

286. *Edgar, étincelant de furie, dominait tous les autres de sa voix plus claire.* (Madame Bovary, Flaubert)

287. *Mais Julien ne se fatiguait pas de tuer, tour à tour bandant son arbalète, dégainant l'épée, pointant du coutelas, et ne pensait à rien, n'avait souvenir de quoi que ce fût.* (Trois contes, Flaubert)

-La coupe flaubertienne

Ce patron joue également un rôle dans la coupe syntaxique flaubertienne. Dans les deux phrases suivantes, le patron est placé à la fin de la phrase. L'effet d'insistance est créé par la brièveté relative de ce patron, isolé du reste de la phrase par une virgule. Cet usage particulier de Flaubert est une des manifestations de son appétence pour la coupe syntaxique spécifique, dont l'effet est proche du rejet³⁰⁵ dans la versification.

³⁰⁵ « Rejet », <https://www.cnrtl.fr/definition/rejet>, consulté le 20 juin, 2021.

288. *Il baisse la figure, écrasé de honte.* (La Tentation de Saint-Antoine, Flaubert)

289. *La nuit, devant eux, s'étendait comme un immense voile sombre, piqué d'argent.* (L'Éducation sentimentale, Flaubert)

290. *Le paquebot, que des femmes halaient en chantant, sortit du port.* (Trois contes, Flaubert)

Le quatrième exemple présente deux occurrences en parallélisme de ce patron vers la fin de la phrase : l'accent est mis sur ces séquences en raison de leur construction brève, répétitive et isolée par la virgule. Nous retrouvons la coupe flaubertienne qui crée un rythme oratoire et harmonieux.

291. *Et la petite fille se jetait à genoux pour les rattraper, riait de plaisir, criait d'effroi.* (L'Éducation sentimentale, Flaubert)

Ce patron peut aussi être employé devant le dernier syntagme, pour l'isoler du reste de la phrase. Il sert ainsi à rallonger la phrase avant la chute brève finale :

292. *Enfin, à minuit, Frédéric, brisé de fatigue, regagna sa maison.* (L'Éducation sentimentale, Flaubert)

293. *La terre, mêlée de cailloux, retomba ;* (L'Éducation sentimentale, Flaubert)

-La répétition

Un autre cas remarquable est la présence de deux patrons au début de la proposition (séparés par une virgule), ce qui leur permet de faire partie d'une répétition syntaxique, souvent en parallélisme :

294. *Des hommes nus pétrissaient des pâtes, broyaient des herbes, agitaient des charbons, versaient de l'huile dans des jarres, ouvraient et fermaient les petites cellules ovoïdes creusées tout autour de la muraille et si nombreuses que l'appartement ressemblait à l'intérieur d'une ruche.* (Salammbô, Flaubert)

295. *Mais L'INSPIRÉ frappant du talon, claquant des doigts, hochant la tête, psalmodie sur un rythme furieux, au son des cymbales et d'une flûte aiguë : Viens ! viens ! viens ! sors de ta caverne !* (La Tentation de Saint-Antoine, Flaubert)

296. *À cheval sur des oiseaux, bercés dans des palanquins, trônant sur des sièges d'or, debout dans des niches d'ivoire, ils songent, voyagent, commandent, boivent du vin, respirent des fleurs.* (La Tentation de Saint-Antoine, Flaubert)

L'analyse du patron « , VB de NN » nous permet de croiser à nouveau quatre caractéristiques du style de Flaubert : la phobie de Flaubert pour les pronoms relatifs ; le jeu littéraire sur le temps de narration ; les répétitions syntaxiques avec syntagmes verbaux ; et la coupe flaubertienne proche de la figure de versification : le rejet.

- « , à le NN de le NN », « à le NN de le NN , »

Il s'agit essentiellement du patron prépositionnel « à le NN de le NN », mais coupé, soit par une virgule placée au début de la construction syntaxique, soit par une virgule

à la fin de la construction. La prise en charge des ponctuations dans les patrons nous permet de relever les phénomènes de coupe syntaxique, spécifiques du style de Flaubert.

-la période ternaire

Les constructions spécifiques des deux patrons avec virgule nous permettent de prendre en compte la relation sémantique et syntaxique de ces patrons au sein de la phrase. Nous remarquons qu'ils apparaissent souvent dans les périodes ternaires, une des structures syntaxiques préférées de Flaubert.

297. *Un homme passa, /qui marchait lourdement, /à la manière des bœufs.*
(Salammbô, Flaubert)

298. *Au bord du quai, / d'autres hennissaient, / effrayés par la mer.* (Trois contes, Flaubert)

299. *Tous les enfants des écoles, les chantres et les pompiers marchaient sur les trottoirs, / tandis qu'au milieu de la rue, / s'avançaient premièrement : le suisse armé de sa hallebarde, le bedeau avec une grande croix, l'instituteur surveillant les gamins, la religieuse inquiète de ses petites filles.* (Trois contes, Flaubert)

300. *Au bord de la falaise, / le vieux palmier, avec sa touffe de feuilles jaunes, / devient le torse d'une femme penchée sur l'abîme, et dont les grands cheveux se balançent.* (La Tentation de Saint-Antoine, Flaubert)

-La coupe syntaxique

La coupe flaubertienne, qui est souvent une coupe oratoire, est également présente dans nos résultats pour ces deux patrons. Les trois exemples suivants montrent les occurrences du patron « , à le NN de le NN ». Dans la langue écrite, la virgule antéposée ne relève pas d'une nécessité pour les trois phrases, mais Flaubert utilise cette ponctuation comme marque de pause à l'intention du lecteur.

301. *Ils tournaient dans les rues étroites, **au milieu des ténèbres**.* (Salammbô, Flaubert)

302. *Julien, avec sa chandelle, enflamma un paquet de fougères, **au milieu de la cabane**.* (Trois contes, Flaubert)

303. *Il me semble ... Antoine grimpe sur une roche, **à l'entrée du sentier**.* (La Tentation de Saint-Antoine, Flaubert)

Le dernier exemple montre une phrase ternaire dont le dernier syntagme est séparé du premier par un circonstanciel de lieu. La phrase, déjà très courte, est coupée en deux par le syntagme « tout en haut d'une tourelle ».

304. *Ils travaillaient ensemble, tout en haut d'une tourelle, **à l'écart du bruit**.* (Trois contes, Flaubert)

Notre analyse de ces deux patrons prépositionnels, en tenant en compte de la ponctuation, nous permet de relever des caractéristiques stylistiques et surtout rythmiques des phrases de Flaubert. Nous constatons que la coupe flaubertienne est très

particulière en raison de ses caractéristiques oratoires. Les patrons avec ponctuations sont de bons indicateurs de cette caractéristique.

- « , le NN de le NN »

Ce patron nominal est apparu 187 fois dans les romans de Flaubert, il est considéré comme suremployé d'après les résultats de l'AFC. Dans les analyses des quatre patrons précédents (« , VB de le NN », « , VB de NN », « , à le NN de le NN » et « à le NN de le NN , »), nous avons déjà présenté les caractéristiques rythmiques liées à la virgule dans les constructions syntaxiques, par conséquent, nous ne répétons pas les mêmes phénomènes rythmiques pour ce patron nominal, qui est également construit avec une virgule au début. Nous présentons uniquement les nouvelles fonctions stylistiques propres à ce patron.

-La nominalisation d'un adjectif

La nominalisation d'un adjectif chez Flaubert est souvent employée pour garder la symétrie syntaxique. Dans les trois exemples suivants, au lieu d'écrire « le NN ADJ », l'emploi du patron « , le NN de le NN » met en place le parallélisme avec les autres syntagmes à proximité dont le premier nom n'est pas le produit d'une nominalisation. Les adjectifs tels que « indifférent », « féroce » et « transparent » sont transformés en nom pour créer le rythme répétitif.

305. Mais l'air bestial des figures, le tapage des métiers, l'indifférence des propos glaçaient son cœur. (Trois contes, Flaubert)

306. *Leurs habits mouillés fumaient devant l'âtre ; et, quand ils étaient repus, ils racontaient leurs voyages : les erreurs des nef s sur la mer écumeuse, les marches à pied dans les sables brûlants, la férocité des païens, les cavernes de la Syrie, la Crèche et le Sépulcre.* (Trois contes, Flaubert)

307. ANTOINE : *Comme c'est bon, le parfum des palmiers, le frémissement des feuilles vertes, la transparence des sources !* (La Tentation de Saint-Antoine, Flaubert)

-Les répétitions ternaires

Le rythme ternaire est caractéristique du style de Flaubert. Ainsi, dans nos résultats, nous retrouvons ce phénomène avec le patron en question.

308. *On criait la dévastation des villes, / l'incendie des campagnes, / le massacre des soldats.* (Salammbô, Flaubert)

309. *L'odeur du sang, / le spectacle du carnage / et le vacarme des clairons avaient fini par lui faire bondir le cœur.* (Salammbô, Flaubert)

310. *Battez le pauvre qui envie la housse de l'âne, / le repas du chien, / le nid de l'oiseau, et qui se désole parce que les autres ne sont pas des misérables comme lui.* (La Tentation de Saint-Antoine, Flaubert)

Le dernier exemple montre une phrase avec répétition ternaire et deux autres figures de style. Nous retrouvons d'abord l'ellipse dans les trois occurrences du patron « , le NN de le NN » avec l'omission du verbe « provient ». Les trois occurrences sont aussi enchaînées par la reprise d'un mot du syntagme précédent. Il s'agit de la

concaténation, figure qui consiste à enchaîner les propositions d'une période en reprenant un mot de la proposition précédente³⁰⁶.

311. Comme l'existence provient de la corruption, / la corruption du désir, / le désir de la sensation, / la sensation du contact, j'ai fui toute action, tout contact ;
(La Tentation de Saint-Antoine, Flaubert)

L'analyse des patrons syntaxiques spécifiques de Flaubert nous permet de relever plusieurs caractéristiques du style de l'auteur. Rappelons que les patrons tels que « ADJ et ADJ », « de le NN de le NN » ne font pas objet de nos analyses parce qu'ils sont relativement sous-employés dans les textes de Flaubert.

Nous concluons que les patrons qui se construisent avec une virgule sont très spécifiques de Flaubert. La virgule permet d'exploiter le rôle grammatical et stylistique d'un certain patron dans les phrases. Le rythme préféré de Flaubert est le rythme ternaire, que ce soit pour la phrase en général ou pour les répétitions au sein de la phrase, cela confirme ce que disent les critiques littéraires sur le style de Flaubert. La coupe flaubertienne est souvent oratoire, les virgules servant à transcrire la pause oratoire comme si les phrases étaient destinées à être lues à haute voix. Les répétitions de Flaubert sont souvent ternaires, elles sont moins abondantes que celles de Victor Hugo, mais également moins discrètes que celles de Chateaubriand qui engagent souvent moins de trois syntagmes répétitifs. Les analyses nous permettent de confirmer la capacité des patrons syntaxiques à caractériser le style de Flaubert.

³⁰⁶ « Concaténation », <https://www.cnrtl.fr/definition/concatenation>, consulté le 1 juillet, 2021.

C) Stendhal

Les patrons identificatoires de Stendhal sont très marqués par la présence de patrons adjectivaux. Presque tous les patrons adjectivaux (et adverbiaux) sont suremployés dans les romans de Stendhal par rapport aux trois autres auteurs de notre corpus, tandis que les patrons nominaux, comme « le NN de le NN », sont tous sous-employés. Nous procédons à l'analyse des occurrences de ces patrons adjectivaux et adverbiaux dans cette section afin de trouver leurs fonctions stylistiques.

- « **ADV de NN** », « **un ADV ADJ** », « **un NN ADV ADJ** »

Ce sont trois des patrons suremployés dans les romans de Stendhal, dont la construction syntaxique comprend un adverbe. Ces patrons sont statistiquement spécifiques uniquement pour Stendhal, aucun d'entre eux n'est spécifique pour les autres auteurs de notre corpus. Nous relevons les caractéristiques stylistiques suivantes.

