



**HAL**  
open science

# Les phraséologismes autour de la boisson dans le roman contemporain : un moyen pour distinguer les sous-genres romanesques ?

Aneta Marzec

## ► To cite this version:

Aneta Marzec. Les phraséologismes autour de la boisson dans le roman contemporain : un moyen pour distinguer les sous-genres romanesques ?. Sciences de l'Homme et Société. 2020. dumas-02972164

**HAL Id: dumas-02972164**

**<https://dumas.ccsd.cnrs.fr/dumas-02972164>**

Submitted on 20 Oct 2020

**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.



# **Les phraséologismes autour de la boisson dans le roman contemporain : un moyen pour distinguer les sous-genres romanesques?**

**Aneta  
MARZEC**

Sous la direction de Mme Julie Sorba

Laboratoire : LIDILEM

UFR LLASIC

Département des Sciences du Langage et Français Langue Étrangère

---

Mémoire de master 2 recherche - 30 crédits - Mention Sciences du Langage

Spécialité : Linguistique

Année universitaire 2019-2020



**Les phraséologismes autour de la boisson  
dans le roman contemporain :  
un moyen pour distinguer  
les sous-genres romanesques?**

**Aneta  
MARZEC**

Sous la direction de Mme Julie Sorba

Laboratoire : LIDILEM

UFR LLASIC

Département des Sciences du Langage et Français Langue Étrangère

---

Mémoire de master 2 recherche - 30 crédits - Mention Sciences du Langage

Spécialité : Linguistique

Année universitaire 2019-2020





## **Remerciements**

Je tiens à remercier toutes les personnes qui m'ont aidée dans la réalisation de ce mémoire.

Je tiens premièrement à remercier ma directrice de mémoire, Madame Julie Sorba, pour m'avoir encouragée, encadrée, et donné de précieux conseils tout au long de la rédaction de ce mémoire. J'ai beaucoup apprécié sa patience, sa disponibilité et sa sympathie.

Je voudrais exprimer ma gratitude à Madame Teresa Muryn et Madame Małgorzata Niziołek de l'Université Pédagogique de Cracovie pour m'avoir donné l'occasion de participer dans le programme de double diplôme avec l'Université Grenoble-Alpes.

Enfin, j'adresse mes plus sincères remerciements à ma famille et à mes amis de Pologne qui m'ont encouragée et épaulée de près ou de loin pendant la rédaction de ce mémoire.

### **DÉCLARATION ANTI-PLAGIAT**

1. Ce travail est le fruit d'un travail personnel et constitue un document original.
2. Je sais que prétendre être l'auteur d'un travail écrit par une autre personne est une pratique sévèrement sanctionnée par la loi.
3. Personne d'autre que moi n'a le droit de faire valoir ce travail, en totalité ou en partie, comme le sien.
4. Les propos repris mot à mot à d'autres auteurs figurent entre guillemets (citations).
5. Les écrits sur lesquels je m'appuie dans ce mémoire sont systématiquement référencés selon un système de renvoi bibliographique clair et précis.

PRENOM : ANETA.....

NOM : MARZEC.....

DATE : 27/08/2020.....



.....	1
Remerciements.....	3
Introduction.....	7
Partie 1 - Cadre théorique .....	10
<b>CHAPITRE 1. LA LINGUISTIQUE DE CORPUS .....</b>	<b>11</b>
1. LA NOTION DE LINGUISTIQUE DE CORPUS .....	12
2. LA NOTION DE CORPUS .....	13
3. LE TRAITEMENT AUTOMATIQUE DES LANGUES (TAL) .....	15
<b>CHAPITRE 2. LA PHRASEOLOGIE .....</b>	<b>18</b>
1. LE DOMAINE DE LA PHRASEOLOGIE.....	18
2. LA COLLOCATION .....	19
3. LA COLLOCATION GRAMMATICALE ET LA COLLIGATION .....	23
<b>CHAPITRE 3. LA QUESTION DU GENRE TEXTUEL .....</b>	<b>26</b>
1. LA NOTION DE GENRE ET LE GENRE ROMANESQUE .....	26
2. LES SOUS-GENRES TEXTUELS.....	28
Partie 2 - Méthodologie .....	35
<b>CHAPITRE 4. LE LEXICOSCOPE .....</b>	<b>36</b>
1. PRESENTATION DE L'OUTIL ET DU CORPUS .....	36
2. CONCORDANCE ET COOCCURRENCE.....	38
3. L'ARBRE LEXICO-SYNTAXIQUE RECURRENT (ALR) .....	41
<b>CHAPITRE 5. METHODOLOGIE DU TRAITEMENT DE DONNEES .....</b>	<b>45</b>
1. L'ETUDE DE LA COMBINATOIRE LEXICO-SYNTAXIQUE .....	45
2. LE CHOIX DE PIVOTS .....	46
Partie 3 - Analyse .....	47
<b>CHAPITRE 7. LA COMBINATOIRE DES NOMS DE BOISSON : APPROCHE QUANTITATIVE .....</b>	<b>48</b>
1. LES COLLOCATIFS DES NOMS DE BOISSON .....	48
2. ÉTUDE DES COLLOCATIFS PAR SOUS-GENRE ROMANESQUE.....	56
<b>CHAPITRE 8. LA COMBINATOIRE DES NOMS DE BOISSONS : APPROCHE QUALITATIVE (BOIRE DU VIN ET PRENDRE UNE BIÈRE) .....</b>	<b>62</b>
1. L'ANALYSE DE LA COLLOCATION BOIRE DU VIN .....	62
2. L'ANALYSE DE LA COLLOCATION PRENDRE UNE BIÈRE .....	78
Conclusion.....	85
Bibliographie .....	87
Sigles et abréviations utilisés .....	95
Table des annexes.....	97
Table des matières.....	118



## Introduction

Le présent mémoire s'inscrit dans la continuité des travaux effectués par le laboratoire LIDILEM (Laboratoire de Linguistique et Didactique des Langues Etrangères et Maternelles) de l'Université Grenoble-Alpes sur la phraséologie spécifique aux romans contemporains français, anglais et allemands à travers des constructions lexico-syntaxiques récurrentes (CLS), dans le cadre du projet PhraseoRom<sup>1</sup>. Le corpus d'interrogation est constitué de textes littéraires français contemporains et partitionné aux plusieurs sous-corpus. Notre recherche est basée sur les données issues de corpus de romans policiers, fantastiques, de science-fiction, historiques, sentimental et de littérature générale ou blanche.

L'objectif de ce travail sera d'analyser les unités phraséologiques contenant les noms des boissons (*champagne, vin, bière, vodka, café et thé*) apparaissant dans le sous-corpus de différents genres romanesques. Le but de ce travail de recherche est l'étude des expressions spécifiques au sein des domaines de la linguistique de corpus et de la stylistique de genre. Le point clé de notre analyse est l'introduction de la phraséologie à la recherche sur la caractérisation et la distinction des genres littéraires. La question de notre recherche est de savoir si les phraséologismes autour de la boisson permettent d'améliorer la caractérisation de sous-genres romanesques et de les distinguer entre eux.

Nos travaux s'inscrivent dans la continuité des recherches antérieures menées par des chercheurs du laboratoire LIDILEM dans cette tendance sur les phraséologismes dans les sous-corpus romanesques. Nous approfondirons dans notre travail les recherches de Gonon & Sorba (2019) sur les phraséologismes autour du vin et du champagne dans le roman contemporain français. Notre motivation pour étudier les propriétés syntaxiques des collocations est liée à la nouveauté de cette approche. Ce mémoire a pour objectif d'observer les constructions syntaxiques des collocations autour de noms de boissons afin de comprendre leur fonctionnement dans les sous-corpus et d'essayer d'indiquer si ces expressions nous permettent de distinguer les genres romanesques entre eux.

Selon Montémont (2009 : 1) « l'alcool et les alcools sont des thèmes étroitement associés au monde littéraire ». Nous avons choisi les noms de boissons comme le coeur de notre recherche parce que ce sujet présente un grand potentiel de recherche sur les phraséologismes dans le roman contemporain. C'est particulièrement important en raison de la nature et de la fonction des noms des boissons dans le texte : « les modes de consommation

---

<sup>1</sup> Voir le site : <https://phraseorom.univ-grenoble-alpes.fr/>

et le regard que l'on porte sur l'alcool dans la littérature sont liés à des pratiques sociales, des habitudes culturelles » (Montémont, 2009: 1). L'alcool, qui apparaît dans les romans sous la forme de divers pivots, fournit un arrière-plan intéressant pour la recherche sur la phraséologie des romans. L'acte du boire est souvent associé à un moment de transition dans l'histoire, permettant au personnage de se déplacer et au lecteur de passer d'une action à une autre dans la narration (Grossmann et al., 2019). L'acte du boire, selon Grossmann et al. (2019), contribue à l'organisation du récit et sert à créer une ambiance ou une atmosphère dans un texte littéraire. Pour cette raison, nous constatons que le sujet de la consommation d'alcool peut être un domaine de recherche intéressant dans la littérature. La recherche sur les constructions statistiquement significatives des noms de boissons peut être un moyen de distinguer et de caractériser différents genres romanesques.

Pour ce qui est de la méthodologie de notre recherche, grâce à l'outil d'extraction des unités polylexicales récurrentes dans différents sous-corpus, le *Lexicoscope*, nous pouvons étudier les profils combinatoires des expressions polylexicales sur le plan linguistique. Notre analyse sera fondée sur six sous-corpus romanesques du *Lexicoscope* qui nous permettront d'extraire les ARL à partir de nos pivots (noms de boissons) et d'étudier des expressions spécifiques pour les sous-genres.

Dans la première partie de notre recherche, nous allons présenter le cadre théorique et nous aborderons quelques notions importantes pour notre étude. Premièrement, nous allons montrer l'intérêt de la linguistique de corpus en rapprochant en même temps la notion de *corpus* et *Traitement Automatique des Langues*. Ensuite, nous nous focalisons sur les notions de collocation et colligation. Enfin, nous nous plongerons dans la notion du genre textuel en présentant les sous-genres de notre corpus.

Puis, dans la deuxième partie, nous allons montrer la méthodologie de notre travail. Nous allons présenter l'outil d'extraction des données, le *Lexicoscope*, avec notre corpus de recherche. Cette partie servira à éclaircir le concept de *concordance*, *cooccurrence* et d'*arbre lexico-syntaxique* (ALR). Ensuite, nous traiterons de l'intérêt de la combinatoire lexico-syntaxique et nous expliquerons notre choix de pivots.

La troisième partie sera consacrée à l'analyse des résultats obtenus. Nous allons examiner dans une approche quantitative la combinatoire des noms de boissons en indiquant le nombre d'occurrences de chaque expression dans chaque sous-corpus. Ensuite, nous allons aborder l'aspect qualitatif- nous allons analyser deux expressions- *boire du vin* dans le corpus du roman général en comparant avec les autres sous-corpus, et l'expression *prendre une*

*bière* dans le sous-corpus de romans policiers. Nous essayerons d'analyser leur fonctionnement linguistique, leurs variations et leur entourage.

## **Partie 1**

-

## **Cadre théorique**

Dans la première partie de notre travail, nous allons présenter le cadre théorique de nos recherches. Nous définirons les notions sur lesquelles s'appuie notre travail. Dans un premier temps, nous allons expliquer l'intérêt de la linguistique de corpus, définir la notion de corpus et montrer les objectifs du Traitement Automatique de Langues (TAL). Dans le chapitre suivant, nous allons explorer le domaine de la phraséologie, particulièrement les notions de collocation et de colligation. Enfin, nous nous focaliserons sur la notion de genre textuel et présenterons les traits caractéristiques des sous-genres romanesques choisis. Ce cadre théorique nous permettra de mener notre recherche sur les phraséologismes dans un corpus romanesque français contemporain.

## **Chapitre 1. La linguistique de corpus**

La linguistique de corpus est devenue, au cours des dernières années, un domaine-clé en termes d'analyse et d'exploitation d'informations, généralement sous forme textuelle grâce à des outils informatiques. Les vastes corpus électroniques permettent de récupérer des exemples attestés d'un phénomène donné, de mener des observations en prenant en compte des faits non accessibles manuellement (Eshkol-Taravella & Lefevre-Halftermeyer, 2017). Les nouvelles approches dans la linguistique au cours du dernier demi-siècle se sont développées grâce aux contacts avec l'informatique et les mathématiques.

Une linguistique formelle s'est développée sur la base d'une linguistique descriptive. La linguistique formelle prend en considération non seulement les aspects lexicaux, mais aussi syntaxiques ou sémantiques. La linguistique formelle propose des modèles s'appuyant sur une double validation-explicative dans un contexte linguistique et opératoire d'un point de vue informatique (Pierre, 2005). Du point de vue explicatif, la linguistique formelle couvre aussi bien les aspects lexicaux que syntaxiques ou sémantiques, en permettant d'aller au-delà de l'accumulation de faits de langue et de confronter les théories à l'usage effectif de la langue. Les progrès des outils informatiques, par exemple dans le champ de stockage ou de puissance de calcul, permettent d'évaluer des modèles opératoires dans un corpus (Habert et al., 1997).

Nous pouvons observer un développement accéléré des approches statistiques et probabilistes sur corpus au cours des années 1990. Les raisons principales de cet engouement sont la disponibilité de ressources textuelles électroniques de grandes tailles (corpus, bases de données textuelles, dictionnaires et lexiques) et les progrès de l'informatique, tant en matière de stockage que de puissance de calcul (Habert et al., 1997). Un nouveau champ de

recherches, la linguistique de corpus, permet au linguiste de travailler sur un grand volume de faits de langue et de confronter les hypothèses à l'usage effective de la langue. Les besoins applicatifs de la linguistique de corpus ont conduit à de nombreux travaux en TAL (Traitement Automatique de la Langue).

Dans les sous-parties de ce chapitre, nous allons traiter les notions de *linguistique de corpus*, de *corpus* et de *Traitement Automatique des Langues*.

### **1. La notion de linguistique de corpus**

Le terme *Corpus Linguistics*, utilisé dans la linguistique contextualiste anglo-saxonne, désigne le projet de recherche de Sinclair (1996) ou de Leech (1992), notamment. Leech emploie aussi le terme de *Computational Corpus Linguistics (CCL)* après avoir remarqué les liens particulièrement inséparables entre *Corpus Linguistics* et usage des ordinateurs. Selon Leech, la *Computer Corpus Linguistics* est « a new research enterprise, and in fact a new philosophical approach to the subject » (Leech, 1992 : 106).

L'usage du terme *linguistique de corpus* apparaît en France, dans les travaux en lien avec la tradition britannique notamment ceux de Habert et al. (1997), de Bilger (2000) et de Williams (2005). Néanmoins, même s'il est assez difficile de trouver chez eux une définition précise du terme de *linguistique de corpus*, Williams en propose la définition suivante :

La linguistique de corpus est une discipline qui s'est développée à partir des années dix-neuf cent quatre-vingts mais qui a vraiment pris son essor avec l'arrivée sur le marché d'ordinateurs personnels. Largement issue de l'école contextualiste, la discipline s'intéresse à la langue en contexte sous la forme de grands ensembles de textes, les corpus. C'est une discipline qui est très liée à l'utilisation de l'informatique, mais qui reste une discipline des sciences humaines et non de l'informatique. Les maîtres mots sont linguistique et corpus. (Williams et al., 2005 : 13).

Rastier complète cette définition en ajoutant que :

Au milieu des années 1990, la linguistique de corpus, issue des humanités (et notamment de la collectivité réunie autour de la revue *Computers and the Humanities*), l'a emporté sur le courant orthodoxe néochomskien de la linguistique computationnelle. (Rastier, 2005: 41)

Jacques (2005 : 21) essaie de définir la linguistique de corpus en opposition à une autre linguistique, qu'elle appelle « la linguistique introspective ». Cette deuxième linguistique se base sur la capacité qu'a tout locuteur de dire si un énoncé est bien ou mal formé selon le système de sa langue. Néanmoins, l'introspection ne suffit pas toujours et les jugements d'acceptabilité peuvent varier. Comme indique Jacques (2005), la fiabilité limitée des jugements de grammaticalité et l'impuissance de la linguistique introspective à capter de façon satisfaisante la variation sont les principales raisons de critiquer la linguistique



introspective. En opposition à cette linguistique, la linguistique de corpus met l'accent sur des données authentiques et observables. La linguistique de corpus se caractérise par une approche empirique qui souligne l'importance de l'expérimentation sur des données.

Le terme *linguistique de corpus* se compose de deux éléments essentiels : la *linguistique* et le *corpus*. Dans la section suivante, nous allons définir la notion de corpus.

## 2. *La notion de corpus*

Le *corpus* en tant qu'objet fabriqué selon une méthode et une théorie établies au préalable afin d'orienter une recherche déterminée est une notion qui date des années 1950 sur la base de travaux de John R. Firth et de la *London School*. Celui-ci est au départ de la « linguistique de corpus », qui considère que le sens d'un mot se situe dans ses usages en contexte et n'est pas un donné a priori.

On peut distinguer deux approches principales. La première, *corpus-based*, utilise le corpus pour une vérification d'une théorie ou une hypothèse afin de la valider ou l'améliorer. La deuxième, *corpus-driven*, est plus inductive et elle revendique des recherches sans a priori faisant émerger du corpus des observables. Nous pouvons considérer que le sens d'un mot est trouvé à l'échelle du corpus. D'après Mayaffre (2005), nous pouvons décrire le corpus comme « la forme maximale du contexte ».

Le corpus a plusieurs définitions, mais celle de Sinclair est utilisée le plus souvent : un corpus est « une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon de langage » (Sinclair, 1996 : 4). À partir de cette définition, nous pouvons remarquer qu'un ensemble de données collectées ici et là sans réflexion sur la nature des documents n'est pas un corpus. Des critères linguistiques explicites servent à créer un ensemble de données homogènes qui appartiennent à un genre particulier. Nous pouvons déduire que pendant la constitution du corpus, les données ne sont pas sélectionnées accidentellement. Il faut précisément définir des principes et des critères de sélection. Dans l'étape finale d'élaboration du corpus, nous devons indiquer les types et les genres des textes à choisir et aussi la taille de corpus. Gibbon (1998) propose une définition complémentaire du corpus :

A corpus is any collection of speech recordings which is accessible in computer readable form and which comes with annotation and documentation sufficient to allow re-use of the data in-house, or by people in others organisations (Gibbon et al., 1998, cité par Delais-Roussarie & Durand, 2003 : 93).

Gibbon met l'accent sur la disponibilité de données sous format électronique. Gibbon, comme Sinclair, souligne que les données du corpus doivent être correctement organisées et annotées pour un usage efficace.

Une autre définition de *corpus* est proposée par le linguiste britannique Baker : « A corpus now means primarily a collection of texts held in machine readable form and capable of being analyzed automatically or semi-automatically in a variety of ways » (Baker, 1995: 225). Baker note, comme Gibbon, que le corpus est composé de textes sous une forme lisible par les machines. Un point important est la possibilité d'analyser automatiquement les données collectées dans le corpus. Les annotations faites sur le corpus permettent ce processus.

Même si le mot *corpus* est utilisé en singulier, il désigne une réalité plurielle. On peut distinguer différents types de corpus en linguistique. De manière hiérarchique et d'après Bommier-Pincemin (1999a, 1999b), on peut distinguer trois grands niveaux de corpus :

1) *les corpus lexicographiques ou sacs de mots* - ils visent à l'exhaustivité complète. Il n'existe en effet plus de difficultés techniques à recueillir et à traiter l'ensemble des données. Les principaux avantages de ces corpus sont la portabilité et la rapidité de consultation supérieures à celles des dictionnaires papier. Les corpus lexicographiques donnent la possibilité de stocker une quantité considérable d'information dans un volume assez réduit. (Béjoint, 2007)

2) *les corpus phrastiques de grammairiens ou de syntacticiens* - une de leurs particularités est de pouvoir recueillir des exemples non pas attestés, mais forgés.

3) *les corpus textuels*- ils ne peuvent aspirer ni à l'exhaustivité ni même à la représentativité, ils concernent toujours des données attestées. La représentativité du corpus est définie par les métadonnées du corpus, soit la documentation du corpus. Leech (2007) indique que aucun corpus n'est par acclamation reconnu comme représentatif d'une langue. La principale raison de cela sont les méthodologies employées pour représenter la langue dans son ensemble qui ne font pas encore l'objet du plein consentement.

La linguistique de corpus dans le sens général repose sur l'affirmation forte que le texte constitue l'objet du linguiste. En effet, de nombreux linguistes contemporains constatent que l'objet accompli de la linguistique n'est ni le signe ni la phrase « artefact des grammairiens » (Rastier, 2005) mais que le sens trouve sa source dans un texte. Le texte doit donc être considéré comme l'unité fondamentale d'une linguistique aboutie : « le texte est pour une linguistique évoluée l'unité minimale, et le corpus l'ensemble dans lequel cette unité prend son sens » (Rastier, 2005: 31). Rastier traite les textes comme les seuls objets empiriques. Le

texte, selon Rastier, doit trouver sa place dans le corpus pour être interprété dans un plus vaste contexte. La contextualisation joue un rôle important dans le processus de l'émergence d'un sens construit à partir d'une analyse d'ensemble de corpus.

Mayaffre aussi met l'accent sur cet aspect en définissant le corpus comme « le lieu linguistique où se construit et s'appréhende le sens des textes » (Mayaffre, 2005:12). Rastier fait aussi attention à l'*intertextualité*, c'est-à-dire la relation d'intégration et de transformation entre les textes dans le corpus. En plaçant un texte dans un corpus, nous allons analyser son sens dans une nouvelle perspective qui est plus vaste. Cette approche nous permet de trouver un sens non pas donné mais repéré sur la base de l'ensemble de textes : « tout texte placé dans un corpus en reçoit des déterminations sémantiques, et modifie potentiellement le sens de chacun des textes qui le composent. » (Rastier, 2001:92).

Adam (2001) souligne également le fait que le sens est né dans le texte. En allant plus loin dans la réflexion, nous remarquerons que pour la linguistique des grandes unités, c'est le contexte qui garde le sens : « Si, comme le dit Saussure, *la langue n'est créée qu'en vue du discours*, la linguistique a non seulement pour objet empirique mais pour objet théorique cette unité de communication-interaction langagière qu'on appelle un TEXTE (ou un DISCOURS...) » (Adam, 2001: 216),

Le corpus de notre recherche est un corpus textuel qui contient des romans contemporains français publiés après 1950 et classés aux six sous-genres : romans policiers, fantastiques, de science-fiction, historiques, sentimental et de littérature générale ou blanche. Nous le présenterons en détail dans la troisième section du présent chapitre.

### ***3. Le Traitement Automatique des Langues (TAL)***

Comme nous l'avons vu, la linguistique de corpus peut être utilisée pour révéler les régularités linguistiques dans des langues à travers des corpus informatisés. Le développement de la linguistique de corpus a sa source dans l'essor de l'informatique et plus particulièrement du TAL ou Traitement Automatique des Langues.

Le TAL a pour objectif principal d'examiner les corpus du point de vue statistique. Il s'agit de traiter de grandes quantités de données textuelles sur des supports électroniques (Condamines, 2005). Le TAL mène une analyse de données textuelles à l'aide d'outils informatiques. Ces outils ont pour but d'analyser des données textuelles grâce aux méthodes statistiques. D'après ces remarques, nous pouvons voir que les corpus sont devenus le principal matériau du TAL.