-La répétition

Pour le patron « ADV de NN », l'adverbe est majoritairement l'adverbe d'intensité. 84% des occurrences se présentent sous les deux formes suivantes : « tant de NN » et « beaucoup de NN ». Toutes les occurrences du patron « un ADV ADJ » se matérialisent sous la forme suivante : « un peu ADJ », il s'agit encore d'un adjectif d'intensité, « peu ». Les phénomènes stylistiques sont rares au sein de ces patrons, nous en avons remarqué un seul : la répétition, présente, mais peu fréquente dans nos résultats.

312. Il y avait **tant de malheur, tant d'assurance**, d'obéir ponctuellement dans ce mot, que madame de Chasteller en fut comme désarmée. (Lucien Leuwen, Stendhal)

313. Lucien, obligé de se taire, entendit **tant de sottises** et surtout vit **tant de sentiments** bas étalés avec orgueil, qu'il eut le sentiment d'être dans l'antichambre de son père. (Lucien Leuwen, Stendhal)

On trouve une seule occurrence de répétition pour le patron « un NN ADV ADJ », dont la reprise du patron est partielle :

314. Au lieu d'**une civilité fort agréable**, mais **fort prudente** au fond et **fort méticuleuse**, le ton de chacun de ces propos disait avec gaieté : « Je me moque de tout au monde, et je compte sur moi. (Lucien Leuwen, Stendhal)

-La comparaison et l'antithèse

Une seule occurrence de comparaison est liée au patron « un NN ADV ADJ » :

315. Mais dans la vie d'aucun homme on ne saurait trouver **une époque aussi heureuse** que le mois de septembre qu'il passa avec Mina sur le lac Majeur. (Mina de Wangel, Stendhal)

Et une seule occurrence d'antithèse est présente dans la construction du patron « un NN ADV ADJ ».

316. Ils furent réduits à répéter entre eux, d'un air **tristement satisfait** : « Après tout, ce n'est qu'un bourgeois, né on ne sait où, et qui ne peut jouir que de la

noblesse personnelle que lui confère son épauvette de sous-lieutenant. (Lucien Leuwen, Stendhal)

Nous concluons que les patrons spécifiques de Stendhal comprenant un adverbe ont une fonction stylistique limitée. Contrairement aux adjectifs dans les écrits de Victor Hugo, discutés dans la section 4.5.2 (Chapitre 4), les adverbes et les adjectifs dans les patrons spécifiques de Stendhal sont peu liés à d'autres phénomènes stylistiques. La seule conclusion à laquelle nous aboutissons, concernant ces trois patrons, est que Stendhal préfère utiliser un adverbe d'intensité pour désigner ces adjectifs, ce qui est prouvé statistiquement par l'AFC, mais cette préférence comprend peu de fonctions stylistiques. La faible présence de fonctions stylistiques est une des caractéristiques du style de Stendhal.

- « **le plus ADJ** », « **de le plus ADJ NN** », « **de le NN le plus ADJ** », « **le NN le plus ADJ** »

Les quatre patrons adjectivaux comprennent tous un adjectif au superlatif. D'après nos résultats de l'AFC, ils sont surreprésentés dans les romans de Stendhal. Nous avons remarqué un seul phénomène stylistique remarquable dans les occurrences de ces quatre patrons.

-La répétition du superlatif

Le superlatif, en lui-même, est un outil de mise en relief. Stendhal met l'accent sur les adjectifs dans ses phrases en faisant répéter les superlatifs. Au lieu de confirmer que le style de Stendhal est clair et naturel, nous trouvons plutôt que son style trahit une simplicité, surtout au niveau des techniques de mise en relief.

317. Mais pour s'amuser lui-même, tout en lui faisant **la cour la plus assidue**, il avait grand soin toutes les fois qu'il trouvait l'occasion de lui parler en particulier, de ne lui adresser que les mots **les plus indifférents** et les **propos les plus décolorés**. (Mina de Wangel, Stendhal)

318. La vulgarité **la plus audacieuse** et **la plus plate** faisait, au contraire, la physionomie de ses associés. (Lucien Leuwen, Stendhal)

319. Le docteur savait donner aux choses **les plus personnelles**, aux conseils intimes **les moins sollicités**, et qui eussent été **les plus impertinentes** chez tout autre, un tour si vif, si amusant, si peu offensant, tellement éloigné de l'apparence de vouloir prendre un ton de supériorité, qu'il fallait tout lui passer. (Lucien Leuwen, Stendhal)

Si Zola témoigne que Flaubert « poursuit les mots répétés jusqu'à trente ou quarante lignes de distance³⁰⁷ » (Zola, 1895, p. 506). Nous remarquons que cela n'est pas le cas pour Stendhal. Dans les trois phrases suivantes, l'adjectif au superlatif est déjà présent dans la même phrase, et la répétition ne crée pas d'effet esthétique remarquable.

320. Pendant que le baron Thérance faisait ce **triste** tableau de la ville de Nancy, le 27^e régiment de lanciers s'en approchait, parcourant la plaine **la plus triste** du monde ; (Lucien Leuwen, Stendhal)

321. Il faut convenir, se disait-il un soir en regagnant son logement et dormant presque debout dans la rue, il faut convenir que des gens cent fois plus **nobles** que

³⁰⁷ Zola, Émile. 1895. *Les romanciers naturalistes*. Paris, France : Bibliothèque-Charpentier, p. 506.

moi daignent m'adresser la parole avec les formes les plus nobles et les plus flatteuses, mais ils m'assomment, les cruels ! (Lucien Leuwen, Stendhal)

322. La seconde fois que Leuwen la vit après le bal, il en fut traité à peine comme une simple connaissance, même il lui sembla qu'elle ne répondait pas au peu de mots qu'il lui adressait autant que la politesse la plus simple aurait semblé l'exiger. (Lucien Leuwen, Stendhal)

Ces sept patrons sont les patrons spécifiques suremployés dans les romans de Stendhal. Nous remarquons que les fonctions stylistiques de ces patrons sont moins claires que pour les autres auteurs de notre corpus : la faible présence de figures de style étant une des caractéristiques de Stendhal. Nous constatons également la différence entre Stendhal et les autres auteurs : les répétitions peu stylistiques ne sont pas soigneusement éliminées ; elles consistent souvent à répéter une même technique de mise en relief (le superlatif, l'adverbe d'intensité). Ainsi, la différence entre Stendhal et Flaubert est évidente, et la densité relativement moins importante de figures stylistiques écarte Stendhal de Chateaubriand et Hugo.

5.6 Conclusion

Nous avons démontré dans ce chapitre la capacité des patrons syntaxiques à caractériser les styles littéraires de Chateaubriand, Flaubert, Hugo et Stendhal. Les différences stylistiques entre les quatre auteurs sont visibles non seulement dans les résultats statistiques des patrons syntaxiques, mais aussi dans les phénomènes stylistiques liés à ces patrons. Pour accomplir la tâche d'identification et de caractérisation, les analyses quantitatives et qualitatives doivent être combinées et l'ensemble des patrons spécifiques doivent être pris en compte.

La section 4.3 montre que les phrases en prose de Chateaubriand et de Stendhal sont syntaxiquement complexes. Si ce phénomène s'explique pour le premier qui cherche à obtenir un style « voluptueux³⁰⁸ » (Barrès, 1931, t. III, p. 189), le résultat pour Stendhal contredit sa recherche d'un style clair et naturel³⁰⁹ (Talbot, 1985). Les phrases de Stendhal trahissent une complexité qui se manifeste par l'existence de phrases incluant souvent plusieurs propositions en situation de coordination, de juxtaposition et de subordination. Ces caractéristiques sont mesurables grâce à notre méthode qui comprend les trois paramètres suivants : la « largeur » de l'arbre syntaxique de dépendance, la « profondeur » et le « LPIndice ».

Les résultats statistiques du repérage des patrons identificatoires et caractérisants montrent que Chateaubriand est l'auteur qui emploie le plus de figures de style dans ses écrits, tandis que Stendhal présente une densité nettement inférieure en termes de présence de figures de style dans ses romans. La ressemblance, révélée par les patrons syntaxiques, entre Flaubert et Victor Hugo est également visible sur l'AFC.

L'étude qualitative des patrons caractérisants, mais aussi des patrons statistiquement spécifiques, montre, d'abord, que les phénomènes d'archaïsme et d'innovation sont bien présents chez Chateaubriand : l'auteur, comme Victor Hugo,

³⁰⁸ Barrès, Maurice. 1931. *Op. cit.*, p. 189.

³⁰⁹ Talbot, Emile J. 1985. *Op. cit.*.

peut être considéré comme représentatif du romantisme, qui relie la tradition littéraire à l'innovation langagière. Mais Chateaubriand se distingue de Hugo par ses répétitions souvent discrètes et ses métaphores moins audacieuses. Ensuite, la préférence de Flaubert pour le rythme ternaire est très marquée dans nos résultats. De plus, la coupe très particulière, qui suit souvent le rythme oratoire, rend son style harmonieux et bien équilibré. À l'exception de la coupe flaubertienne, le style de Flaubert est proche de celui de Hugo par l'existence de répétitions abondantes et par la combinaison habituelle de plusieurs figures de style. Le style de Stendhal, enfin, se définit à travers son opposition avec les autres auteurs. Les figures de style sont relativement rares dans ses écrits, la coupe syntaxique est moins équilibrée que celle de Flaubert, ses techniques d'insistance et d'analogie sont relativement moins variées que celles de Chateaubriand et Hugo.

Les manifestations formelles des styles littéraires ne se limitent pas au niveau syntaxique, mais notre étude montre que les patrons d'ordre syntaxique sont capables de relever ces phénomènes stylistiques formels. S'il reste difficile de procéder à des conclusions générales à partir des patrons caractérisants, en raison de leurs fonctions stylistiques limitées et leurs fréquences relativement faibles, les patrons identificatoires les complètent utilement, avec leur capacité à différencier les styles des quatre auteurs. Nous insistons enfin sur le fait que les patrons par eux-mêmes, en dehors du contexte, sont souvent difficiles à interpréter, mais l'étude sur des occurrences dans le texte original permet en général de révéler leurs caractéristiques stylistiques.

Chapitre 6 Conclusion

6.1 Introduction

Nous avons proposé dans notre thèse une méthodologie pour répondre à la problématique suivante : comment l'analyse syntaxique automatique peut-elle être utile pour identifier et caractériser les styles littéraires ? Nous avons implémenté et appliqué des méthodes d'extraction supervisée et non-supervisée de patrons syntaxiques pour caractériser le style de Chateaubriand, Flaubert, Hugo et Stendhal. Nous avons ainsi montré que cette méthode est robuste pour caractériser les styles littéraires.

Dans ce chapitre de conclusion, nous résumons les résultats des analyses quantitatives et qualitatives à partir des patrons syntaxiques extraits automatiquement par notre méthode. Nous insistons sur le fait que les patrons syntaxiques peuvent être considérés comme un marqueur stylistique performant en accomplissant à la fois la tâche d'identification et de caractérisation des styles littéraires. De plus, les études comparatives sur le changement de style (entre les premiers et les derniers romans de Hugo, ainsi que sur l'analyse du contraste entre les styles romanesques des quatre auteurs de notre corpus) montrent la nécessité de combiner les analyses statistique et littéraire des patrons syntaxiques pour obtenir des résultats fiables et interprétables.

6.2 Les résultats d'analyse

6.2.1 La complexité syntaxique

La méthode pour mesurer la complexité syntaxique que nous avons présentée au Chapitre 3 de la thèse est essentiellement une méthode qui mesure la complexité des phrases, à partir de l'analyse des relations de dépendance. Elle se base sur l'analyse

syntaxique en dépendances automatique fournie par UDPipe. Les paramètres utilisés sont l'expansion des relations syntaxiques horizontales et verticales au sein de l'arbre syntaxique de dépendance : la « largeur » de l'arbre (son expansion horizontale) rend compte de la présence de propositions en situation de coordination et de juxtaposition, alors que la « profondeur » (son expansion verticale) rend compte de la présence de propositions en situation de subordination.