L'analyse de corpus a de nombreux objectifs. Parmi les plus importants, nous pouvons mentionner les analyses morphologiques, syntaxiques et sémantiques pour approfondir les connaissances sur la langue et améliorer l'extraction d'informations ou la recherche d'informations sur un sujet donné (Condamines, 2005). Pierrel (2005) utilise des caractéristiques similaires à propos de la matière du corpus. Il remarque qu'aujourd'hui, traitement automatique des langues et linguistique de corpus structurent un nouveau champ disciplinaire aux finalités multiples (Pierrel, 2005: 2-3):

- modélisation de la langue, de sa structure et de son usage conduisant la linguistique à des exigences d'opérationnalité effective sur les formes d'usage de la langue, par opposition aux exemples construits, encore trop souvent utilisés en linguistique ;
- mise en place d'applications concrètes : indexation et accès à l'information, résumé de textes, extraction de connaissances, dialogue homme-machine, par exemple .

Le traitement des corpus textuels et les corpus d'exemples permettent de construire des modèles dans la langue (Pierrel, 2005). La recherche sur les corpus peut également contribuer à construire des bases de connaissances sur un domaine donné et sa terminologie spécifique (Condamines, 2005). Lorsque l'on parle de corpus, leur fiabilité et leur disponibilité doivent être prises en compte (Pierrel, 2005). La qualité d'un corpus doit être indiscutable. Pour construire et valider des modèles pertinents, il faut travailler sur un corpus propre c'est-à-dire sans erreur. L'homogénéité de références dans le corpus est un autre facteur très important. Pour obtenir l'homogénéité de références, le corpus devra consister en textes de mêmes formes, genres et niveaux de langue qui obéissent à des critères de choix très précis. L'assurance de la pérennité de la disponibilité d'un corpus permet de maintenir l'objectivité de recherches menées sur ce corpus. La constitution du corpus sur lequel nous travaillons s'est conformée à ces exigences (projet PhraseoRom).

Les études et les recherches en linguistique de corpus et en TAL exigent l'usage de vastes ressources linguistiques (corpus, si possible annotés et dictionnaires). Les outils de gestion et d'analyse de ces ressources sont aussi indispensables. Les champs de la linguistique de corpus et du TAL portent l'enjeu essentiel de travailler à une meilleure connaissance et à une meilleure modélisation de la langue. Le développement dans ce domaine nous permet de mener une exploitation en profondeur du contenu des corpus.

Pour mener notre recherche et extraire les données de notre corpus, nous utiliserons un outil issu du TAL – le Lexicoscope – que nous présenterons dans la partie 2 de ce mémoire.

Pour résumer, nous pouvons remarquer que les corpus numérisés et les outils lexicométriques ont profondément révolutionné la recherche linguistique, en élargissant les possibilités d'observations du point de vue des domaines comme la lexicologie, la syntaxe ou la sémantique. C'est dans ce cadre théorique profondément renouvelé que nous inscrivons notre recherche sur le lexique. Dans le chapitre suivant, nous allons traiter de la phraséologie et de la notion de « collocation » en approfondissant cet aspect de la lexicologie.

## Chapitre 2. La phraséologie

La phraséologie est un domaine de la linguistique qui traite des unités lexicales préconstruites. Dans la première section, nous fournissons des éléments de définition de la phraséologie. Ensuite, dans la deuxième section de ce chapitre, nous présenterons l'une de ces unités préconstruites, la « collocation ». Enfin, nous aborderons la « colligation », une autre de ces unités. Nous mènerons notre recherche sur les constructions adjectivales, nominales et verbales qui se composent de deux ou trois mots et ces deux types d'unités préconstruites nous semblent susceptibles de répondre à notre question de recherche.

### 1. *Le domaine de la phraséologie*

La phraséologie peut être définie comme le domaine qui traite les séquences lexicales perçues comme préconstruites. Les unités préconstruites sont des suites préfabriquées qui intègrent à la fois lexicale et grammaticale, comme par exemple *bien que*, *pomme de terre* ou *tout se passe bien*. Les unités préconstruites sont usuellement traitées comme les exceptions auxquelles nous ne pouvons pas appliquer les règles habituelles donc de façon spécifique. González Rey et Fernández González soulignent aussi que dans le processus d'identification des unités phraséologiques leur « polylexicalité, figement, répétition et institutionnalisation » sont les caractéristiques les plus importantes. (González Rey & Fernández González, 2006 : 1017). Depuis que la phraséologie a commencé à occuper une place significative dans la linguistique, nous pouvons observer un nouveau développement et un élargissement de ces objets d'étude. Traditionnellement, la phraséologie évoque une approche syntaxico-sémantique et combinatoire du lexique<sup>2</sup>.

Les expressions figées suscitaient peu l'intérêt des linguistes parce qu'elles étaient synonymes de folklore, classées seulement dans le registre familier ou populaire. Néanmoins, González-Rey (2002) indique que les expressions phraséologiques n'apparaissent pas seulement dans les registres familier, vulgaire ou argotique, mais également dans les autres registres de la langue. Comme exemples elle donne les expressions décrivant le rire très fort dans les différents registres : *rire à gorge déployée* pour le registre

---

<sup>2</sup> Selon Blumenthal (2007), le *profil combinatoire* d'un mot désigne le comportement d'un mot de base avec l'ensemble de ses collocatifs.

soutenu, *rire aux larmes* pour le registre courant, *rire comme un bossu* pour le registre familier ou *se fendre la poire* pour le registre vulgaire. González Rey souligne que la phraséologie « doit être étudiée non comme un domaine à part, mais comme un ensemble porteur d'idiosyncrasies d'une culture, d'une société, d'une façon collective de voir les choses, d'une façon idiomatique de parler » (González Rey, 2002 : 35).

Les recherches actuelles en phraséologie se concentrent sur le « principe phraséologique du langage » (Legallois & Tutin, 2013) et incluent des phénomènes tels que les collocations, les schémas syntaxiques ou textuels. Nous pouvons observer l'élargissement de ce champ de recherches.

Selon González Rey (2008 : 5), « placée à mi-chemin entre le lexique et la syntaxe, la phraséologie s'occupe des signes polylexicaux composés d'au moins deux mots (parfois même d'un seul si son emploi est idiomatique), stables, répétés et souvent figurés ». Cette notion de phraséologie varie selon les linguistes, comme le souligne Lengert (Lengert, 2001 : 802), qui note également que « dans le cadre d'une conception restreinte, l'objet d'étude de cette discipline est susceptible d'interprétations différentes » (Lengert, 2001 : 803).

Pour notre part, nous allons inscrire nos recherches dans le cadre de la *phraséologie étendue*, la phraséologie incluant désormais des objets d'étude divers, allant des collocations aux séquences discursives (Legallois & Tutin, 2013). Les objets de la phraséologie étaient auparavant considérés comme des anomalies ou des exceptions. Actuellement, ils deviennent des éléments essentiels dans les modèles linguistiques où la notion de principe phraséologique de la langue se développe.

Un des problèmes rencontrés actuellement est que l'on trouve des définitions assez variées pour désigner les mêmes phénomènes. Nous pouvons parler d'« une trop grande hétérogénéité des termes renvoyant à l'objet d'étude de cette discipline » (Zaharieva & Kaldieva-Zaharieva, 2017: 16). Nous devons prendre en considération l'absence de consensus terminologique pour désigner les unités phraséologiques. Il nous faut donc préciser dès maintenant le cadre théorique et terminologique que nous utiliserons dans notre étude et définir plus précisément l'une des unités phraséologiques au cœur de notre étude : la collocation.

## **2. La collocation**

La phraséologie est devenue un centre d'intérêt pour les linguistes au début du XX<sup>e</sup> siècle. En effet, Charles Bally est considéré comme le précurseur qui a proposé une première

typologie des phénomènes phraséologiques. En 1909, il a distingué deux types de *locutions phraséologiques* :

1) *des unités phraséologiques* – ces *unités* sont indécomposables et se caractérisent par leur non-compositionnalité sémantique (ex. *cordon bleu, broyer du noir*)

2) *des séries phraséologiques* - chaque constituant conserve son autonomie et elles postulent des « affinités » entre les constituants ;

Il y a série ou groupement usuel lorsque les éléments du groupe conservent leur autonomie, tout en laissant voir une affinité évidente qui les rapproche, de sorte que l'ensemble présente des contours arrêtés et donne l'impression du « déjà vu » . (Bally, 1909 : 70)

Bally remarque l'existence des groupes intermédiaires de figement entre les expressions mémorisées et les combinaisons libres (Tutin & Grossmann, 2002) dont les prototypes sont les *séries d'intensité (diamétralement opposé)* et les *périphrases verbales (prendre une décision)*.

Entre les extrêmes (groupements passagers et unités indécomposables) se placent des groupes intermédiaires *appelés séries phraséologiques* (p.ex. les séries d'intensité ou les périphrases verbales). (Bally, 1909)

Sur la base des exemples de ces *séries phraséologiques* comme *grièvement blessé* ou *gravement malade*, il montre que les adverbes ne sont pas facilement interchangeables. Ces éléments de définition des séries phraséologiques nous permettent de passer à la notion de collocation.

La notion de *collocation* est généralement définie comme une « cooccurrence lexicale » et naît dans le terreau de la didactique du lexique (Palmer, 1933, cité par Tutin, 2013). Dans les années 50, Firth, le linguiste de la tradition contextualiste anglaise, a défini la collocation comme une attraction mutuelle entre des mots sans prendre en compte l'ordre des mots ni leur proximité <sup>3</sup>. Cette conception de la collocation, introduite par Firth en 1957, a été reprise par d'autres linguistes. Dans leur approche fonctionnelle, Halliday & Hasan (1976) constatent que si une paire d'éléments a tendance à partager le même environnement (par exemple : *rire - blague ; malade – médecin ; juge - tribunal*), des collocations contribuent à la cohésion du texte. Généralement, quand deux éléments lexicaux ont tendance à apparaître dans les contextes similaires, ils généreront une force cohésive :

---

<sup>3</sup> « Collocations of a given word are statements of the habitual or customary places of that word in collocational order but not in other contextual order and emphatically not in any grammatical order. The collocation of a word or a 'piece' is not to be regarded as mere juxtaposition, it is an order of mutual expectancy » (Firth, 1957: 12).



*laugh ... joke, blade ... sharp, ill... doctor [...]* The cohesive effect of such pairs depends not so much on any systematic relationship as on their tendency to share the same lexical environment, to occur in COLLOCATION with one another. In general, any two lexical items having similar patterns of collocation – that is, tending to appear in similar contexts – will generate a cohesive force if they occur in adjacent sentences. (Halliday & Hasan, 1976: 285-286).

Plus tard, Sinclair (1991), un autre linguiste de la tradition contextualiste britannique, inscrit la notion de collocation dans le cadre de la linguistique de corpus. Il défend l'idée du *principe idiomatique* selon laquelle les productions langagières des locuteurs sont constituées d'une série de choix préconstruits, qui intègrent à la fois lexicale et grammaticale : « Pour expliquer l'existence de formulations nouvelles, Sinclair postule *un choix ouvert*, une dimension dans laquelle chaque choix lexical mène à une série restreinte de nouvelles constructions grammaticales » (Gledhill & Frath, 2007: 68). Il développe d'une part l'idée que les éléments linguistiques ont tendance à apparaître de manière irrégulière dans l'usage et d'autre part l'idée d'une combinatoire libre des unités linguistiques :

This is a way of seeing language text as the result of a very large number of complex choices. At each point where a unit is completed (a word, a phrase or clause), a large range of choice opens up and the only constraint is grammaticality. (Sinclair, 1991: 109).

Sinclair indique que ces deux principes se présentent en parallèle, mais ils sont contradictoires :

For normal texts we can put forward the proposal that the first mode to be applied is the idiom principle [...] The open-choice analysis could be imagined as an analytical process which goes on in principle all the time, but whose results are only intermittently called for. (Sinclair, 1991 : 114).

Toutefois, en prenant en considération la perspective lexico-grammaticale, les contraintes de *grammaticalité* sur le choix ouvert et celles qui régissent le principe idiomatique sont identiques. Nous pouvons donc constater l'absence d'un choix véritablement ouvert et que chaque mot entre dans des relations lexico-grammaticales (Gledhill & Frath, 2007).

En utilisant le critère de *fréquence collocationnelle* (Sinclair, 1991), la phraséologie statistique examine des unités polylexicales même les moins contraintes que nous pouvons repérer à partir de leur fréquence de cooccurrences. Dans cette perspective, la fréquence est considérée comme un indice caractéristique du figement langagier dans le contexte des relations syntaxiques et des affinités catégorielles qu'entretiennent les termes (Bolly, 2010). La collocation peut être alors définie comme une cooccurrence qui est statistiquement significative et dont la co-apparition des termes n'est pas aléatoire.

Dans la tradition continentale, les linguistes adoptent une conception de la collocation lexicale, décrite comme une expression binaire (Hausman, 1989) Heid, 1994 ; Mel'čuk, 1988, cité par Tutin, 2013). Cette conception repose sur deux principes : le premier, c'est une **binarité catégorielle** qui joint deux types de constituants, qui sont généralement des unités lexicales ; le deuxième, c'est une **binarité fonctionnelle** qui indique un fonctionnement dissymétrique, antagonique entre les deux éléments (Tutin, 2013). Selon Tutin, qui reformule le critère de binarité catégorielle, « La collocation met en jeu une unité lexicale simple ou complexe en relation syntaxique avec un constituant linguistique » (Tutin, 2013 : 50).

Le statut des deux constituants combinés est différent (Hausman, 1989:1001, cité par Tutin, 2013). En effet, *la base* garde son sens habituel pendant que *le collocatif* est « sémiotiquement dépendant et choisi en fonction de la base » (Tutin, 2013). Le contexte module le sens du collocatif (*peur bleue* vs. *\*crainte bleue*). Dans les collocations, le deuxième élément de la structure, le *collocatif*, est le plus idiomatique et le plus inattendu. On retrouve dans cette expression compositionnelle une restriction sémantique ou lexicale sur le collocatif (Tutin, 2013). De plus, cette dissymétrie « est liée à la structure sémantique de la collocation qui s'analyse comme l'association d'un prédicat, le collocatif, associé à un argument plus stable sur le plan lexical et sémantique, la base » (Tutin, 2013: 50).

Tutin & Grossmann (2002) définissent formellement la notion de collocation sur la base des études théoriques (Cruse, 1986; Hausman, 1989; Mel'čuk, 1984) et proposent cinq critères pour la fonder (Tutin & Grossmann, 2002: 3-4) :

- L'aspect arbitraire (la non-prédictibilité de la collocation) — *appétit d'ogre* ou *faim de loup* (vs. *\*appétit de loup* et *\*faim d'ogre*)
- La transparence et le non-figement sémantiques de la collocation — l'interprétation de la collocation par le locuteur est possible même si la forme est non prévisible. Néanmoins, le locuteur non-natif est incapable de les produire : *célibataire endurci* ou *feuilleter un livre*.
- Le caractère binaire de la collocation — la collocation est composée de deux mots ou lexies (*avoir peur*, *peur bleue*).
- La dissymétrie des composants de la collocation (un élément conserve son sens habituel, l'autre non) : *la base* existe dans son sens habituel pendant que le *collocatif* est dépendant du premier élément (*peur bleue*).

- La notion de cooccurrence restreinte, de sélection lexicale — la sélection du *collocatif* est imposée par la *base*, la cooccurrence n'est pas libre.

Selon Tutin & Grossmann (2002), trois de ces critères (le caractère binaire, la dissymétrie et la notion de sélection lexicale) définissent la collocation de la façon la plus précise :

Une collocation est l'association d'une lexie (mot simple ou phrasème L et d'un constituant C (généralement une lexie, mais parfois un syntagme par exemple à *couper au couteau* dans *un brouillard à couper au couteau*) entretenant une relation syntaxique telle que :

- C (le collocatif) est sélectionné en production pour exprimer un sens donné en cooccurrence avec L (la base) ;
- Le sens de L est habituel (Tutin & Grossmann, 2002:5).

Tutin & Grossmann présentent aussi une classification sémantique des collocations prenant en compte l'opacité du collocatif. Ainsi, ils distinguent :

- Les collocations opaques dans lesquelles le sens du collocatif est arbitraire et non transparent (*peur bleue, colère noire*).
- Les collocations transparentes qui sont difficilement prédictibles sur le plan lexical, mais elles sont interprétables en compréhension (*triste à pleurer, bête à manger du foin*).
- Les collocations régulières — transparentes et simples de prédire (*gros mangeur, tristesse affreuse*).

Nous pouvons situer les collocations à la frontière de la combinatoire libre et de la combinatoire figée. Nous distinguons en linguistique deux approches vers les collocations : l'approche *quantitative* et l'approche *qualitative* (Hausmann & Blumenthal, 2006). La première se fonde sur la description statistique de corpus, il s'agit d'un repérage des cooccurrences statistiquement fréquentes. En ce qui concerne l'approche qualitative, nous pouvons entendre la collocation comme une association lexicale syntagmatique restreinte entre deux éléments entretenant une relation syntaxique. Ce sont ces deux approches que nous allons combiner dans nos analyses.

### ***3. La collocation grammaticale et la colligation***

Il existe aussi une autre catégorie de collocations - **les collocations grammaticales**. La collocation lexicale, présentée plus haut, est une construction lexicale où un verbe, un nom, un adjectif, ou un adverbe existe dans une relation qui est tout à fait prédictible avec

un autre mot. La collocation grammaticale, quant à elle, est un type de construction dans laquelle un verbe ou un adjectif est suivi d'une préposition particulière ou bien une construction dans laquelle un nom est suivi d'une forme verbale particulière (*spécifique à, dépendre de, un argument contre, différent de*). Selon C. Bolly (2010), elle se définit comme :

une combinaison lexicalement contrainte constituée d'un mot lexical (de contenu) qui sélectionne de manière arbitraire un mot grammatical (mot-outil), généralement une préposition, ou une structure syntaxique, par exemple une proposition infinitive (Bolly, 2010: 16-17).

La collocation grammaticale est aussi utilisée comme le synonyme de la notion de colligation. Cette notion a également été introduite dans la linguistique contextualiste britannique en 1957 par Firth. Même si ces deux notions sont utilisées comme des synonymes, nous devons rendre compte qu'il existe des différences entre elles. Legallois (2012) définit la colligation comme « un phénomène d'association entre un mot lexical ou grammatical et une catégorie grammaticale (partie du discours, fonction syntaxique, marqueurs aspectuels, modaux, temporels, marqueurs grammaticaux des catégories de la négation, de la propriété, etc.) » (Legallois, 2012: 39).

Selon Francis (Francis, 1991), la colligation est un système dans lequel toute unité lexicale possède sa propre grammaire. Il a mené les recherches sur les quelques noms fréquents en anglais prenant en compte leur comportement. Sur la base des résultats obtenus, Francis détermine la fréquence des fonctions des syntagmes dont les noms sont la tête (Legallois, 2012). Il indique qu'il existe des mots dont le comportement syntaxique peut différer selon les acceptions du mot. Francis (Francis, 1991) donne comme un exemple le nom *impact* qui conserve deux acceptions : la première se manifeste en fonction « objet » - c'est l'effet de quelque chose. La deuxième se manifeste en fonction « adjectif » qui signifie l'action d'un objet en contact avec un autre. Francis constate que le comportement syntaxique est différent selon l'acceptation.

Sinclair remarque que la colligation participe, avec d'autres propriétés phraséologiques, à la constitution de l'« *unité lexicale étendue* » c'est-à-dire une séquence linguistique, qui possède un « coeur lexical » avec des caractéristiques sémantiques, syntaxiques et pragmatiques. Il en propose la définition suivante :

Colligation is similar to collocation in that they both concern the cooccurrence of linguistic features in a text. Colligation is the occurrence of a grammatical class or structural pattern with another one, or with a word or phrase. « Negative », « possessive » and « modal » are the kinds of largely grammatical categories that figure in colligation. The term was first used by J. R. Firth, and has been widened a little for corpus work. (Sinclair, 2003: 145)

Hoey (2005) enrichit la notion de colligation en défendant sa thèse du *lexical priming*. Selon lui, l'emploi d'un mot est amorcé par d'autres mots, dans certaines positions ou fonctions grammaticales, à un ou des endroits particuliers de la phrase, du paragraphe ou du texte. Legallois reformule ainsi la définition de la colligation proposée par Hoey :

- la compagnie grammaticale qu'un mot ou qu'une séquence de mots privilégie ou évite,
- les fonctions grammaticales préférées ou évitées par le syntagme auquel appartient le mot ou la séquence de mots,
- la position dans une séquence (phrase, paragraphe, texte) qu'un mot ou séquence de mots préfère (ou évite). (Legallois, 2012: 42).

Pour résumer, dans cette partie du chapitre, nous avons présenté les grandes lignes du domaine de la phraséologie. Ensuite, nous avons présenté un état des recherches actuelles sur deux unités phraséologiques qui seront au centre de nos analyses : la collocation et la colligation. Nous avons mené une analyse des travaux pour mieux comprendre ces notions et choisir une approche pertinente dans notre travail.

Dans le chapitre suivant, nous allons présenter un état des lieux sur la notion de genre textuel et sur le roman contemporain. Nous nous concentrerons sur les définitions et les caractéristiques de chaque sous-genre romanesque constituant notre corpus.

## Chapitre 3. La question du genre textuel

### 1. La notion de genre et le genre romanesque

La définition du « genre », un concept désignant une détermination suprasegmentale du sens d'un texte (ou d'un discours), reçoit des traitements différents par les chercheurs. Bouquet fait ce constat, d'après Charaudeau et Maingueneau (2002) :

La diversité des points de vue montre la complexité de la question des genres, jusques et y compris les dénominations puisque certains parlent de « genres du discours », d'autres de « genres de textes », d'autres encore de « types de textes » : J.-M. Adam oppose « genres » et « types de textes » (1999) ; J.-P. Bronckart oppose « genres de textes » et « types de discours » (1996) ; D. Maingueneau distingue, dans les rapports d'enchâssement, « types de textes », « hypergenre » et « genre de discours » (1998) ; P. Charaudeau distingue « genres et sous-genres situationnels » et, à l'intérieur de ceux-ci, des variantes de genres de discours (Bouquet, 2004 : 7).

La notion de genre textuel est particulièrement traitée par Bakhtine : « Chaque sphère d'utilisation de la langue élabore ses propres types relativement stables d'énoncés, et c'est ce que nous appelons les genres » (Bakhtine et al., 1984: 256). Cependant, la notion de genre était déjà présente dans l'antiquité, dans une vision prescriptive apparentée à la rhétorique. Branca-Rosoff remarque l'évolution de cette approche vers un point de vue plus descriptif. Elle indique aussi que la « modernité » provoque une « déstabilisation des genres » (Branca-Rosoff, 1999).

Bakhtine (1986) lie la notion de style au genre de discours et à l'empreinte individuelle d'un locuteur. Il existe des genres qui sont plus enclins à montrer le caractère individuel d'un locuteur, comme les genres littéraires. Pour Bakhtine, les faits grammaticaux et les faits stylistiques se trouvent au même niveau et se séparent dans tout fait concret. Du point de vue de la langue, un fait concret est un fait de grammaire. On identifie le fait concret avec un fait stylistique du point de vue de l'énoncé individuel. Ces deux niveaux de faits doivent se combiner organiquement à partir de l'unité réelle, un représentent du fait de langue (Bakhtine, 1984). Dans tout énoncé concret, la dimension stylistique peut se cacher sous un terme abstrait de « genre » : « La langue pénètre dans la vie à travers des énoncés concrets (qui la réalisent), et c'est encore à travers des énoncés concrets que la vie pénètre dans la langue. » (Bakhtine 1986; cité par Freyermuth, 2011: 174). La notion *énoncé concret* se définit comme le fait de produire un énoncé dans un cadre de communication précis. L'énoncé concret trouve sa place dans le processus de l'interaction sociale des participants de l'énoncé.