Les résultats sur le corpus Victor Hugo montrent que la complexité syntaxique se réduit progressivement et chronologiquement au sein des neuf romans de Victor Hugo. Les derniers romans de Hugo, ceux publiés après l'exil de l'auteur, contiennent souvent des phrases plus simples que ses premiers romans. Concernant les phrases dans le corpus des quatre auteurs du 19^e siècle considérés, les résultats statistiques montrent que Chateaubriand et Stendhal sont deux auteurs qui privilégient les phrases complexes, tandis que Hugo et Flaubert préfèrent les phrases relativement simples. Les résultats confirment, par exemple, les études littéraires sur le style de Chateaubriand, en fournissant les preuves concrètes de sa tendance à rallonger les apodoses de ses phrases, ce dont rend compte la complexité syntaxique. Notons aussi, à l'inverse, que nos résultats contredisent, d'une certaine manière, ce qui est souvent mis en avant concernant le style clair et naturel de Stendhal : ses phrases sont en fait globalement complexes et travaillées, contrairement à ce à quoi on pouvait s'attendre.

Si la complexité syntaxique est une des manifestations formelles permettant de caractériser grossièrement le style d'auteur, elle ne révèle que partiellement les caractéristiques stylistiques des phrases. Il s'agit d'une analyse macroscopique, qui donne une vue assez générale : ce paramètre purement statistique doit toujours être combiné avec d'autres paramètres, et notamment une étude minutieuse au niveau du texte même, pour comprendre la syntaxe d'un certain auteur.

6.2.2 Les patrons caractérisants.

Les patrons caractérisants sont des structures syntaxiques qui possèdent déjà une fonction stylistique précise. Nous avons choisi trois des figures de style de nature syntaxique (l'inversion du sujet, le clivage et la répétition), extraites de manière automatique et supervisée dans les deux corpus.

L'analyse quantitative révèle des préférences propres à chaque auteur dans l'emploi de ces trois figures de style. Les résultats statistiques montrent en effet que :

1. L'utilisation de certaines figures de style en contexte est propre à un auteur particulier : Chateaubriand et Hugo sont les seuls à utiliser les inversions concessives dans leurs romans, les deux auteurs ont aussi une préférence marquée pour l'utilisation conjointe de l'inversion du sujet avec le clivage, une autre figure d'insistance. Stendhal est le seul auteur à utiliser l'inversion du sujet pour exprimer la condition, il est également l'auteur qui emploie l'inversion complexe (avec la reprise du sujet par un pronom personnel, placé derrière le verbe) qui ne relève pas d'une nécessité grammaticale.
2. Les romans de Chateaubriand présentent une densité supérieure dans l'emploi de ces trois figures de style, alors que Stendhal est l'auteur qui emploie le moins de figures de style dans ses écrits.
3. Pour le corpus Victor Hugo, nous constatons la ressemblance entre *Les Misérables* et *Notre-Dame de Paris* en matière de fréquences et de typologie concernant ces trois figures de style. Plus globalement, la fréquence des figures de style augmente dans les derniers romans de Hugo : les romans tels que *Quatrevingt-treize* et *Les Travailleurs de la mer* présentent une densité importante de figures de style alors que *Bug-Jargal*, le premier roman de Hugo, en contient relativement moins par exemple.

L'analyse qualitative des occurrences de ces trois figures est fructueuse :

1. La combinaison de multiples figures de style est courante dans les romans de Chateaubriand et de Hugo. Plusieurs figures d'insistance et d'analogie telles que l'image, le clivage et la répétition sont souvent présentes dans une seule phrase.
2. Les répétitions hugoliennes sont très caractéristiques par leur abondance. Ce phénomène a été également constaté dans les romans de Flaubert, où l'auteur emploie souvent plus de trois occurrences du même patron syntaxique. En revanche, les répétitions de Chateaubriand sont « discrètes³¹⁰ » (Brunot et al., p. 324) : elles n'engagent majoritairement que deux ou trois patrons répétitifs.
3. Nous constatons également que l'étude des phrases qui comprennent au moins une des trois figures de style est capable de révéler le rythme équilibré des phrases de Flaubert, ainsi que le jeu sur les sonorités, très typique, qui les rapproche de l'art oratoire.
4. La faible présence de figures de style dans les romans de Stendhal est souvent accompagnée d'une certaine simplicité des techniques stylistiques utilisées. Malgré la complexité syntaxique élevée que nous avons constatée chez Stendhal, nos analyses qualitatives montrent que les techniques stylistiques sont relativement peu variées et simples sous la plume de l'auteur. Les inversions se combinent rarement avec d'autres figures de style, la typologie des clivages est moins variée et les répétitions se limitent majoritairement aux patrons adjectivaux.

L'analyse des figures de style permet de révéler les particularités de chaque auteur dans leur emploi de ces outils stylistiques particuliers. Il nous faut toutefois admettre que la prise en charge de trois figures de style seulement (les trois figures d'ordre syntaxiques qui sont repérables de manière complètement ou largement automatique) ne permet pas une étude exhaustive des patrons caractérisants. Néanmoins, notre étude

³¹⁰ Brunot, Ferdinand, et Charles Bruneau. 1948. *Op. cit.*, p. 324.

permet déjà de mettre en avant deux points cruciaux pour toute démarche stylistique : premièrement, les patrons caractérisants ne révèlent que partiellement les caractéristiques stylistiques d'un texte en raison de leurs fréquences peu élevées, qui rendent le processus de généralisation difficile ; deuxièmement, l'analyse des patrons caractérisants, malgré leurs fonctions stylistiques indéniables, doit toujours tenir compte du contexte. Les phénomènes stylistiques que nous avons présentés dans l'analyse qualitative ne sont détectables que par l'analyse et l'interprétation qui en est faite par le lecteur (ou le critique), par le truchement du retour au texte, c'est-à-dire par l'étude des figures de style en contexte.

6.2.3 Les patrons identificatoires

La notion de patron identificatoire est complémentaire de celle de patron caractérisant et permet de pallier le manque de généralité de ceux-ci. Les patrons identificatoires sont extraits d'une manière non-supervisée et automatique. La thèse propose d'étudier les patrons syntaxiques en tenant compte des relations de dépendance, mises en évidence par l'analyseur syntaxique UDPipe. La méthode d'extraction que nous avons élaborée réduit considérablement le nombre de patrons non-interprétables par rapport aux méthodes traditionnelles à base de n-grams, discutées au Chapitre 3. La spécificité d'un certain patron dans un corpus par rapport à un autre est calculée sur une base statistique : nous avons recours à l'Analyse factorielle des correspondances (AFC), basée sur le test du χ^2 .

L'analyse quantitative des patrons identificatoires révèle un certain nombre de particularités. Par exemple, presque tous les patrons spécifiques des romans de Chateaubriand sont des patrons nominaux, alors que Stendhal suremploie essentiellement des patrons adjectivaux et adverbiaux ; les patrons spécifiques de Flaubert possèdent souvent une virgule au début ou à la fin de la construction, marquant

une coupe phrastique et potentiellement oratoire ; Victor Hugo suremploie un grand éventail de patrons issus de catégories grammaticales différentes, ce qui confirme son style « hétérogène » tel que le qualifient souvent les stylisticiens.

L'étude des occurrences de ces patrons identificatoires dans leur contexte donne accès aux caractéristiques stylistiques propres à chaque auteur :

1. Chateaubriand est un auteur qui emploie un grand nombre d'outils stylistiques différents dans ses écrits. L'analyse qualitative des patrons identificatoires est conforme à ce qui avait déjà été constaté pour l'analyse des patrons caractérisants de cet auteur, à savoir une grande densité dans l'emploi des figures de style dans ses romans. Les résultats confirment aussi les caractéristiques stylistiques de ses textes, qui relient souvent tradition littéraire et innovation langagière, ce qui est représentatif d'après les stylisticiens du style romantique de cette époque³¹¹ (Sainte-Beuve, 1926). Les figures de style telles que l'image et l'antithèse sont omniprésentes dans ses textes. Chateaubriand se différencie malgré tout de Hugo par ses répétitions discrètes et ses métaphores moins surprenantes.
2. L'étude des patrons identificatoires de Flaubert met en avant sa coupe oratoire typique : ses phrases sont soigneusement coupées, le rythme y est équilibré, les propositions dans les phrases longues sont minutieusement agencées par les pronoms relatifs, les conjonctions et les adverbes. L'effet de fluidité est souvent constaté dans nos résultats. Les figures de style préférées de Flaubert, telles que l'image, l'ellipse et le parallélisme sont aussi faciles à détecter par les patrons syntaxiques. « L'harmonie³¹² » (Thibaudet, 1935) qui existe dans ses phrases d'après les stylisticiens est confirmée par notre étude.
3. Le style « hétérogène » de Victor Hugo se manifeste également dans notre analyse des patrons identificatoires : l'antithèse est fréquente, la métaphore est

³¹¹ Sainte-Beuve. 1926. *Op. cit.*

³¹² Thibaudet, Albert. 1935. *Op. cit.*

surprenante, reliant souvent des mots issus de domaines sémantiques différents, la répétition est abondante. Le style de Hugo se différencie de celui des autres par ces trois caractéristiques, évoquées fréquemment dans les études littéraires sur son style. Notre étude présentée dans le Chapitre 4 montre aussi que ces caractéristiques sont plus présentes dans les derniers romans, ceux écrits après l'exil de l'auteur.

4. L'analyse des patrons spécifiques de Stendhal révèle la faible présence de figures de style dans ses écrits : elle confirme la détestation de Stendhal du style « voluptueux³¹³ » (Barrès, 1931, t. III, p. 189) de ses contemporains, orné d'un grand nombre de figures d'insistance et d'analogie. Pourtant, certains résultats d'analyse semblent contredire la recherche par l'auteur d'un style clair et naturel. Nous constatons que ses phrases sont souvent longues, avec peu de marqueurs de pause, les coupes ne sont pas soigneusement insérées dans la phrase pour créer un rythme équilibré, les répétitions lexicales qui ne relèvent pas d'une nécessité stylistique ne sont pas supprimées, et enfin, les techniques stylistiques (que ce soit pour exprimer l'insistance ou pour exprimer le superlatif) ne sont pas variées.

Notre étude des patrons identificatoires extraits des romans des quatre auteurs de notre corpus a permis de relever avec succès leurs caractéristiques stylistiques respectives. Elle a également permis de repérer des phénomènes stylistiques qui ont été peu discutés par les critiques littéraires jusqu'ici : par exemple, les jeux sur les sons chez Victor Hugo, la surreprésentation des patrons adjectivaux dans les écrits de Stendhal, les répétitions abondantes de Flaubert et la nominalisation des adjectifs dans les romans de Chateaubriand. Notre démarche partant de l'étude des patrons syntaxiques a permis de confirmer et parfois d'infirmer certaines caractéristiques stylistiques mises en avant dans la littérature à propos de ces quatre auteurs ; notre

³¹³ Barrès, Maurice. 1931. *Op. cit.*, p. 189.

démarche a également permis de mettre au jour de nouvelles caractéristiques stylistiques, qui pourront à leur tour alimenter des études à venir.

Néanmoins, la notion de patrons syntaxiques présente également des limites, qu'il faut évoquer. En particulier, certains phénomènes stylistiques, bien que présents au niveau des phrases, ne peuvent pas être repérés directement ou uniquement par les patrons syntaxiques. Par exemple, la tendance de Chateaubriand à rallonger les apodoses (c'est-à-dire les phrases dont la principale est en position finale) est constatée dans nos résultats, mais ce phénomène n'est pas le produit d'un patron syntaxique particulier ; c'est le travail d'observation et d'interprétation à partir des patrons extraits qui permet de l'obtenir. Nous pouvons donc dire, d'après notre analyse des patrons identificatoires, que les patrons sont souvent difficiles à interpréter hors contexte : l'interprétation dépend fortement de l'analyse du reste de la phrase ; il est aussi difficile d'utiliser les patrons syntaxiques pour aller vers une analyse plus globale au niveau des phrases. Si le rythme ternaire de Flaubert et le rallongement excessif de l'apodose sont présents dans nos résultats, ces phénomènes ne sont pas directement visibles en inspectant uniquement les patrons extraits.