Nous pouvons envisager la notion de genre de deux points de vue - linguistique et

social. Bakhtine constate qu'il existe une sorte d'interdépendance entre des genres et des activités humaines. Il mentionne par exemple que l'utilisation de la langue se réalise au sein des domaines d'activités humaines et c'est la raison pour laquelle des liens existent entre la langue et le domaine d'activité : « Each separate utterance is individual, of course, but each sphere in which language is used developed its own relatively stable types of the utterances. These we may call speech genres. » (Bakhtine, 1986 : 60). Le nombre de genres peut être donc presque infini à cause de la multitude des champs d'activité humaine.

D'après Bronckart (1996), de nombreuses questions sont soulevées par la notion de genre textuel. Premièrement, on observe la difficulté de distinguer des genres particuliers. Plusieurs genres peuvent être à l'œuvre dans un texte : par exemple l'étiquette « drame romantique » recouvre des éléments de la tragédie et de la comédie. Une autre remarque met l'accent sur l'évolution continue des genres. Bronckart indique aussi qu'un point de vue peut lui-même avoir une influence sur la façon de considérer le genre. Bronckart attire l'attention sur l'impossibilité de classement stable et définitif des genres : « la finalisation actuelle d'un genre constituant généralement un mixte dépendant, et des décisions originelles d'une formation sociale, et de processus ultérieurs de "récupération" ou de "travestissement" » (Bronckart, 2008: 86). La détection de régularités linguistiques peut être la clé pour améliorer la distinction entre les (sous-)genres textuels. Bronckart (2004 : 105) constate que « les genres combinent des modes de structuration particulièrement hétérogènes, de sorte qu'ils ne peuvent jamais être entièrement définis par un ensemble donné d'opérations cognitives, qui seraient matérialisées par un ensemble donné d'unités et de règles linguistiques ».

L'un des facteurs les plus importants qui indique si un texte appartient à un genre est le fait qu'il possède une cohérence stable. Le texte doit avoir une structure spécifique permettant de classer le texte dans un genre. Les critères de sélection sont indispensables dans le processus de la classification des genres littéraires. Comme les critères principaux de catégorisation du genre littéraires, nous pouvons indiquer : la forme, le contenu et l'effet créé sur le lecteur. Dans la poésie ou le théâtre pour lesquels la forme (souvent versifiée) joue un rôle important, cette distinction pose peu de problèmes. Dans le cas du roman, pour lequel le contenu et l'effet créé sont des critères retenus, la distinction entre sous-genres devient plus délicate.

Schaeffer attire l'attention sur l'identification du genre romanesque qui peut être différent pour presque n'importe quel roman :

L'identité d'un genre est fondamentalement celle d'un terme général identique appliqué

à un certain nombre de textes. « Ce baptême » peut être collectif et unique pour toute la classe [...], ou plus souvent, individuel et donc multiple (c'est le cas, de manière exemplaire, de la dénomination roman qui est avant tout un élément paratextuel, c'est-à-dire un acte de baptême toujours lié à une œuvre individuelle). (Schaeffer, 1989: 65)

Le roman, selon Schaeffer, est un genre littéraire dont les frontières sont assez incertaines. Le roman manque d'une forme particulière, c'est-à-dire que nous ne pouvons pas identifier de structures formelles spécifiques comme par exemple la forme versifiée dans la poésie. L'une des rares particularités du roman est qu'il s'agit d'un texte d'une longueur considérable. Le contenu du roman, qui essaie d'obtenir un effet réaliste, est considéré comme le facteur le plus important pour déterminer l'appartenance à un sous-genre spécifique.

Dans la section suivante, nous présenterons les caractéristiques des six sous-genres romanesques qui constituent notre corpus de travail : les romans policiers, fantastiques, de science-fiction, historiques, sentimentaux et de littérature générale ou blanche.

## **2. *Les sous-genres textuels***

### **2.1. *Le roman policier***

Le genre policier désigne « une catégorie de roman dont la trame narrative est une enquête policière » (Boutevin & Principalli, 2008: 205). Poslaniec & Houyel (2001) indique que, dans un roman policier, se produit toujours une transgression de la loi comme par exemple un « crime ». Une dimension proprement juridique ou judiciaire doit être impliquée. Dans ce genre littéraire, on peut trouver des caractéristiques récurrentes, comme la présence d'un schéma narratif standard comprenant les éléments suivants : un délit, un mobile, un coupable, une victime, une enquête, des indices, etc. (Poslaniec & Houyel, 2001 : 31).

Pour trouver un coupable, une enquête est menée, ce qui implique la présence du personnage du détective. Dubois (2005) parle du « carré sémiotique parfait » avec le détective ingénieux et méthodique, le criminel, la victime et le suspect qui constitue une nouvelle particularité du genre. Une autre caractéristique consiste en la présence d'un raisonnement par hypothèses et déductions qui permet la reconstitution à rebours du méfait. Nous pouvons donc observer l'existence d'une infinité de combinaisons possible entre ces différentes caractéristiques.

Le genre policier se caractérise par une grande multiplicité de catégories très différentes avec leurs traits propres. Todorov (1980) réduit le nombre de formes de ce genre à trois principales : le récit d'énigme dit « classique », le roman noir et le roman à suspense.

- *Le récit d'énigme « classique »* : Lits (1999) le définit comme les romans « qui sont



caractérisés par un crime initial qu'un détective va tenter d'expliquer en menant une enquête qui remontera, à rebours, vers les antécédents du crime, pour faire parler les indices, démêler le vrai du faux et identifier le coupable, avant que le lecteur, également en quête, n'y parvienne » (Lits, 1999: 9).

Cette forme classique du roman policier se focalise sur l'énigme à résoudre et sur l'enquête du héros qui cherche à réunir et interpréter les indices. Il a été créé et utilisé dans les années 1880-1940 par les auteurs comme Poe, Conan Doyle, Christie, Gaboriau par exemple. En raison du fait que notre corpus comprend des romans contemporains écrits après 1950, il ne contient pas de romans de cette période.

- *Le roman noir ou polar* : dans ce type de roman policier, on trouve un mélange de l'histoire du crime et de l'enquête, parfois l'histoire du crime est supprimée. On observe un changement d'orientation de l'histoire. Dans le roman noir, la prospection se substitue à la rétrospection. Le personnage du détective se trouve en danger : arriver vivant à la dernière page du roman devient plus important que la résolution du mystère. L'entrée dans la psychologie du personnage détective fait aussi une grande différence. Le roman noir (ou polar) propose aussi une analyse sur la noirceur de la société qui souffre au travers des thèmes concernant la transgression de la loi. (Lits, 1999)

- *Le roman à suspense* : ce type semble être l'union des deux précédents. Il garde la présence des deux histoires (comme le récit d'énigme) mais comme dans le roman noir, la deuxième histoire a une place centrale. L'accent est mis sur le suspens, qui est fortement privilégié : « il est plutôt un point de départ, l'intérêt principal venant de la seconde histoire, celle qui se déroule au présent » (Todorov, 1971 : 63). Le roman à suspense montre la tension entre la vie quotidienne des personnages et l'inconnu qui y fait brusquement irruption.

Nous n'utiliserons pas dans notre travail une telle distinction entre ces trois catégories car elle n'est pas opératoire pour notre corpus.

L'étude du genre policier apparaît alors beaucoup plus complexe qu'il n'y paraît. Les variations au sein du genre sont nombreuses. Une classification des romans policiers implique une simplification créée sur la base de leurs caractéristiques discursives (Roa & Santiago, 2011).

La liste exhaustive des auteurs de romans policiers est disponible en annexe.

## 2.2. *Le roman fantastique*

Le fantastique (ou fantasy) est un genre littéraire qui exploite les potentialités de l'imaginaire. Selon Castex (1994: 8) « le fantastique se caractérise par l'irruption brutale du surnaturel dans le monde réel ». Les événements invraisemblables n'ont pas d'explication logique. C'est la raison du manque d'acceptation de l'irrationnel. Le fantastique joue sur l'ambiguïté sans affirmer les suppositions du lecteur qui trouve les choses inexplicables et inadmissibles.

Todorov essaie de définir le fantastique avec ses thèmes et réalisations variées comme par exemple l'étrange, le merveilleux ou l'allégorie. Il indique les trois conditions essentielles du fantastique :

D'abord, il faut que le texte oblige le lecteur à considérer le monde des personnages comme un monde de personnes vivantes et à hésiter entre une explication surnaturelle des événements évoqués. Ensuite, cette hésitation peut être ressentie également par un personnage ; ainsi le rôle de lecteur est pour ainsi dire confié à un personnage et dans le même temps l'hésitation se trouve représentée, elle devient un des thèmes de l'œuvre ; dans le cas d'une lecture naïve, le lecteur réel s'identifie avec le personnage. Enfin, il importe que le lecteur adopte une certaine attitude à l'égard du texte : il refusera aussi bien l'interprétation allégorique que l'interprétation "poétique" (Todorov, 1970: 37-38).

Todorov montre aussi une correspondance entre ces conditions et trois aspects suivants : le verbal, le syntaxique-sémantique, et le pluriel des niveaux de lecture. Chez Todorov, nous pouvons trouver le schéma de subdivisions du fantastique avec ses traits particuliers (Todorov, 1970: 49-62) :

- a) étrange pur - des événements insolites expliqués par la raison (coïncidences, « expérience des limites ») ;
- b) fantastique-étrange - des faits surnaturels recevant des explications rationnelles ;
- c) fantastique-merveilleux - acceptation du surnaturel ;
- d) merveilleux pur- événements surnaturels sans provoquer de la surprise.

Le fantastique est une littérature non réaliste qui « offre l'image renversée de l'union de l'âme et du corps » (Sartre, 1947: 115). Il s'agit d'une image dans laquelle l'esprit et la matière se subvertissent mutuellement, en créant une réalité à l'envers. En définissant le fantastique, on impose une référence simultanée à des termes proches, comme le merveilleux, le surnaturel ou la science-fiction. C'est un genre fondé sur l'inquiétude qui exploite les peurs du lecteur. Nous pouvons résumer ainsi ce genre comme l'irruption de l'irréel dans le réel, soit l'irruption d'un phénomène inexplicé dans le quotidien. On peut

observer la concurrence frontale de ce genre avec celui de la science-fiction (SF), que nous présenterons dans la sous section suivante, mais une différence essentielle entre les deux repose sur la nature des actions. Dans la fantasy, c'est la magie qui occupe le rôle le plus important, tandis que pour la SF, c'est la science.

La liste exhaustive des auteurs de romans fantastiques est disponible en annexe.

### ***2.3. Le roman de science-fiction***

La science-fiction (SF) est un genre romanesque qui raconte des fictions ayant pour cadre ou pour objet des progrès scientifiques et techniques dans un futur plus ou moins lointain. Selon Bozzetto (2007), l'auteur, en utilisant des images issues de savoirs actuels, extrapolées ou imaginaires mais vraisemblables dans le contexte scientifique de l'époque, essaie de créer des « mondes possibles ». Il les situe dans des futurs, des présents ou des passés parallèles. La SF « situe les actions des personnages dans un environnement cohérent, bien qu'éloigné du nôtre » (Bozzetto, 2007: 8).

La SF soulève le sujet des bouleversements introduits par le développement des sciences et des techniques en montrant leurs potentialités. Le sentiment moderne de l'incertitude joue un rôle important dans ce genre romanesque. La SF vise à explorer ou créer les rapports qu'entretient l'humanité avec « un environnement technologique, médiatique et même psychique, issu de découvertes réelles ou supposées » (Bozzetto, 2007 : 9). Souvent, la SF représente l'humanité occidentale. Ce genre romanesque repose sur l'appropriation mentale du futur en utilisant des métaphores empruntées aux technologies nouvelles. La SF décrit ainsi des fictions « mimétiques du virtuel » (Bozzetto, 2007) et utilise les mêmes stratégies narratives qui privilégient les narrations facilitant les métamorphoses et anamorphoses de la réalité.