La combinaison d'analyses statistique et littéraire des patrons identificatoires donne des résultats satisfaisants pour caractériser les styles littéraires, en fournissant à la fois la trace des préférences de chaque auteur pour des patrons syntaxiques particuliers, et la possibilité d'interpréter les phénomènes stylistiques liés à ces patrons.

6.3 Perspectives et suggestions pour de futures recherches

6.3.1 Pour les patrons caractérisants

La thèse a proposé une méthode d'extraction supervisée pour trois figures de style d'ordre syntaxique. Cette méthode, ainsi que l'analyse de ces trois figures, pourraient être poussés plus avant. La reconnaissance d'autres figures de style, impliquant par

exemple une analyse sémantique fine, serait évidemment un plus pour l'étude des patrons caractérisants dans les textes. Par exemple, l'annotation sémantique pourrait permettre de relever l'oxymore, une figure de style qui vise à rapprocher deux termes de sens éloigné, dans une formule en apparence contradictoire³¹⁴. Les métaphores et les antithèses, que nous avons relevées manuellement dans l'analyse qualitative, pourraient aussi bénéficier d'une annotation sémantique particulière.

Par conséquent, nous pensons que l'approfondissement et l'enrichissement des méthodes d'annotations, en combinant avec relations syntaxiques (en dépendance ou en constituants) et sémantiques, est une bonne stratégie pour extraire avec précision plus de patrons caractérisants dans le texte et donc approfondir l'étude que nous avons présentée.

6.3.2 Pour les patrons identificatoires

Notre méthode d'extraction non-supervisée de patrons syntaxiques ne demande pas une prédéfinition du nombre de composants dans un patron, mais elle exige une prédéfinition de la profondeur d'extraction. Nous croyons qu'il serait possible d'éliminer ce paramètre en extrayant tous les patrons contigus dans un arbre syntaxique. La stratégie consisterait alors à extraire tous les patrons complets (c'est-à-dire correspondant à des séquences linguistiquement complètes et valides) sans critère de profondeur. Enfin, d'autres stratégies de filtrage statistique pourraient être appliquées aux résultats pour calculer la spécificité des patrons dont la fréquence est faible, mais néanmoins statistiquement significative, à l'instar de la méthode utilisée par A. Salem pour extraire ce qu'il appelle les « segments répétés³¹⁵ » (Lebart et al. 1994).

Nous avons évoqué dans Chapitre 4 que les résultats *Du Dernier jour d'un*

³¹⁴ <https://fr.wikipedia.org/wiki/Oxymore>, consulté le 1 août, 2021.

³¹⁵ Lebart, Ludovic, André Salem, et Christian Préfacier Baudelot. 1994. *Statistique textuelle*. Paris, France: Dunod.

condamné de Victor Hugo sont exceptionnels, parce que cette œuvre se situe entre le genre romanesque et théâtral. Ainsi, nous croyons que ce serait intéressant d'appliquer notre méthode sur les œuvres de genres littéraires différents pour mettre en évidence la capacité de la syntaxe à différencier et à caractériser les genres littéraires.

6.4 Conclusion

La thèse exploite le rôle de la syntaxe pour différencier et caractériser les styles littéraires en français. L'application de la méthode sur les romans de Victor Hugo et sur les romans des quatre auteurs français du 19^{ème} siècle démontre la capacité des patrons syntaxiques (identificatoires et caractérisants) à identifier des phénomènes stylistiquement significatifs dans les textes. Il est également important d'insister sur le rôle primordial du contexte et de l'interprétation pour ce type d'étude.

En conclusion, nous pouvons affirmer que les patrons syntaxiques sont bien un marqueur stylistique performant, d'une part grâce aux fréquences observées, qui permettent de caractériser à un niveau macroscopique le style des auteurs considérés ; d'autre part grâce à l'étude qualitative des patrons en contexte, qui permet de fournir une interprétation des fonctions stylistiques dans le texte. Nous insistons sur le fait que l'étude stylistique ne doit pas se limiter à un seul patron ou à une seule catégorie de patrons, et que la stylistique de corpus gagne à adopter un point de vue combinatoire, mêlant analyse quantitative et qualitative. Ainsi, les patrons syntaxiques accomplissent-ils avec succès la tâche d'identification et de caractérisation de la stylométrie outillée.

Annexe 1 Deux échantillons de phrases pour tester la performance des analyseurs syntaxiques

- **50 phrases extraites du roman *Les Travailleurs de la mer***

sent_id = 1

text = Pourquoi Gilliatt se donnait-il toute cette peine ?

sent_id = 2

text = On verra.

sent_id = 3

text = Il dut refaire plusieurs fois le tranchant de sa hache et les dents de sa scie.

sent_id = 4

text = Il s'était, pour la scie, fabriqué un tiers-point.

sent_id = 5

text = Il se servait dans l'occasion du cabestan de la Durande.

sent_id = 6

text = Le crochet de la chaîne cassa.

sent_id = 7

text = Gilliatt en reforga un autre.

sent_id = 8

text = A l'aide de sa pince et de sa tenaille, et en se servant de son ciseau comme d'un tournevis, il entreprit de démonter les deux roues du navire ;

sent_id = 9

text = il y parvint.

sent_id = 10

text = On n'a pas oublié que ce démontage était exécutable ;

sent_id = 11

text = c'était une particularité de la construction de ces roues.

sent_id = 12

text = Les tambours qui les avaient couvertes, les emballèrent ;

sent_id = 13

text = avec les planches des tambours, Gilliatt charpenta et menuisa deux caisses où il déposa, pièce à pièce, les deux roues soigneusement numérotées.

sent_id = 14

text = Son morceau de craie lui fut précieux pour ce numérotage.

sent_id = 15

text = Il rangea ces deux caisses sur la partie la plus solide du pont de la Durande.

sent_id = 16

text = Ces préliminaires terminés, Gilliatt se trouva face à face avec la difficulté suprême.

sent_id = 17

text = La question de la machine se posa.

sent_id = 18

text = Démonter les roues avait été possible ; démonter la machine, non.

sent_id = 19

text = D'abord Gilliatt connaissait mal ce mécanisme.

sent_id = 20

text = Il pouvait, en allant au hasard, lui faire quelque blessure irréparable.

sent_id = 21

text = Ensuite, même pour essayer de le défaire morceau à morceau, s'il eût eu cette imprudence, il fallait d'autres outils que ceux qu'on peut fabriquer avec une caverne pour forge, un vent coulis pour soufflet, et un caillou pour enclume.

sent_id = 22

text = En tentant de démonter la machine, on risquait de la dépecer.

sent_id = 23

text = Ici on pouvait se croire tout à fait en présence de l'impraticable.

sent_id = 24

text = Il semblait que Gilliatt fût au pied de ce mur : l'impossible.

sent_id = 25

text = Que faire ?

sent_id = 26

text = Gilliat avait son idée.

sent_id = 27

text = Depuis ce maçon charpentier de Salbris qui, au seizième siècle, dans le bas âge de la science, bien avant qu'Amontons eût trouvé la première loi du frottement, Lahire la seconde et Coulomb la troisième, sans conseil, sans guide, sans autre aide qu'un enfant, son fils, avec un outillage informe, résolut en bloc, dans la descente du « gros horloge » de l'église de la Charité-sur-Loire, cinq ou six problèmes de statique et de dynamique mêlés ensemble ainsi que des roues dans un embarras de charrettes et faisant obstacle à la fois, depuis ce manoeuvre extravagant et superbe qui trouva moyen, sans casser un fil de laiton et sans décliqeter un engrenage, de faire glisser tout d'une pièce, par une simplification prodigieuse, du second étage du clocher au premier étage, cette massive cage des heures, toute en fer et en cuivre, « grande comme la chambre du guetteur de nuit », avec son mouvement, ses cylindres, ses barillets, ses tambours, ses crochets et ses pesons, son orbe de canon et son orbe de chaussée, son balancier horizontal, ses ancras d'échappement, ses écheveaux de chaînes et de chaînettes, ses poids de pierre dont un pesait cinq cents livres, ses sonneries, ses carillons, ses jacquemarts ;

sent_id = 28

text = depuis cet homme qui fit ce miracle, et dont on ne sait plus le nom, jamais rien de pareil à ce que méditait Gilliat n'avait été entrepris.

sent_id = 29

text = L'opération que rêvait Gilliat était pire peut-être, c'est-à-dire plus belle encore.

sent_id = 30

text = Le poids, la délicatesse, l'enchevêtrement des difficultés, n'étaient pas moindres de la machine de la Durande que de l'horloge de la Charité-sur-Loire.

sent_id = 31

text = Le charpentier gothique avait un aide, son fils ;

sent_id = 32

text = Gilliatt était seul.

sent_id = 33

text = Une population était là, venue de Meung-sur-Loire, de Nevers, et même d'Orléans, pouvant, au besoin, assister le maçon de Salbris, et l'encourageant de son brouhaha bienveillant ;

sent_id = 34

text = Gilliatt n'avait autour de lui d'autre rumeur que le vent et d'autre foule que les flots.

sent_id = 35

text = Rien n'égale la timidité de l'ignorance, si ce n'est sa témérité.

sent_id = 36

text = Quand l'ignorance se met à oser, c'est qu'elle a en elle une boussole.

sent_id = 37

text = Cette boussole, c'est l'intuition du vrai, plus claire parfois dans un esprit simple que dans un esprit compliqué.

sent_id = 38

text = Ignorer invite à essayer.

sent_id = 39

text = L'ignorance est une rêverie, et la rêverie curieuse est une force.

sent_id = 40

text = Savoir déconcerte parfois et déconseille souvent.

sent_id = 41

text = Gama, savant, eût reculé devant le cap des Tempêtes.

sent_id = 42

text = Si Christophe Colomb eût été bon cosmographe, il n'eût point découvert l'Amérique.

sent_id = 43

text = Le second qui monta sur le mont Blanc fut un savant, Saussure ;

sent_id = 44

text = le premier fut un pâtre, Balmat.

sent_id = 45

text = Ces cas, disons-le en passant, sont l'exception, et tout ceci n'ôte rien à la science, qui reste la règle.

sent_id = 46

text = L'ignorant peut trouver, le savant seul invente.

sent_id = 47

text = La panse était toujours à l'ancre dans la crique de l'Homme, où la mer la laissait tranquille.

sent_id = 48

text = Gilliatt, on s'en souvient, avait tout arrangé de façon à se maintenir en libre pratique avec sa barque.

sent_id = 49

text = Il y alla, et en mesura soigneusement le bau à plusieurs endroits, particulièrement le maître-couple.

sent_id = 50

text = Puis il revint à la Durande, et mesura le grand diamètre du parquet de la machine.

● **50 phrases extraites du roman *Madame Bovary***

sent_id = 1

text = Le coude sur la longue planche où elle repassait, il considérait avidement toutes ces affaires de femmes étalées autour de lui : les jupons de basin, les fichus, les collerettes, et les pantalons à coulisse, vastes de hanches et qui se rétrécissaient par le bas.

sent_id = 2

text = — À quoi cela sert-il ?

sent_id = 3

text = demandait le jeune garçon en passant sa main sur la crinoline ou les agrafes.

sent_id = 4

text = — Tu n'as donc jamais rien vu ? répondait en riant Félicité ;

sent_id = 5

text = comme si ta patronne, Mme Homais, n'en portait pas de pareils.

sent_id = 6

text = — Ah bien oui !

sent_id = 7

text = Mme Homais !

sent_id = 8

text = Et il ajoutait d'un ton méditatif : — Est-ce que c'est une dame comme Madame ?

sent_id = 9

text =, Mais Félicité s'impatientait de le voir tourner ainsi tout autour d'elle.

sent_id = 10

text = Elle avait six ans de plus, et Théodore, le domestique de M. Guillaumin, commençait à lui faire la cour.

sent_id = 11

text = — Laisse-moi tranquille ! disait-elle en déplaçant son pot d'empois.

sent_id = 12

text = Va-t'en plutôt piler des amandes ;

sent_id = 13

text = tu es toujours à fourrager du côté des femmes ;

sent_id = 14

text = attends pour te mêler de ça, méchant mioche, que tu aies de la barbe au menton.

sent_id = 15

text = — Allons, ne vous fâchez pas, je m'en vais vous faire ses bottines.

sent_id = 16

text = Et aussitôt, il atteignait sur le chambranle les chaussures d'Emma, tout empâtées de crotte — la crotte des rendez-vous — qui se détachait en poudre sous ses doigts, et qu'il regardait monter doucement dans un rayon de soleil.

sent_id = 17

text = — Comme tu as peur de les abîmer !

sent_id = 18

text = disait la cuisinière, qui n'y mettait pas tant de façons quand elle les nettoyait elle-même, parce que Madame, dès que l'étoffe n'était plus fraîche, les lui abandonnait.

sent_id = 19

text = Emma en avait une quantité dans son armoire, et qu'elle gaspillait à mesure, sans que jamais Charles se permît la moindre observation.

sent_id = 20

text = C'est ainsi qu'il déboursa trois cents francs pour une jambe de bois dont elle jugea convenable de faire cadeau à Hippolyte.

sent_id = 21

text = Le pilon en était garni de liège, et il y avait des articulations à ressort, une mécanique compliquée recouverte d'un pantalon noir, que terminait une botte vernie.

sent_id = 22

text =, Mais Hippolyte, n'osant à tous les jours se servir d'une si belle jambe, supplia Mme Bovary de lui en procurer une autre plus commode.

sent_id = 23

text = Le médecin, bien entendu, fit encore les frais de cette acquisition.

sent_id = 24

text = Donc, le garçon d'écurie peu à peu recommença son métier.

sent_id = 25

text = On le voyait comme autrefois parcourir le village, et quand Charles entendait de loin, sur les pavés, le bruit sec de son bâton, il prenait bien vite une autre route.

sent_id = 26

text = C'était M. Lheureux, le marchand, qui s'était chargé de la commande ;

sent_id = 27

text = cela lui fournit l'occasion de fréquenter Emma.

sent_id = 28

text = Il causait avec elle des nouveaux déballages de Paris, de mille curiosités féminines, se montrait fort complaisant, et jamais ne réclamait d'argent.

sent_id = 29

text = Emma s'abandonnait à cette facilité de satisfaire tous ses caprices.

sent_id = 30

text = Ainsi, elle voulut avoir, pour la donner à Rodolphe, une fort belle cravache qui se trouvait à Rouen dans un magasin de parapluies.

sent_id = 31

text = M. Lheureux, la semaine d'après, la lui posa sur sa table.

sent_id = 32

text =, Mais le lendemain il se présenta chez elle avec une facture de deux cent soixante et dix francs, sans compter les centimes.

sent_id = 33

text = Emma fut très embarrassée : tous les tiroirs du secrétaire étaient vides ;

sent_id = 34

text = on devait plus de quinze jours à Lestibouois, deux trimestres à la servante, quantité d'autres choses encore, et Bovary attendait impatiemment l'envoi de M. Derozerays, qui avait coutume, chaque année, de le payer vers la Saint-Pierre.

sent_id = 35

text = Elle réussit d'abord à éconduire Lheureux ; enfin il perdit patience ;

sent_id = 36

text = on le poursuivait, ses capitaux étaient absents, et, s'il ne rentrait dans quelques-uns, il serait forcé de lui reprendre toutes les marchandises qu'elle avait.

sent_id = 37

text = Il s'était tant de fois entendu dire ces choses, qu'elles n'avaient pour lui rien d'original.

sent_id = 38

text = Emma ressemblait à toutes les maîtresses ;

sent_id = 39

text = et le charme de la nouveauté, peu à peu tombant comme un vêtement, laissait voir à nu l'éternelle monotonie de la passion, qui a toujours les mêmes formes et le même langage.

sent_id = 40

text = Il ne distinguait pas, cet homme si plein de pratique, la dissemblance des sentiments sous la parité des expressions.

sent_id = 41

text = Parce que des lèvres libertines ou vénales lui avaient murmuré des phrases pareilles, il ne croyait que faiblement à la candeur de celles-là ;

sent_id = 42

text = on en devait rabattre, pensait-il, les discours exagérés cachant les affections médiocres ;

sent_id = 43

text = comme si la plénitude de l'âme ne débordait pas quelquefois par les métaphores les plus vides, puisque personne, jamais, ne peut donner l'exacte mesure de ses besoins, ni de ses conceptions, ni de ses douleurs, et que la parole humaine est comme un chaudron fêlé où nous battons des mélodies à faire danser les ours, quand on voudrait attendrir les étoiles.

sent_id = 44

text =, Mais, avec cette supériorité de critique appartenant à celui qui, dans n'importe quel engagement, se tient en arrière, Rodolphe aperçut en cet amour d'autres jouissances à exploiter.

sent_id = 45

text = Il jugea toute pudeur incommode.

sent_id = 46

text = Il la traita sans façon.

sent_id = 47

text = Il en fit quelque chose de souple et de corrompu.

sent_id = 48

text = C'était une sorte d'attachement idiot plein d'admiration pour lui, de voluptés pour elle, une béatitude qui l'engourdissait ;

sent_id = 49

text = et son âme s'enfonçait en cette ivresse et s'y noyait, ratatinée, comme le duc de Clarence dans son tonneau de malvoisie.

sent_id = 50

text = Par l'effet seul de ses habitudes amoureuses, Mme Bovary changea d'allures.

Annexe 2 Liste des verbes courants dans la construction d'incise

être	relever	remarquer	appeler	récrier	agréer
poursuivre	glisser	rectifier	pardoner	marteler	confirmer
confier	finir	recommander	mentir	époumoner	maintenir
terminer	remarquer	reconnaître	ordonner	rugit	extasier
enchaîner	rectifier	imaginer	avertir	tonner	recommander
conter	reconnaître	rebondir	accuser	beugler	assurer
répliquer	rassurer	râler	fulminer	égosiller	accorder
répondre	prier	protester	invectiver	clamer	trancher
lancer	prévenir	proposer	grommeler	proclamer	tempérer
dit	faire	prévenir	menacer	trancher	solliciter
déclarer	exposer	prétendre	aboyer	crier	obstiner
faire	exprimer	offrir	prévenir	exclamer	persévérer
annoncer	expliciter	mentionner	étrangler	gronder	opiner
rétorquer	expliquer	insinuer	hurler	rugir	obtempérer
demander	enchaîner	grincer	couper	rectifier	noter
remarquer	cracher	grogner	tempêter	rapporter	garantir
reprandre	décider	gronder	assener	tonner	corriger
relever	continuer	formuler	éructer	tempêter	convenir
commenter	confier	fanfaronner	râler	extasier	constater
recommander	commencer	exposer	rabrouer	époumoner	confirmer

expliquer	avertir	exprimer	intimer	emporter	concéder
confier	récapituler	expliciter	marteler	écrier	cautionner
rapporter	résumer	expliquer	époumoner	égosiller	assurer
proposer	arguer	escompter	rugir	réfuter	acquiescer
suggérer	conclure	espérer	tonner	proclamer	accorder
renseigner	argumenter	essayer	égosiller	invectiver	accepter
offrir	invectiver	dévoiler	cingler	interroger	affirmer.
indiquer	préciser	décider	insurger	insister	objecter
rappeler	alléguer	confier	renfrogner	exhorter	renfrogner
enquérir	terminer	avertir	bougonner	exiger	interrompre
promettre	suggérer	avancer	asséner	exploser	agacer
vouloir	souligner	étonner	rouspéter	crier	ruminer
assurer	récrier	surprendre	attaquer	croasser	protester
terminer	corriger	tiquer	agacer	confirmer	rectifier
susurrer	excuser	siffler	exploser	commander	défendre
finir	relever	récrier	vociférer	avancer	riposter
empresser	remarquer	consterner	récrier	appeler.	indigner
révéler	relater	offusquer	offusquer	marmonner	plaindre
résumer	réciter	insurger	râler	chuchoter	défiler
répéter	récapituler	indigner	rabrouer	murmurer	sommer
répliquer	recommander	excuser	mugir	soupirer	vouloir
répondre	réclamer	étrangler	trembler	ronronner	réfuter
reprendre	recommencer	agacer	trancher	couiner	pester
relater	rebondir	rougir	tonner	glisser	grimacer

réciter	rassurer	ricaner	tempêter	tempérer	renfrogner
énoncer	rappoter	reconnaître	sommer	susurrer	braquer
récapituler	quémander	jeter	siffler	soupirer	agacer
rappeler	questionner	douter	sermonner	souffler	ruminer
rappoter	proposer	interroger	braquer	souhaiter	rouspéter
rappoter	prétendre	hoqueter	insurger	renfrogner	repartir
parler	préconiser	grimacer	indigner	plaindre	objecter
mentionner	noter	avancer	acharner	étrangler	fulminer
lâcher	mentionner	haleter	riposter	rougir	corriger
lancer	lâcher	gémir	renchérir	ronchonne r	défendre
faire	lancer	trembler	rembarer	regretter	lâcher
exposer	jeter	susurrer	protester	prier	supplier
exprimer	formuler	soupirer	pester	marmonne r	corriger
expliciter	finir	plaindre	ordonner	grommele r	riposter
expliquer	faire	désoler	intimer	concéder	rectifier
dire	exposer	inquiéter	grincer	bredouille r	offusquer
discourir	exprimer	indigner	grogner	bougonner	crisper
discuter	expliciter	étrangler	gronder	bredouille r	excuser
dissenter	expliquer	regretter	fulminer	balbutier	corriger
continuer	éclaircir	râler	feuler	embrouill er	plaindre
confier	demander	prier	éructer	risquer	défiler
conclure	conclure	murmurer	cracher	bégayer	obstiner

compléter	compléter	implorer	bondir	rougir	enhardir
communiquer	communiquer	hoqueter	beugler	hésiter	assurer
commencer	argumenter	haleter	insister	rougir	rougir
ajouter	constater	hasarder	obstiner	présumer	rembarrer
informer	assurer	gémir	articuler	bafouiller	regretter
entendit-il	décréter	glapir	renchérir	hésiter	nier
enquit	commenter	cracher	piailler	enfonça	mentir
prévint	noter	chuchoter	souligner	tiquer	implorer
suggérer	observer	chevroter	persévérer	tergiverser	feuler
mentionner	supposer	bredouiller	persister	taire	fanfaronner
préciser	présumer	bégayer	acharner	songer	éluder
renseigner	inquiéter	confesser	marteler	plaindre	bondir
solliciter	excuser	avouer	recommencer	éluder	bafouiller
renseigner	exclamer	songer	sommer	insinuer	balbutier
enquérir	revendiquer	saluer	braquer	supposer	baragouiner
empresser	réclamer	obstiner	scander	solliciter	alléguer
révéler	reconnaître	insurger	insurger	désoler	plaindre
quémander	promettre	réclamer	exiger	étrangler	gémir
questionner	juger	rassurer	exclamer	embrouiller	glapir
mentionner	féliciter	promettre	acharner	agacer	couiner
lâcher	dénoncer	maintenir	riposter	objecter	plaindre
lancer	ajouter	glisser	réprimander	nier	sursauter
indiquer	inquiéter	garantir	regretter	mentir	crisper
informer	excuser	éclaircir	réfuter	implorer	piailler
formuler	exclamer	confesser	réclamer	hoqueter	divulguer