Bozzetto (2007) souligne les différences entre la SF et les autres genres romanesques. Selon lui, la SF en faisant référence à l'imaginaire se distingue du genre fantastique, qui renvoie à l'inimaginable. Elle se différencie aussi du genre fantastique par le fait que le savoir dominant dans la narration de ce genre provient de la connaissance et de l'utilisation de la magie. Comme l'indique Bozzetto (2007: 8), « Parmi les littératures de l'imaginaire, la SF relève de la perspective dite « réaliste » ou « mimétique ». La SF se distingue du roman historique dans son rapport au temps parce que ce dernier est « mimétique du passé » et se réfère à des événements réalistes et du roman de littérature générale qui est considéré comme « mimétique de l'univers du présent ». Le roman de littérature générale montre un monde

compréhensible qui n'exige pas les explications supplémentaires alors que cela est nécessaire dans le cas de la SF.

#### **2.4. Le roman historique**

Dans le roman historique, les relations entre histoire et fiction sont constamment à interroger. Selon Gengembre, « le roman historique éclaire l'histoire. Il est même "plus vrai" que l'histoire parce qu'il la donne à comprendre de manière vivante » (Gengembre, 2006: 87). Ce genre romanesque mêle généralement des événements et des personnages réels et fictifs. Gengembre indique que le roman historique est un genre « qui prétend restituer une image relativement fidèle d'un passé daté, à l'aide d'une fiction mettant en scène des comportements et des mentalités de l'époque considérée, voire des personnage ayant réellement existé. » (Gengembre, 2006: 73). La ligne de partage entre des faits historiques et les œuvres d'imagination étant souvent discutable, nous pouvons donc constater la dualité du roman historique.

Gengembre propose la classification terminologique complexe de l'historiographie et du roman historique. En définissant le roman historique, il distingue les espaces de l'Histoire de ceux du roman et les termes *novel* et *romance* par rapport aux Histoire et roman :

Deux tendances se dessinent alors, bien différenciées en anglais : *romance* et *novel*. Le genre *romance* raconte des histoires fort éloignées de la vie quotidienne. Le genre *novel* place ses personnages dans un contexte présenté comme conforme à une réalité reconnaissable (Gengembre, 2006: 20).

La définition du roman historique indique qu'il a pour fonction la représentation fictionnelle du passé (Bernard, 1996). Lukács remarque que « la particularité des personnages dérive de la spécificité historique de leur temps » (Lukács, 2000: 17). Il met aussi l'accent sur l'« expérience vécue des masses » (Lukács, 2000: 21) et les événements comme « quelque chose d'historiquement conditionné » (Lukács, 2000: 23). Les romans historiques mettent donc en jeu les fonctions narrative, cognitive et anthropologique. La fonction narrative désigne l'entourage entier d'un récit donc la description d'un cadre, d'un personnage, d'une ambiance etc. (Schiff, 2012). La fonction cognitive fait référence à notre connaissance sur le sujet tandis que la fonction anthropologique montre l'homme dans tous les aspects, à la fois physiques et culturels (Huemer, Gibson, & Pocci, 2007).

Nous pouvons observer, dans le roman historique, la présence massive du passé sous des formes différentes. Ce genre romanesque a pour but de permettre une meilleure

compréhension des époques historiques ou des événements individuels de l'histoire en éclairant le présent. Il vise aussi à observer et à présenter les personnages dans les différentes modalités narratives.

À la différence du récit historique, dans le roman historique, les personnages principaux sont généralement fictifs. Le récit historique vise à être le plus proche possible des événements du passé car, comme l'indique Duby, « l'historien n'est pas libre. Il lui faut transcrire fidèlement ce qu'il extrait des documents, de toutes les traces du passé qu'il repère et qu'il lui est strictement interdit de manipuler » (Duby, 1996: 185). Le roman historique, au contraire du récit historique, n'a pas besoin d'être autorisé ni vérifié.

La liste des auteurs de sous-corpus de romans historiques se trouve en annexe.

### ***2.5. Le roman sentimental***

Le genre du roman sentimental prend sa source dans les romans grecs du début de notre ère. Il est considéré comme le plus ancien des genres romanesques occidentaux (Constans, 1999). Ce genre romanesque se consacre au développement de la relation romantique entre deux personnages et à la façon dont ils gèrent cette relation. Dans le roman sentimental, nous trouvons des intrigues parallèles, mettant en scène d'autres personnages ou d'autres événements. Souvent, l'intrigue est arrangée pour développer à la fois les émotions et les actions des personnages. Comme l'indique Larsson (1991):

le roman sentimental désigne une forme de réalisation romanesque qui conjugue, sur des plans textuels différents, les traits suivants : a) sur le plan thématique : mise en scène de la passion romanesque, celle-ci pouvant recouvrir des manifestations diverses ; amour ou haine ; indifférence (ni amour ni haine) ; amour et haine ; b) sur le plan actanciel : la présence d'actant engagé dans des relations passionnelles amoureuses ; c) sur le plan narratif, une organisation syntaxique régie, déterminée par les avatars subis par la seule passion amoureuse. (Larsson, 1991: 77)

Selon Constans (1999), le plan narratif du roman sentimental détermine sa forme :

le roman sentimental raconte une histoire d'amour. Le programme narratif global est structuré par trois invariants qui, à la fois, permettent de jalonner la narration et de distinguer le genre d'autres où l'amour, le désir amoureux, la passion, etc., fournissent aussi la matière première : le roman libertin, le roman érotique, le roman pornographique. Trois étapes aussi dans le programme narratif : la rencontre, la disjonction du fait d'obstacles, épreuves..., la réunion ou conjonction des amants dans le bonheur ou le malheur. (Constans, 1999:6).

Cette structure narrative particulière met l'accent sur un vaste éventail d'émotions et d'attitudes caractéristiques des personnages. Le genre du roman sentimental est surtout

marginalisé au sein de la littérature populaire. Une des raisons est le fait que ce genre semble surtout destiné aux femmes (Constans, 1999). Le roman sentimental se caractérise par l'inventaire des formes très partielles qui se limite à quelques locutions ou collocations, donc à des unités relativement figées. Legallois (2016) remarque que ce sous-genre romanesque est lié aux clichés. Ces raisons positionnent le roman sentimental dans une catégorie plutôt méprisée par les critiques et considérée comme un genre dévalorisé de seconde classe.

Nous incluons la liste des auteurs de roman sentimental de notre corpus en annexe.

Notre présentation des genres romanesques s'appuie sur des études qui traitent essentiellement des phénomènes extra-linguistiques comme les thèmes traités, les types de personnage ou les émotions suscitées chez le lecteur. Notre recherche s'inscrit dans le prolongement du projet PhraseoRom et l'approche linguistique de ce projet représente l'originalité de notre travail. Nous essayerons de déterminer des observables linguistiques spécifiques à chaque sous-genre afin de les distinguer les uns des autres.

## **Partie 2**

-

## **Méthodologie**

## Chapitre 4. Le Lexicoscope

Dans ce chapitre, nous allons présenter la méthodologie de notre recherche. On va présenter l'outil le *Lexicoscope* grâce auquel nous pouvons mener des recherches sur les corpus romanesques constitués dans le cadre du projet PhraseoRom.

### 1. *Présentation de l'outil et du corpus*

Pour effectuer notre recherche, nous avons utilisé un outil développé pour le traitement de corpus et l'exploration de la combinatoire lexico-syntaxique : le *Lexicoscope*. Cet outil a été initialement développé par Kraif & Diwersy (2012) dans le cadre du projet franco-allemand EMOLEX<sup>4</sup>. Le *Lexicoscope* et ses corpus sont librement accessibles en ligne<sup>5</sup>.

Dans notre recherche, nous allons comparer les différents phraséologismes de la boisson qui apparaissent dans six sous-genres romanesques contemporains en langue française: fantasy, sentimental, policier, général, science-fiction et historique. Le tableau 1 présente les caractéristiques du corpus<sup>6</sup>:

---

<sup>4</sup> Les recherches dans le cadre du projet EMOLEX ont permis de mener l'analyse contrastive du lexique des émotions en 5 langues : allemand, anglais, espagnol, français, russe.

<sup>5</sup> Voir le site : <http://phraseotext.univ-grenoble-alpes.fr/lexicoscope>

<sup>6</sup> La liste exhaustive des auteurs de chaque sous-genre de notre corpus se trouve en annexe.



	Nombre de tokens	Nombre d'auteurs	Nombre d'oeuvres
FY	13 323 976	43	105
SENT	9 802 410	39	112
POL	17 859 351	83	194
GEN	34 334 554	170	444
SF	13 173 618	39	151
HIST	14 868 273	39	113

Tableau 1 La composition du corpus romanesque PhraseoRom partitionné en sous-genres

Le *Lexicoscope* utilise le concept de *lexicogramme*, proposé par M. Tournier et utilisé dans le logiciel WebLex (Heiden & Tournier, 1998). Son objectif est de constituer la liste des cooccurents les plus fréquents pour un pivot donné, par calcul des mesures d'association statistiques et grâce à l'intégration des fréquences de cooccurrences (comme p. ex. le rapport de vraisemblance). Le *Lexicoscope* permet à l'utilisateur de construire les requêtes et de « définir lui-même les unités de cooccurrences : formes, lemmes, catégories morphosyntaxiques, traits additionnels (p.ex. sémantiques), relations syntaxiques (dans le cas des colligations) ou des combinaisons de ces informations » (Kraif & Diwersy, 2012: 400). Nous pouvons aussi appliquer des contraintes sur le contexte d'un pivot. Ce fonctionnement permet à l'utilisateur de neutraliser l'ambiguïté des contextes et de mener les analyses sur les aspects lexicaux et syntaxiques de ces associations. Dans la section suivante, nous présentons les fonctionnalités *concordance* et *cooccurrence* qui permettent d'analyser les corpus dans le *Lexicoscope* et d'extraire les données sur lesquelles notre analyse s'appuiera.