éclaircir	suggérer	assurer	prétendre	hasarder	cracher
demander	éluder	asséner	finir	éluder	poursuivre
communiquer	proposer	chantonner	recommander	douter	persévérer
aviser	insinuer	rire	reconnaître	corriger	intimer
arguer	requérir	sourire	rapporter	chevroter	intervenir
argumenter	concéder	encourager	proposer	calculer	attendre
apostropher	trancher	roucouler	promettre	bredouiller	morigéner
appeler	terminer	rassurer	prévenir	balbutier	juger
annoncer	supposer	esclaffer	pester	baragouiner	insister
observer	suggérer	féliciter	persévérer	ricaner	imposer
maintenir	souhaiter	siffler	noter	ironiser	escompter
formuler	vouloir	extasier	ordonner	glousser	espérer
demander	solliciter	exclamer	menacer	amuser	essayer
ajouter	préconiser	amuser	juger	moquer	discourir
empresser	noter	rire	invectiver	persifler	dissenter
apostropher	jeter	glousser	interroger	invectiver	corriger
enchaîner	intimer	ronronner	imposer	imaginer	interrompre
terminer	renseigner	plaisanter	garantir	glisser	couper
souligner	renfrogner	pouffer	exhorter	fanfaronner	insinuer
saluer	corriger	offrir	exploser	convenir	insister
inquiéter	scander	féliciter	défendre	accepter	imposer
excuser	exclamer	apprécier	continuer	approuver	héler
étrangler	étonner	relativiser	constater	reconnaître	glisser

étonner	rétorquer	imaginer	hurler	admettre	fanfaronner
enhardir	renchérir	encourager	vociférer	défendre	attendre
empresser	rembarrer	divulguer	pesta	accorder	
risquer	relever	désirer	jeter	acquiescer	

Annexe 3 Les algorithmes d'extraction

L'algorithme d'extraction pour les inversions du sujet

Algorithm 1 l'algorithm pour extraire les inversion du sujet

```
1: function←column 8
2: dependancy←column 7
3: index←column 1
4: word←column 2
5: lemma←column 3
6: if function= 'nsubj' or 'nsubj: pass' then
7:   if dependancy ≤ index then
8:     if lemma[rank] is not in the list of incise then
9:       print the phrase
10:      print lemma[rank]
11:      print lemma
```

L'algorithme d'extraction pour les clivages

Algorithm 1 l'algorithme pour extraire les clivages

lemma ← column 3

if lemma = 'ce être' or 'ce ne être' or 'être ce' or 'ne être ce' **then**

if 'que' or 'qui' is found behind the four expressions **then**

 print the phrase

L'algorithme d'extraction pour les répétitions

Algorithm 1 l'algorithme pour extraire les parallélismes

```
functions←column 8
dependancys←column 7
indexes←column 1
words←column 2
lemmas←column 3
for every lemma in lemmas do
  if lemma is in the list of fonctional words then
    ChangedTags←lemma
  else if lemma is in the list of fonctional words then
    ChangedTags←POS

for every index in indexes do
  save all the nodes of the dependancy tree
for every node saved do
  save all the subtrees in the dependancy tree for every node
while Depth of extraction equals input do
  find all the subtrees with the reaquired depth
  for every subtree in subtrees do
    for i in range of the length of subtree do
      if index[i]-index[i-1] ≡ 1 then
        save all the subtrass in a dictionnay and count their frequency
        if frequency ≥ 2 then
          print the phrase
          print the subtree by word order in ChangedTags
          print the subtree by word order in original text
```

L'algorithme d'extraction non-supervisée des patrons syntaxiques

Algorithm 1 l'algorithme pour extraire les parallélismes

```
functions←column 8
dependancys←column 7
indexes←column 1
words←column 2
lemmas←column 3
POS←column 4
for every lemma in lemmas do
  if lemma is in the list of fonctional words then
    ChangedTags←lemma
  else if lemma is in the list of fonctional words then
    ChangedTags←POS

for every index in indexes do
  save all the nodes of the dependancy tree
for every node saved do
  save all the subtrees in the dependancy tree for every node
while Depth of extraction equals input do
  find all the subtrees with the reaquired depth
  for every subtree in subtrees do
    for i in range of the length of subtree do
      if index[i]-index[i-1] ≡ 1 then
        print the phrase
        print the subtree by word order in original text
        print the subtree by word order in ChangedTags
```

Bibliographie

La date indiquée entre crochets correspond à la date de première publication de l'ouvrage, ou dans le cas des correspondances, à la date d'écriture de la lettre.

Aarts, Bas. 2008. *English syntax and argumentation*. Basingstoke, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, Pays multiples.

Aljawarneh, Shadi, Muneer Bani Yassein, et Mohammed Aljundi. 2019. « An enhanced J48 classification algorithm for the anomaly intrusion detection systems ». *Cluster Computing* 22 (septembre): 1-17. <https://doi.org/10.1007/s10586-017-1109-8>.

Archer, Dawn, Jonathan Culpeper, et Paul Rayson. 2009. « Love - “a familiar or a devil”? An Exploration of Key Domains in Shakespeare’s Comedies and Tragedies », janvier.

Argamon, Shlomo, et Levitan Shlomo. 2005. « Measuring the Usefulness of Function Words for Authorship Attribution ».

Argamon, Shlomo, Casey Whitelaw, Paul Chase, Sobhan Raj Hota, Navendu Garg, et Shlomo Levitan. 2007. « Stylistic Text Classification Using Functional Lexical Features ». *Journal of the American Society for Information Science and Technology* 58 (6): 802-22. <https://doi.org/10.1002/asi.20553>.

Argamon-Engelson, Shlomo, Moshe Koppel, et Galit Avneri. s. d. « Style-Based Text Categorization: What Newspaper Am I Reading? », 4.

Aristote, et Myriam Hecquet-Devienne. 2008 [IVe siècle av. J.-C.]. *Métaphysique gamma: édition, traduction, études*. Édité par Annick Stevens. Louvain-la-Neuve, Belgique, France, Etats-Unis d'Amérique.

- Baayen, H, H van Halteren, et F Tweedie. 1996. « Outside the cave of shadows: using syntactic annotation to enhance authorship attribution ». *Literary and Linguistic Computing* 11 (3): 121-32. <https://doi.org/10.1093/lc/11.3.121>.
- Bally, Charles. 1905. *Précis de stylistique: esquisse d'une methode fondée sur l'étude du français moderne*. Genève, Suisse: A. Eggimann,.
- Barrès, Maurice. 1931. *Mes cahiers*. Paris, France: Librairie Plon.
- Brault-Scappini, Sophie-Anne. 2007. « Étude du dispositif d'extraction en « c'est...qu », différenciation entre une relative en « c'est...qu » et une proposition clivée ». *L'information grammaticale* 114 (1): 53-56. <https://doi.org/10.3406/igram.2007.4453>.
- Brombert, Victor, et Institut d'études françaises. 1954. *Stendhal et la voie oblique , l'auteur devant son monde romanesque, par Victor Brombert*. New Haven Paris: Yale University press Presses universitaires de France (Mayenne, impr. de Floch).
- Brombert, Victor, et Héloïse Neefs. 1985. *Victor Hugo et le roman visionnaire*. Paris, France: PUF.
- Brown, Edward James. 1973. *Major Soviet writers: essays in criticism*. London, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord.
- Brunot, Ferdinand, et Charles Bruneau. 1948. *L'époque romantique*. 1 vol. Histoire de la langue française, des origines à 1900 12. Paris: A. Colin.
- Buffon, Georges-Louis Leclerc. 1886. *Discours sur le style*. Paris, France: C. Delagrave.
- Burrows, J. F. 1986. « Modal Verbs and Moral Principles: An Aspect of Jane Austen's Style ». *Literary and Linguistic Computing* 1 (1): 9-23. <https://doi.org/10.1093/lc/1.1.9>.
- Cahen, D. 1966. *Entretien avec Roland Barthes*. S.l.: Aletheia.
- Candito, Marie, et Djamé Seddah. 2012. « Le corpus Sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical ». In *TALN 2012 -*

- 19e conférence sur le Traitement Automatique des Langues Naturelles*. Grenoble, France. <https://hal.inria.fr/hal-00698938>.
- Carter, Ronald. 2010. « Methodologies for Stylistic Analysis: Practices and Pedagogies », janvier. https://doi.org/10.1007/978-1-137-06574-2_4.
- Carter, Ronald, et Peter Stockwell, éd. 2008. *The Language and Literature Reader*. 1 vol. London: Routledge.
- Cer, Daniel, Marie-Catherine de Marneffe, Dan Jurafsky, et Chris Manning. 2010. « Parsing to Stanford Dependencies: Trade-offs between Speed and Accuracy ». In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valletta, Malta: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2010/pdf/730_Paper.pdf.
- Chapman, Raymond. 1973. *Linguistics and literature: an introduction to literary stylistics*. London, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord: Edward Arnold, 1973.
- Chateaubriand, François-René de. 2013. *Les oeuvres intégrales - François-René de Chateaubriand - Ligram éditions - ebook (ePub) - Le Hall du Livre NANCY*. Ligram éditions. <https://halldulivre.com/ebook/9791027400058-francois-rene-de-chateaubriand-les-oeuvres-completes-edition-augmentee-francois-rene-chateaubriand/>.
- Chatman, Seymour Benjamin, éd. 1971. *Literary style: a symposium*. London, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, Etats-Unis d'Amérique.
- Chomsky, Noam. 1957. *Syntactic structures*. The Hague, Pays-Bas, France.
- Colas-Blaise, Marion, et Claire Stolz. 2016. « Comment faire dialoguer stylistique et sémiotique ? Éléments pour une pensée de la frontière disciplinaire ». In *Stylistiques ?*, édité par Laurence Bougault et Judith Wulf, 99-110. Interférences. Rennes: Presses universitaires de Rennes. <http://books.openedition.org/pur/40061>.

- Covington, Michael A. 2009. *Syntactic Theory in the High Middle Ages: Modistic Models of Sentence Structure*. Cambridge: Cambridge University Press.
- Culpeper, Jonathan. 2009. « Keyness: Words, Parts-of-Speech and Semantic Categories in the Character-Talk of Shakespeare's Romeo and Juliet ». *International Journal of Corpus Linguistics* 14 (1): 29-59. <https://doi.org/10.1075/ijcl.14.1.03cul>.
- Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, et Richard Harshman. 1990. « Indexing by Latent Semantic Analysis ». *Journal of the American Society for Information Science* 41 (6): 391-407. [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<391::AID-ASII>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASII>3.0.CO;2-9).
- Dessaintes, Maurice. 1960. *Éléments de linguistique descriptive en fonction de l'enseignement du français*. Namur : La Procure. <http://archive.org/details/elementsdelingui0000unse>.
- Diederich, Joachim, Jörg Kindermann, Edda Leopold, et Gerhard Paass. 2003. « Authorship Attribution with Support Vector Machines ». *Applied Intelligence* 19 (1): 109-23. <https://doi.org/10.1023/A:1023824908771>.
- Dmitriev, Aleksander. 2002. « Le contexte européen (français et allemand) du formalisme russe ». *Cahiers du monde russe. Russie - Empire russe - Union soviétique et États indépendants* 43 (43/2-3): 423-40. <https://doi.org/10.4000/monderusse.8505>.
- Dubois, Jean, Mathée Giacomo-Marcellesi, Louis Guespin, Christiane Marcellesi, Jean-Baptiste Marcellesi, et Jean-Pierre Mével. 2012. *Le dictionnaire de linguistique et des sciences du langage*. Paris, France: Larousse.
- Dubremetz, Marie. 2017. « Detecting Rhetorical Figures Based on Repetition of Words: Chiasmus, Epanaphora, Epiphora ». <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-334486>.
- Dufour, Philippe. 2004. *La pensée romanesque du langage*. Paris: Seuil.

- Eder, Maciej, Jan Rybicki, et Mike Kestemont. 2016. « Stylometry with R: A Package for Computational Text Analysis ». *The R Journal* 8 (1): 107-21.
- Empson, William. 1953. *Seven Types of Ambiguity*. London, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord: Chatto and Windus.
- Fahnestock, Jeanne. 1999. *Rhetorical Figures in Science*. Oxford University Press.
- Ferrère, Étienne-Louis. 1913. « L'esthétique de Gustave Flaubert ». Thèse, France: Université de Paris (1896-1968). Faculté des lettres.
- Fish, Stanley E. 1979. « What is Stylistics and Why Are They Saying Such Terrible Things about It?-Part II ». *boundary 2 boundary 2* 8 (1): 129.
- Fish, Stanley Eugene. 1980. *Is There a Text in This Class?: The Authority of Interpretive Communities*. Cambridge, Mass.: Harvard University Press.
- Flaubert, Gustave. 1852. « Flaubert à Louise Colet, Croisset, 24 avril 1852 ». 1852. <https://flaubert.univ-rouen.fr/jet/public/correspondance/trans.php?corpus=correspondance&id=9920>.
- . 1926. *Correspondance*. Paris, France: L. Conard.
- . 2021. *Flaubert l'intégrale : Œuvres complètes, 69 titres*. 1er édition. Ink book.
- Fowler, Roger. 1981. *Literature as Social Discourse: The Practice of Linguistic Criticism*. Bloomington: Indiana university press.
- Fromilhague, Catherine, et Claude Thomasset. 2007. *Les figures de style*. Paris, France: A. Colin, DL 2007.
- Frontini, Francesca, Mohamed Amine Boukhaled, et Jean-Gabriel Ganascia. 2017. « Mining for characterising patterns in literature using correspondence analysis: an experiment on French novels ». *Digital Humanities Quarterly*, Göttingen Dialog in Digital Humanities 2015, 11 (2). <https://hal.archives-ouvertes.fr/hal-01527780>.
- Fuchs, Catherine. 2019. « La postposition du sujet nominal : paramètres linguistiques et effets stylistiques ». In *L'Ordre des mots à la lecture des textes*, édité par Agnès

- Fontvieille-Cordani et Stéphanie Thonnérioux, 27-44. Textes & Langue. Lyon: Presses universitaires de Lyon. <http://books.openedition.org/pul/2472>.
- Galperin, I. R. 1997. *Stylistics*. La Habana: s.n.
- Gamon, Michael. 2004. « Linguistic correlates of style: authorship classification with deep linguistic analysis features ». In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, 611-17. Geneva, Switzerland: COLING. <https://www.aclweb.org/anthology/C04-1088>.
- Ganascia, Jean-Gabriel. 2001. « Extraction automatique de motifs syntaxiques ». In *TALN 2001*. Tours, France. <https://hal.archives-ouvertes.fr/hal-01570833>.
- Garvin, Paul Lucian. 1964. *A Prague School Reader on Esthetics, Literary Structure, and Style. Selected and Translated ... by Paul L. Garvin. (Third Printing.)*. Georgetown University Press: Washington.
- Gawryjolek, Jakub Jan. 2009. « Automated Annotation and Visualization of Rhetorical Figures », mai. <https://uwspace.uwaterloo.ca/handle/10012/4426>.
- Gide, André Éditeur scientifique Préfacier. 1949. *Anthologie de la poésie française*. Paris, France: Gallimard.
- Givón, Talmy. 2009. *The Genesis of Syntactic Complexity: Diachrony, Ontogeny, Neuro-Cognition, Evolution*. Amsterdam Philadelphia: John Benjamins Pub. Co.
- Graesser, Arthur C, Danielle S Mcnamara, Max M Louwerse, et Zhiqiang Cai. 2004. « Coh-Matrix: Analysis of Text on Cohesion and Language ». *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc.* 36 (2): 193.
- Green, Georgia M. 1982. « Linguistics and the Pragmatics of Language Use ». *Poetics* 11 (1): 45-76. [https://doi.org/10.1016/0304-422X\(82\)90004-3](https://doi.org/10.1016/0304-422X(82)90004-3).
- Groden, Michael, Martin Kreiswirth, Imre Szeman, et Professor of English Film Studies and Sociology Imre Szeman. 2005. *The Johns Hopkins Guide to Literary Theory and Criticism*. Johns Hopkins University Press.

- Grossman, Kathryn M. 1986. *The Early novels of Victor Hugo: towards a poetics of harmony*. Genève, Suisse: Droz.
- Hallen, Cynthia L, et Jennifer Shakespear. 2002. « The T-Unit as a Measure of Syntactic Complexity in Emily Dickinson's Poems ». *Emily Dickinson Journal* 11 (1): 91-103.
- Halliday, M. 2020. « Linguistic function and literary style: an inquiry into the language of William Golding's *The Inheritors* ». In . <https://doi.org/10.4324/9781003060789-4>.
- Halteren, Hans Van. 2007. « Author verification by linguistic profiling: An exploration of the parameter space ». *ACM Transactions on Speech and Language Processing* 4 (1): 1:1-1:17. <https://doi.org/10.1145/1187415.1187416>.
- Hollingsworth, Charles. 2012. « Syntactic Stylometry: Using Sentence Structure for Authorship Attribution ». 2012. </paper/Syntactic-Stylometry%3A-Using-Sentence-Structure-for-Hollingsworth/a7e194c7f18f2c4f83993207f737733ff31f03b1>.
- Hovasse, Jean-Marc. 2001. *Victor Hugo : 1802-1851 , I: Avant l'exil*. Victor Hugo. Paris: Fayard.
- Hugo, Victor. 1985 [1827] [1901]. *Oeuvres complètes*. Édité par Guy Rosa, Jacques Seebacher, et Groupe Hugo. 16 vol. Paris, France: R. Laffont.
- Hugo, Victor, et ' Arvensa Editions. 2013. *Victor Hugo: Oeuvres complètes - 122 titres (Annotés et illustrés) - Arvensa Editions*. Arvensa Editions.
- Huguet, Edmond. 1905. *Les métaphores et les comparaisons dans l'oeuvre de Victor Hugo*. Paris : Hachette. <http://archive.org/details/lesmtaphoreset00hugu>.
- Jakobson, Roman. 1966. *Grammatical parallelism and its Russian facet*. New York, Etats-Unis d'Amérique.
- Jakobson, Roman, et Tzvetan Todorov. 1977. *Huit questions de poétique*. Paris, France: Éditions du Seuil.

- Jenny, Laurent. 1995. « L'objet singulier de la stylistique ». *Bilan de l'École de Genève : (actes du colloque polono-suisse organisé par l'Institut de Philologie Romane, le Centre de Civilisation Française de l'Université de Varsovie et le Département de langue et de littérature françaises modernes de l'Université de Genève : Varsovie, mai 1992) Cahiers de Varsovie : publication du Centre de Civilisation Française de l'Université de Varsovie*, 19-27.
- Karlgren, Jussi, et Gunnar Eriksson. 2007. « Authors, Genre, and Linguistic Convention ».
- Kaster, Andreas, Stefan Siersdorfer, et Gerhard Weikum. 2005. « Combining Text and Linguistic Document Representations for Authorship Attribution ». In *In SIGIR Workshop: Stylistic Analysis of Text for Information Access (STYLE, 27-35*.
- Kübler, Sandra, Ryan McDonald, et Joakim Nivre. 2008. « Dependency Parsing ». *Synthesis Lectures on Human Language Technologies 2* (1): 1-127. <https://doi.org/10.2200/S00169ED1V01Y200901HLT002>.
- Kukushkina, O. V., A. A. Polikarpov, et D. V. Khmelev. 2001. « Using Literal and Grammatical Statistics for Authorship Attribution ». *Problems of Information Transmission* 37 (2): 172-84. <https://doi.org/10.1023/A:1010478226705>.
- Landauer, Thomas K., Peter W. Foltz, et Darrell Laham. 1998. « An introduction to latent semantic analysis ». *Discourse Processes* 25 (2-3): 259-84. <https://doi.org/10.1080/01638539809545028>.
- Lebart, Ludovic, André Salem, et Christian Préfacier Baudelot. 1994. *Statistique textuelle*. Paris, France: Dunod.
- Lecorvé, Gwénolé. 2020. « Traitement Automatique Du Style Dans Le Langage Naturel : Quelques Contributions et Perspectives », thèse d'habilitation.
- Leech, Geoffrey Neil, et Michael Henry Short. 2013. *Style in fiction: a linguistic introduction to English fictional prose*. Harlow, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord: Longman.

- Legallois, Dominique, Thierry Charnois, et Meri Larjavaara, éd. 2018. *The grammar of genres and styles: from discrete to non-discrete units*. Berlin, Allemagne: De Gruyter Mouton.
- Legallois, Dominique, Thierry Charnois, et Thierry Poibeau. 2016. « Repérer les clichés dans les romans sentimentaux grâce à la méthode des « motifs » ». *Lidil. Revue de linguistique et de didactique des langues*, n° 53 (mai): 95-117. <https://doi.org/10.4000/lidil.3950>.
- Lemon, Lee T. Traducteur Préfacier, et Marion J. Traducteur Préfacier Reis. 1965. *Russian formalist criticism: four essays*. Lincoln (Nebraska), Etats-Unis d'Amérique: University of Nebraska Press.
- Longrée, Dominique, et Sylvie Mellet. 2013. « Le motif : une unité phraséologique englobante ? Étendre le champ de la phraséologie de la langue au discours* ». *Langages*, Vers une extension du domaine de la phraséologie, 1 (189): 65-79.
- Mahlberg, Michaela. 2007. « Corpus Stylistics: Bridging the Gap between Linguistic and Literary Studies ». [/paper/Corpus-stylistics%3A-bridging-the-gap-between-and-Mahlberg/496d26efbc8dadada177748030c0a582bee47807](http://paper/Corpus-stylistics%3A-bridging-the-gap-between-and-Mahlberg/496d26efbc8dadada177748030c0a582bee47807).
- Manning, Christopher, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, et David McClosky. 2014. « The Stanford CoreNLP Natural Language Processing Toolkit ». In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 55-60. Baltimore, Maryland: Association for Computational Linguistics. <https://doi.org/10.3115/v1/P14-5010>.
- Marill-Albérès, Francine. 1956. « Le Naturel chez Stendhal ». Paris, France: Nizet.
- Matthews, Peter H. 2007. *Syntactic Relations: A Critical Survey*. Cambridge: Cambridge Univ. Press.
- Maupassant, Guy de, et Noëlle Benhamou. 2019 [1888]. *Pierre et Jean*. Paris: Flammarion.

- Mccarthy, Philip, Gwyneth Lewis, David Dufty, et Danielle McNamara. 2006. *Analyzing Writing Styles with Coh-Metrix. FLAIRS 2006 - Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference*. Vol. 2006.
- Melcuk, Igor A. 1988. *Dependency Syntax: Theory and Practice*. Albany (N.Y.): State University of New York Press.
- Mendenhall, T. C. 1887. « THE CHARACTERISTIC CURVES OF COMPOSITION ». *Science* ns-9 (214S): 237-46. <https://doi.org/10.1126/science.ns-9.214S.237>.
- Molinié, Georges Organisateur de réunion, Pierre Organisateur de réunion Cahné, Colloque international Qu'est-ce que le style ?, et École doctorale Sciences du langage. 1994. *Qu'est-ce que le style ? : actes du colloque international*. Paris, France: Presses universitaires de France.
- Molinié, Georges, et Alain Viala. 1993. *Approches de la réception*. Presses Universitaires de France. <https://doi.org/10.3917/puf.molin.1993.01>.
- Molino, Jean. 1981. *Sur le parallélisme morpho-syntaxique*. Paris, France: Larousse.
- Morel, Mary-Annick. 1980. « L'inversion du sujet dans certaines propositions concessives du français : Conditions d'emploi et valeur ». *DRLAV. Documentation et Recherche en Linguistique Allemande Vincennes* 22 (1): 105-12. <https://doi.org/10.3406/drlav.1980.959>.
- Mosteller, Frederick, et David Wallace. 1984. *Applied Bayesian and classical inference: the case of The Federalist papers*. New York, Etats-Unis d'Amérique, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord.
- Mourot, Jean. 1960. « Le Génie d'un style, Chateaubriand, rythme et sonorité dans les «Mémoires d'Outre-tombe» : Thèse pour le doctorat és lettres présentée à la Faculté des Lettres de l'université de Paris ». S. l.
- Müller-Hauser, M.-L. 1943. *La mise en relief d'une idée en français moderne*. Genève.