## 2. Concordance et cooccurrence

D'abord, pour extraire des données dans le *Lexicoscope*, il faut sélectionner le corpus de recherche (Figure 1) :

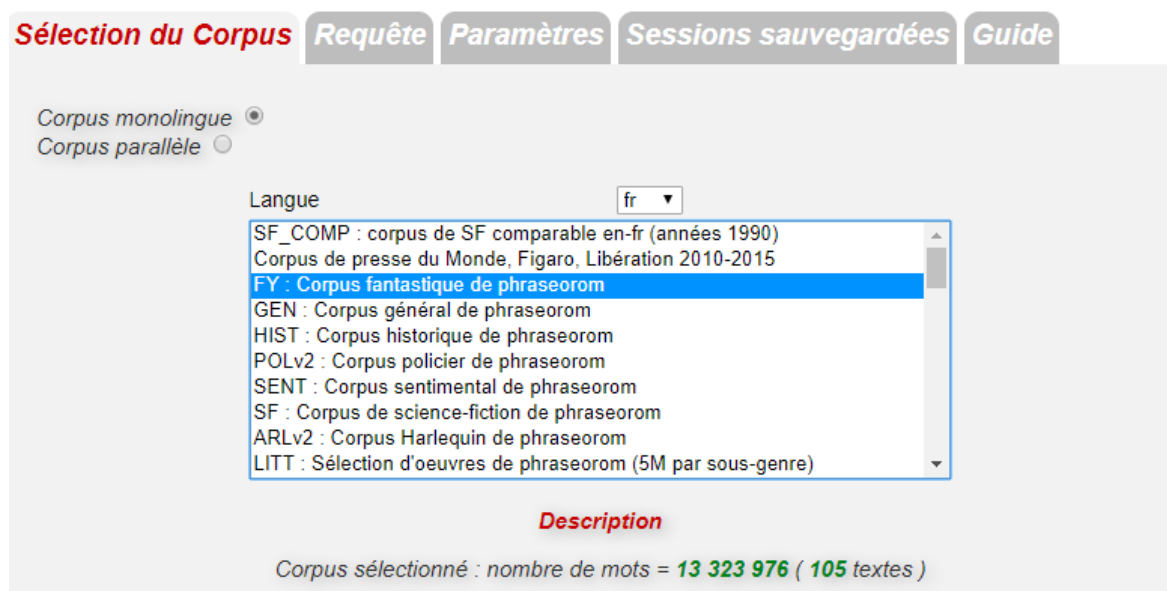


Figure 1 : Le choix du corpus dans le Lexicoscope

Après la sélection du corpus, on peut formuler des requêtes simples ou complexes pour indiquer le pivot. Cette opération nous permet d'extraire tout type de pivot – pas seulement des mots, mais aussi des expressions polylexicales - sous forme de concordances. Cette fonctionnalité nous permet de voir la réalisation en contexte (gauche et droit) du pivot. Le « Contexte gauche » par rapport au pivot représente l'énoncé en amont du pivot (généralement le début de la phrase). Le « Contexte droit » est souvent la fin de l'énoncé après le pivot. Le « Pivot », est isolé en vert, au centre du tableau (voir figure 2). Cette présentation permet une meilleure visualisation des contextes gauches et droits. La colonne « Identifiant » contient la référence codée de chaque segment qui renvoie aux métadonnées (le nom de l'auteur, le nom du livre et la représentation graphique de l'annotation syntaxique réalisée sur le segment en question). En bas de la même page, on peut trouver l'information sur le nombre d'occurrences (1213 dans la figure 2) et la dispersion soit le nombre d'auteurs différents chez lesquels se trouve le lemme (41 dans la figure 2). En bas de la page, on trouve le tableau avec la répartition par fichier donc l'information sur le nombre d'occurrences pour chaque texte (non visible sur la figure 2). Le figure 3 montre les concordances de la requête

simple effectuée avec le pivot *vin* (<w=vin,#1><>?<>?):

Num ^	Identifiant ↕	Contexte gauche	Pivot ↕	Contexte droit
1	s107	Cinquante litres de	vin	, même de mauvais vin, c'est une masse à soulever.
2	s107	Cinquante litres de vin, même de mauvais	vin	, c'est une masse à soulever.
3	s357	Nos verres gisaient sur cette grève, couchés çà et là sur le flanc comme des bateaux à marée basse, et le goulot d'une bouteille de	vin	pointait hors d'un panier ventripotent, beaupré de verre pour galion d'osier.
4	s4656	Derrière leurs quatre-vingt centimètres de robe de bure, leur faciès bonhomme et leur goût du	vin	, ils tenaient davantage du défroqué un peu trop porté sur les choses de la vie que de la sentinelle à l'affût.\n
5	s5764	– S'offrent un repas et un	vin	chaud, on dirait.\n
6	s5868	Il leur sera servi un	vin	chaud.
7	s9790	Nous pénétrâmes dans une ancienne cave à	vin	.
8	s678	Deux paysans buvaient de la bière, trois bourgeois installés sur des bancs comptaient de l'argent autour d'une cruche de	vin	, un adolescent portant un tablier s'affairait avec un plateau autour d'un groupe de nobles élégamment habillés.
9	s680	Ils étaient trois : deux femmes, jeunes et belles, qui sirotaient du	vin	dans des verres presque propres et un jeune homme à la chemise blanche impeccable et aux cheveux coupés à la dernière mode, dévorant une énorme tranche de pain bis.\n
10	s684	Arekh s'assit et commanda un large repas : du pain, de la soupe, de la viande pour l'esclave, du ragoût et une cruche de	vin	rouge pour lui.

Showing 1 to 10 of 1,213 entries

[First](#)
[Previous](#)
[1](#)
[2](#)
[3](#)
[4](#)
[5](#)
[...](#)
[122](#)
[Next](#)
[Last](#)

**Nombre total d'occurrences 1213**

Dispersion

41

Figure 2 : L'exemple des concordances pour le pivot *vin*

L'étiquetage morpho-syntaxique des tokens permet de chercher une forme particulière d'un mot (formalisée par « w= » dans la requête) mais aussi toutes les formes d'un lemme (formalisée par « l= » dans la requête). La formule de requête affiche tous les paramètres de recherche.

Pour présenter les résultats, nous obtenons des tableaux nommés lexicogrammes. On a choisi de montrer les exemples pour le pivot *vin* qui se trouvent dans le corpus FY (figure 3) et dans le corpus POL (figure 4) :

I1	I2	f.deprels	f	f1	f2	N	f.disp	am.log.likelihood	r.log.likelihood
vin_*	de_PREP	PREPOBJ#1	743	3039	654850	19248988	40	1802,8913	1
vin_*	verre_NOUN	~NMOD_POSIT1#1	82	3039	3870	19248988	17	644,6470	2
vin_*	bouteille_NOUN	~NMOD_POSIT1#1 NMOD_POSIT1#1	49	3039	1244	19248988	18	446,0244	3
vin_*	coupe_NOUN	~NMOD_POSIT1#1 NMOD_POSIT1#1	48	3039	1339	19248988	15	427,7249	4
vin_*	gorgée_NOUN	~NMOD_POSIT1#1 NMOD_POSIT1#1	44	3039	987	19248988	13	411,5269	5
vin_*	boire_VERB	~VMOD#1 NMOD_POSIT1#1 ~VMOD_POSIT1#1 ~OBJ#1	58	3039	5692	19248988	19	370,8604	6
vin_*	chaud_ADJ	NMOD_POSIT1#1	42	3039	1406	19248988	13	358,6746	7
vin_*	rouge_ADJ	NMOD_POSIT1#1 ~ADJMOD#1	32	3039	2545	19248988	13	217,6897	8
vin_*	du_DET	DETERM_DEF#1	33	3039	3430	19248988	12	206,9912	9
vin_*	pichet_NOUN	SEQNP#1 ~NMOD_POSIT1#1	18	3039	279	19248988	12	181,8383	10
I1	I2	f.deprels	f	f1	f2	N	f.disp	am.log.likelihood	r.log.likelihood

Figure 3 : Lexicogramme relatif au pivot *vin* dans le FY

I1	I2	f.deprels	f	f1	f2	N	f.disp	am.log.likelihood	r.log.likelihood
vin_NOUN	de_PREP	PREPOBJ	740	5739	782753	24713762	63	1019,7883	1
vin_NOUN	verre_NOUN	~NMOD_POSIT1 SEQNP NMOD_POSIT1	133	5739	10346	24713762	37	811,1349	2
vin_NOUN	blanc_ADJ	NMOD_POSIT1	112	5739	5225	24713762	28	796,5733	3
vin_NOUN	bouteille_NOUN	~NMOD_POSIT1 NMOD_POSIT1	112	5739	5573	24713762	42	782,1366	4
vin_NOUN	tache_NOUN	~NMOD_POSIT1	48	5739	2921	24713762	9	315,4166	5
vin_NOUN	rouge_ADJ	NMOD_POSIT1	49	5739	3967	24713762	24	294,2725	6
vin_NOUN	boire_VERB	~OBJ ~VMOD_POSIT1 NMOD_POSIT1 ~VMOD	55	5739	10629	24713762	22	237,0997	7
vin_NOUN	gorgée_NOUN	~NMOD_POSIT1	33	5739	1807	24713762	18	223,7001	8
vin_NOUN	chaud_ADJ	NMOD_POSIT1	32	5739	1804	24713762	12	215,0598	9
vin_NOUN	du_DET	DETERM_DEF	28	5739	6379	24713762	16	111,7949	10
I1	I2	f.deprels	f	f1	f2	N	f.disp	am.log.likelihood	r.log.likelihood

Figure 4 : Lexicogramme relatif au pivot *vin* dans le POL

Dans le lexicogramme, la colonne « I1 » indique le pivot et sa catégorie grammaticale ; dans la figure 3, le pivot *vin* est un nom. La colonne « I2 » est celle des cooccurents statistiquement significatifs, et indique le cooccurent avec sa catégorie grammaticale, comme par ex. le cooccurent *verre* qui est un nom ou le cooccurent *boire* qui est un verbe pour le pivot *vin*. La colonne « f.deprels » indique la relation de dépendances syntaxiques entre le pivot et le cooccurent par ex. PREPOBJ indique que la préposition *de* introduit un syntagme prépositionnel contenant le pivot *vin*. Ce collocatif est le plus fréquent dans les deux sous-genres POL et FY. La colonne « f » comporte les fréquences de la cooccurrence du pivot et du cooccurent. La colonne « f1 » indique la

fréquence du pivot et la colonne « f2 » celle du cooccurrent. La colonne « f.disp » comporte la dispersion soit le nombre de textes dans lequel l'association pivot + cooccurrent est présente. Le facteur « am.log.likelihood » présente les valeurs de spécificité statistique de chaque collocation en indiquant la force d'association. La colonne « r.log.likelihood » classe les collocations par rang selon le taux de spécificité. Plus la valeur de log est élevée, plus le rang est petit.

Une étude approfondie des cooccurrents nous permettra d'indiquer pour chaque sous-corpus le type de collocations spécifiques. Nous allons présenter dans la section suivante ce que sont les arbres lexico-syntaxiques récurrents (ALR) extraits pour chaque pivot.

### 3. L'arbre lexico-syntaxique récurrent (ALR)

Le Lexicoscope permet de faire une recherche complexe sur les textes en utilisant la syntaxe, indépendamment de l'ordre des mots. On peut aussi chercher une expression qui contient au moins deux mots. Les représentations graphiques des ALR ont pour objectif de visualiser les liens de dépendances syntaxiques entre les unités lexicales. Dans la figure 5, on observe une hiérarchie de relations syntaxiques entre chaque élément :

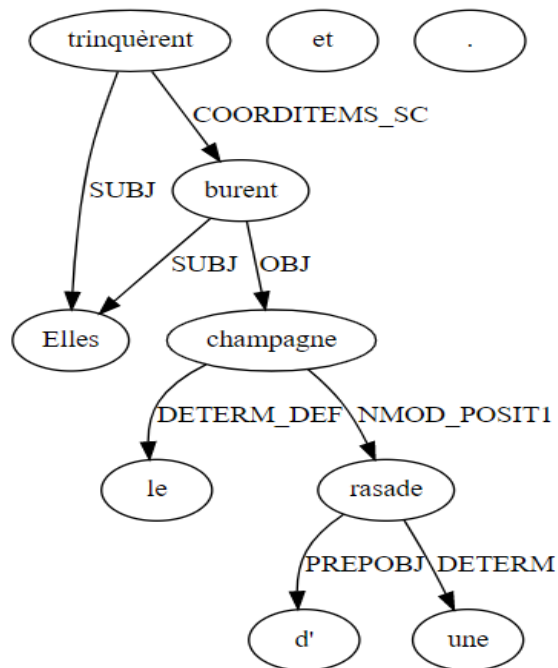


Figure5 : Exemple de l'arbre lexico-syntaxique dans le Lexicoscope: annotation syntaxique de la phrase « Elles trinquèrent et burent le champagne d'une rasade » (Jacques Laurent, *Les Sous-ensembles flous*, 1981)

Les résultats sont également affichés sous forme d'un tableau qui spécifie les informations sur l'ordre des mots, le lemme, la forme, la catégorie, les traits et les relations syntaxiques (figure 6) :

id	forme	lemme	catégorie	Traits	Relations
0	Elles	il	PRON	Nom Fem PL P3 PC	
1	trinquent	trinquer	VERB	avoir PasS PL P3 Verb VERB_P3PL	COORDITEMS_SC->3 SUBJ->0
2	et	et	COORD	Coord COORD	
3	burent	boire	VERB	aSN avoir SN PasS PL P3 Verb VERB_P3PL	OBJ->5 SUBJ->0
4	le	le	DET	Masc SG DefDet DET_SG	
5	champagne	champagne	NOUN	Masc SG Noun NOUN_SG	DETERM_DEF->4 NMOD_POSIT1->8
6	d'	de	PREP	Prep PREP_DE	
7	une	un	DET	Fem SG IndefDet DET_SG	
8	rasade	rasade	NOUN	Fem SG Noun NOUN_SG	PREPOBJ->6 DETERM->7
9	.	.	SENT	SENT	

Figure 6 : Représentation en tableau d'un ALR dans le Lexicoscope

En analysant l'ALR présenté dans la figure 6, nous remarquons huit différentes relations entre les éléments. Premièrement, pour le verbe *trinquer*, on a deux relations :

- 1) COORDITEMS\_SC- : les verbes *boire* et *trinquer* sont coordonnés par *et*.
- 2) SUBJ- *Elles* est le sujet des verbes *trinquer* et *boire*
- 3) OBJ- *champagne* est le pivot du complément d'objet direct du verbe *boire*.