- Novakova, Iva, et Dirk Siepmann. 2020. *Phraseology and Style in Subgenres of the Novel : A Synthesis of Corpus and Literary Perspectives*. Palgrave Macmillan (Springer). <https://hal.univ-grenoble-alpes.fr/hal-02077476>.
- Peacham, Henry. 2001. *The Garden of Eloquence (1593)*. Ann Arbor: Scholar Press.
- Philippe, Gilles. 2005. « Traitement stylistique et traitement idiolectal des singularités langagières ». *Cahiers de praxématique*, n° 44 (janvier): 77-92. <https://doi.org/10.4000/praxematique.1659>.
- . 2013a. *Le rêve du style parfait*. 1 vol. Paris: Presses universitaires de France.
- . 2013b. « Les deux corps du style ». *Les Temps Modernes* n° 676 (5): 144-54.
- . 2017. *French style: l'accent français de la prose anglaise*. Bruxelles, Belgique: les Impressions nouvelles.
- Puková, Lucie. 2018. « Lyrisation et dramatisation dans Les Misérables de Victor Hugo ». Masarykova univerzita, Filozofická fakulta. <https://theses.cz/id/ybgbt0/>.
- Posadas Durán, Juan, Grigori Sidorov, Helena Gomez Adorno, Ildar Batyrshin, Elibeth Mirasol-Mélendez, Gabriela Posadas-Durán, et Liliana Chanona-Hernández. 2017. « Algorithm for Extraction of Subtrees of a Sentence Dependency Parse Tree ». *Acta Polytechnica Hungarica* 14 (novembre): 79. <https://doi.org/10.12700/APH.14.3.2017.3.5>.
- Quintilian. 1921. *The Instituto Oratoria. With an English Translation by H.E. Butler*. London: Heinemann.
- Rastier, François. 2001. *Arts et sciences du texte*. Presses Universitaires de France. <https://doi.org/10.3917/puf.rast.2001.01>.
- Rayson, Paul. s. d. *Wmatrix: A Web-Based Corpus Processing Environment*.
- Richards, Ivor Armstrong. 1965. *The philosophy of rhetoric*. London, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, Etats-Unis d'Amérique.
- Riffaterre, Michael. 1957. « Le style des Pléiades de Gobineau: essai d'application d'une méthode stylistique ». Genève, Suisse, France: E. Droz.

- . 1959. « Criteria for Style Analysis ». *WORD* 15 (1): 154-74. <https://doi.org/10.1080/00437956.1959.11659690>.
- . 1979. *La production du texte*. Paris, France: Éditions du Seuil.
- Riffaterre, Michael, et Daniel Delas. 1971. *Essais de stylistique structurale*. Paris, France: Flammarion.
- Roig, Audrey. 2017. « L'inversion du sujet (pro)nominal : un indice syntaxique ? » *Verbum (Presses Universitaires de Nancy)* XXXIX (2): 193.
- Sainte-Beuve, Charles-Augustin. 1861 [1843]. *Chateaubriand et son groupe littéraire sous l'Empire: cours professé à Liège en 1848-1849*. 2 vol. Paris, France: Garnier frères.
- . 1926. *Mes poisons: cahiers intimes inédits*. Édité par Victor Giraud. Paris, France: Plon-Nourrit et Cie, impr.1926.
- Sartre, Jean-Paul. 1972 [1971]. *L'idiot de la famille*. Paris, France: Gallimard.
- Schopenhauer, Arthur. 2020 [1851]. *Essays of Schopenhauer*. Otbebookpublishing. <http://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9783965379558>.
- Sidorov, Grigori, Francisco Velasquez, Efstathios Stamatatos, Alexander Gelbukh, et Liliana Chanona-Hernández. 2014. « Syntactic N-Grams as Machine Learning Features for Natural Language Processing ». *Expert Systems with Applications, Methods and Applications of Artificial and Computational Intelligence*, 41 (3): 853-60. <https://doi.org/10.1016/j.eswa.2013.08.015>.
- Spitzer, Leo. 1948. *Linguistics and literary history: essays in stylistics*. Princeton, Etats-Unis d'Amérique: Princeton University Press.
- Spitzer, Leo, Jean Starobinski, Éliane Kaufholz-Messmer, Alain Coulon, et Michel Foucault. 1970. *Études de style /Précédé de Leo Spitzer et la lecture stylistique*. Paris: Gallimard.

- Stamatatos, Efstathios, Nikos Fakotakis, et Kokkinakis George. 2000. « Automatic Text Categorization In Terms Of Genre, Author. » *Computational Linguistics* 26 (décembre): 471-95. <https://doi.org/10.1162/089120100750105920>.
- . 2001. « Computer-Based Authorship Attribution Without Lexical Measures ». *Computers and the Humanities* 35 (mai): 193-214. <https://doi.org/10.1023/A:1002681919510>.
- Stendhal. 1999. *Correspondance générale*. Édité par Victor Del Litto. *Correspondance générale*. Paris, France: H. Champion.
- Stendhal, et ' Arvensa Editions. 2014. *Stendhal : Oeuvres complètes*. Arvensa Editions.
- Stendhal, et Romain Préfacier Rolland. 1914 [1814]. *Vies de Haydn de Mozart et de Métastase*. Édité par Daniel Muller. Paris, France: H. Champion.
- Straka, Milan, et Jana Straková. 2017. « Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe ». In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 88-99. Vancouver, Canada: Association for Computational Linguistics. <https://doi.org/10.18653/v1/K17-3009>.
- Suhamy, Henri. 2020. *Les figures de style*. Vol. 14e éd. Que sais-je ? Paris cedex 14: Presses Universitaires de France. <https://www.cairn.info/les-figures-de-style--9782715403413.htm>.
- Szmrecsanyi, Benedikt, Gérard Purnelle, Gérard Fairon, et Anne Dister. 2004. « On Operationalizing Syntactic Complexity ». *Le Poids Des Mots. Proceedings of the 7th International Conference on Textual Data Statistical Analysis. Louvain-La-Neuve, March 10-12, 2004*, 1032-39.
- Talbot, Emile J., éd. 1979. *La critique stendhalienne de Balzac à Zola*. York (S.C.), Etats-Unis d'Amérique: French Literature.
- . 1985. *Stendhal and Romantic Esthetics*. French Forum Monographs 61. Lexington, Ky: French forum.

- Tellier, Christine, et Daniel Valois. 2018. *Constructions méconnues du français. Constructions méconnues du français*. Paramètres. Montréal: Presses de l'Université de Montréal. <http://books.openedition.org/pum/10250>.
- Tesnière, Lucien. 1988 [1959]. *Éléments de syntaxe structurale*. 2e éd. rev. et Corr. Paris: Klincksieck.
- « The Language and Literature Reader ». s. d. Routledge & CRC Press. Consulté le 14 mars 2021. <https://www.routledge.com/The-Language-and-Literature-Reader/Carter-Stockwell/p/book/9780415410038>.
- Thibaudet, Albert. 1935. *Gustave Flaubert*. 4e éd. 1 vol. Paris: Gallimard.
- Ubersfeld, Anne. 2001. *Le roi et le bouffon : étude sur le théâtre de Hugo de 1830 à 1839 (Éd. rev.) / Anne Ubersfeld*. <https://gallica.bnf.fr/ark:/12148/bpt6k3377187p>.
- Ullmann, Stephen. 1957. *Style in the French novel*. Cambridge Eng., Royaume-Uni de Grande-Bretagne et d'Irlande du Nord: University Press.
- . 1966. *Language and Style : Collected Papers*. Language and Style Series 1. Oxford: B. Blackwell.
- Urieli, Assaf. 2013. « Robust French Syntax Analysis : Reconciling Statistical Methods and Linguistic Knowledge in the Talisman Toolkit ». Phdthesis, Université Toulouse le Mirail - Toulouse II. <https://tel.archives-ouvertes.fr/tel-01058143>.
- Urieli, Assaf, et Ludovic Tanguy. 2013. « L'apport du faisceau dans l'analyse syntaxique en dépendances par transitions : études de cas avec l'analyseur Talisman ». In , (publication en ligne). <https://halshs.archives-ouvertes.fr/halshs-00953754>.
- Wake, William C. 1957. « Sentence-Length Distributions of Greek Authors ». *Journal of the Royal Statistical Society. Series A (General)* 120 (3): 331-46. <https://doi.org/10.2307/2343104>.
- Williams, C. B. 1940. « A Note on the Statistical Analysis of Sentence-Length as a Criterion of Literary Style ». *Biometrika* 31 (3/4): 356-61. <https://doi.org/10.2307/2332615>.

- Wulf, Judith. 2014. *Étude sur la langue romanesque de Victor Hugo: le partage et la composition*. Paris, France: Classiques Garnier.
- Wynne, Martin. 2006. *Stylistics : Corpus Approaches*. Elsevier. <https://ora.ox.ac.uk/objects/uuid:fd469124-f662-4dab-942e-91dd16620692>.
- Yule, G. Udny. 1939. « On Sentence-Length as a Statistical Characteristic of Style in Prose: With Application to Two Cases of Disputed Authorship ». *Biometrika* 30 (3/4): 363-90. <https://doi.org/10.2307/2332655>.
- Zhao, Ying, et Justin Zobel. 2007. « Searching with style: authorship attribution in classic literature ». In *Proceedings of the thirtieth Australasian conference on Computer science - Volume 62*, 59-68. ACSC '07. AUS: Australian Computer Society, Inc.
- Zola, Émile. 1895. *Les romanciers naturalistes*. Paris, France: Bibliothèque-Charpentier.

RÉSUMÉ

Comment l'analyse syntaxique automatique peut-elle être utile pour identifier et caractériser les styles littéraires ? Dans cette thèse, nous proposons de répondre à cette question en considérant les patrons syntaxiques comme des marqueurs stylistiques performants pour identifier, mais aussi caractériser les styles littéraires dans les romans français du 19^{ème} siècle. La méthode d'extraction des patrons caractérisants (à savoir les figures de style d'ordre syntaxique) et la méthode d'extraction des patrons identificatoires (dont la spécificité est de nature purement statistique) sont appliquées à l'œuvre romanesque de Victor Hugo d'une part, et à un corpus plus grand, constitué des romans de Chateaubriand, Flaubert, Hugo et Stendhal d'autre part. Notre méthode combinatoire démontre la capacité satisfaisante des patrons syntaxiques à différencier les styles littéraires, et permet de confirmer (ou au contraire, parfois, de contredire) les observations passées sur les auteurs visés.

MOTS CLÉS

stylistique computationnelle, stylométrie, analyse syntaxique en dépendances, patrons syntaxiques, extraction syntaxique, modélisation statistique

ABSTRACT

How can automatic syntactic analysis be useful to identify and characterize literary styles? In this thesis, we propose to answer this question by considering syntactic patterns as efficient stylistic markers to identify, but also to characterize literary styles in 19th century French novels. The method for extracting characterising patterns (i.e. syntactic figures of speech) and the method for extracting identifying patterns (whose specificity is of a purely statistical nature) are applied to the novels of Victor Hugo on the one hand, and to a larger corpus, consisting of the novels of Chateaubriand, Flaubert, Hugo and Stendhal on the other. Our combinatorial method demonstrates the satisfactory capacity of syntactic patterns to differentiate literary styles, and makes it possible to confirm (or, on the contrary, sometimes contradict) past observations on the styles of the authors concerned.

KEYWORDS

computational stylistics, stylometry, dependency parsing, syntactic patterns, syntactic extraction, statistical modeling