Pour le nom *champagne* :

- 1) DETERM\_DEF- *le* est le déterminant défini du nom *champagne*
- 2) NMOD\_POSIT1 (noun : *champagne*, noun : *rasade*) le nom *rasade* est un modifieur du nom *champagne*

Pour le nom *rasade* :

- 1) PREPOBJ- (noun : *rasade*, prep : *de*) – *de* est une préposition postposée à *rasade*
- 2) DETERM- (noun : *rasade*, det : *une*) — *une* est le déterminant du nom *rasade*

La figure 7 montre un exemple de lexicogramme, relatif au pivot *thé* :

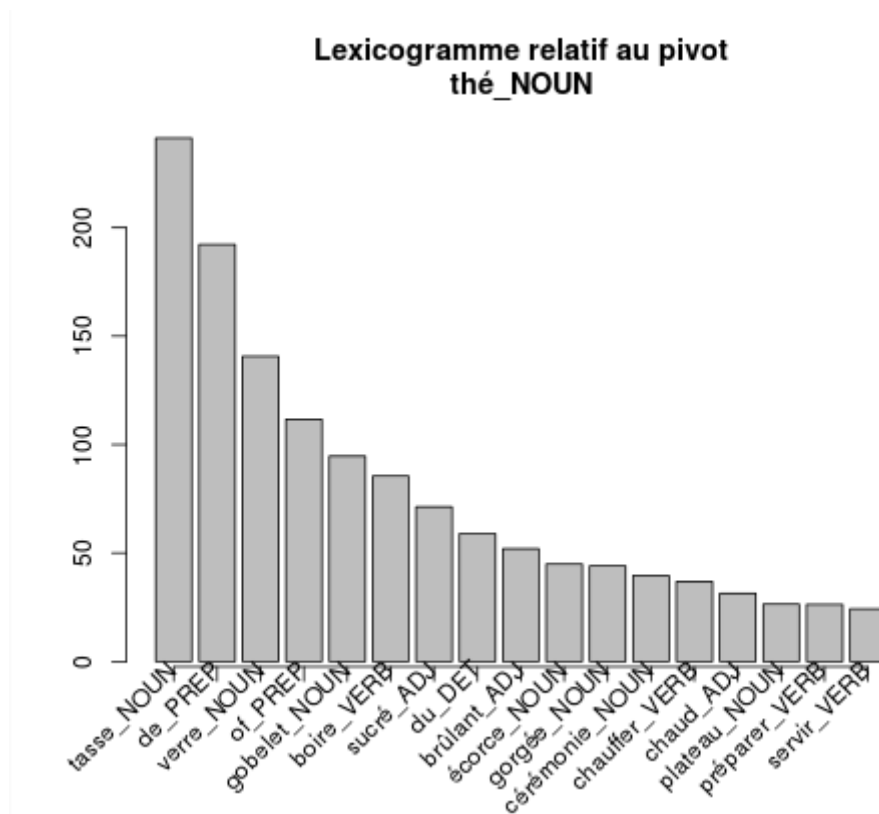


Figure 7. Lexicogramme relatif au pivot *thé* dans FY

La technique d'extraction d'arbres lexico-syntaxiques récurrents (ALR) permet d'identifier des constructions lexico-syntaxiques ayant une récurrence significative sur un plan statistique. On utilise le calcul de rapport de vraisemblance pour la détermination d'un ALR comme spécifique pour un corpus donné. Ce calcul de rapport de vraisemblance (Log Likelihood Ratio ou LLR) est une mesure statistique qui indique la spécificité d'un pivot par rapport à l'ensemble du corpus. La figure 8 montre les collocatifs du pivot *thé* dans le corpus FY. Ils sont classés du plus spécifique au moins spécifique. Le valeur de LLR est l'unité de mesure.

En utilisant la fonction « requête avancée », on peut approfondir notre recherche en obtenant des résultats de plus en plus spécifiques. Pour l'expression « boire le champagne » dans GEN, la requête est la suivante :

```
<l=boire,c=VERB,#1>&&<l=le,c=DET,#2>&&<l=champagne,c=NOUN,#3>:(DETERM
_DEF,3,2) (OBJ,1,3)
```

Il indique que :

- a) on cherche le lemme *boire* (toutes les formes possibles) → l=boire
- b) la catégorie du lemme *boire* est un verbe → c=VERB

- c) le lemme *boire* est le premier élément de la requête → #1
- d) on a un nombre indéterminé d'éléments entre le premier mot et le deuxième mot → >&&<
- e) le lemme *le* est dans la catégorie DET (*déterminant*) et c'est le deuxième élément → <l=le,c=DET,#2>  
le lemme *champagne* est un *nom* et le troisième élément
- f) → <l=champagne,c=NOUN,#3>
- g) les relations entre les éléments) l'élément 2 (*le*) est le déterminant de l'élément 3 (*champagne*) → (DETERM\_DEF,3,2) et l'élément 3 (*champagne*) complète l'élément 1 (le verbe *boire*) ;(OBJ,1,3)

Dans la requête avancée, on a la possibilité de décrire les relations entre les éléments. En formulant la requête, on peut choisir quel type de relations observer : par exemple, quand on indique la relation « ~OBJ », le pivot est en position de dépendance d'un verbe. Le Lexicoscope permet de rechercher en combinant des critères, mais aussi avec des alternatives « | », comme dans l'exemple suivant avec le nom *bière* dont le verbe régisseur est soit *boire* soit *commander* :

<l=boire|commander,c=VERB,#1>&&<l=bière,c=NOUN,#2>

On peut aussi indiquer les éléments facultatifs avec « ? » ou facultatifs et répétables avec « \* ». Dans le Lexicoscope, il est possible de rechercher les mots apparaissant en cooccurrence syntaxique d'un autre mot. Pour ce type de recherche, on utilise le mode « Requête multi-pivots » en indiquant le mot et la relation qui nous intéresse. Par exemple, on demande les cooccurrences du mot *boire* avec n'importe quelle relation et avec la relation OBJ. Dans le premier cas, on entre le mot *boire* et pour n'importe quelle relation on entre « \* ». On obtiendra un lexicogramme avec tous les cooccurrences et les paramètres. Néanmoins, ce sont seulement les mots qui apparaissent avec le lemme *boire*. Pour observer les collocations possibles, on clique sur le pivot pour voir le contexte. On a par exemple la cooccurrence du mot *boire* avec le mot *goulot* et dans le contexte, on voit que c'est la collocation *boire au goulot*. Un autre exemple intéressant est la collocation *boire d'une traite* dans la relation « VMOD\_POSIT1 ». Pour une recherche avec la relation indiquée, notre requête est plus précise donc nous allons obtenir moins de résultats.



Dans le chapitre suivant, nous allons expliquer les pivots choisis pour mener notre recherche. Nous présenterons aussi la méthodologie de l'analyse de la combinatoire lexico-syntaxique que nous allons utiliser.

## Chapitre 5. Méthodologie du traitement de données

### 1. L'étude de la combinatoire lexico-syntaxique

Blumenthal (2007: 19) définit le profil combinatoire d'un mot comme « l'ensemble de ses accompagnateurs stéréotypés, porteurs d'associations typiques ». L'étude de la combinatoire lexico-syntaxique est très souvent associée à des méthodes quantitatives. Cette approche vise à améliorer l'identification de la signification et la typologie des lexies. Comme l'indiquent Novakova & Tutin (2009 : 7-8), « la combinatoire syntaxique renvoie aux structures actanciennes des lexies, la combinatoire lexicale, quant à elle, englobe les cooccurrents lexicaux qui entretiennent une relation sémantique privilégiée avec le mot pivot, comme par ex. aimer passionnément ». Dans notre recherche, nous allons nous occuper de caractériser l'environnement syntaxique et fonctionnel de collocations et colligations issus de notre corpus. Nous essayerons d'examiner comment le fonctionnement dans le texte d'un pivot relevant du champ de la boisson, c'est-à-dire son profil combinatoire, va permettre de distinguer les sous-genres romanesques les uns des autres.

Dans notre étude en phraséologie, nous allons utiliser la méthodologie qui est très souvent pratiquée dans ce type de recherche. Nous classerons les différents composants des unités polylexicales selon leur *structure lexico-syntaxique* (Kraif, Novakova, & Sorba, 2016). Ce classement implique l'identification précise des constituants lexicaux et de leur structure syntaxique. Le classement selon la structure lexico-syntaxique nous permet de former les constructions composées de différentes catégories. Nous mènerons notre recherche sur les constructions verbales, nominales et adjectivales. En prenant compte des critères définitoires des unités phraséologiques que nous avons présenté dans la première partie de nos travaux, nous allons les examiner dans la façon suivante :

- les collocations verbales (N+V) → ex. *boire du vin, préparer du café*
- les collocations nominales (N+N) → ex. *un verre de vin, tasse de café*
- les colligations → ex. *café au lait*

Notre outil nous donne la possibilité d'extraire automatiquement des expressions polylexicales statistiquement significatives dans un certain corpus, dans notre cas, c'est un certain sous-genre romanesque. Dans la section suivante, nous allons présenter le choix de pivots dans notre recherche.

## ***2. Le choix de pivots***

Notre travail s'inscrit dans la lignée des recherches de Gonon & Sorba (2019) sur les phraséologismes autour du vin et du champagne dans le roman contemporain. Sur la base des statistiques générales des boissons consommées en France, nous avons créé une liste de noms de boisson. On utilise les statistiques préparées pour le journal *Les Echos*<sup>7</sup> et les statistiques concernant les boissons préférées des Français pour l'été<sup>8</sup>. Nous avons identifié non seulement les noms de boissons consommées les plus fréquemment par les Français mais aussi ceux qui sont moins courants. Pour effectuer notre recherche, nous avons choisi les noms de boissons suivantes : *champagne, vin, bière, vodka, café et thé*. Nous avons exclu le nom « eau » parce que dans de nombreux contextes, il peut ne pas désigner une boisson. Ces nom pivots peuvent être répartis en deux catégories : les noms de boissons alcoolisées (*champagne, vin, bière, vodka*) et ceux de boissons non alcoolisées (*café, thé*).

Le Lexicoscope nous permet de mener les recherches dans les sous-corpus afin de trouver les collocations les plus spécifiques contenant un nom de boisson. En analysant les lexicogrammes de pivots contenant les noms de boissons mentionnés ci-dessus, on va dégager les collocations les plus spécifiques dans chaque de sous-genre romanesque. Ce type de recherche va permettre d'indiquer si les collocations qui contiennent le nom d'une boisson peuvent être utilisés pour distinguer les genres romanesques.

Ainsi, dans cette partie 2 de notre travail, nous avons montré comment fonctionne l'outil de fouille utilisé, le *Lexicoscope*, et présenté notre méthodologie, les corpus romanesques de notre étude et les observables que nous avons dégagés. Dans la partie 3 de notre travail, nous allons présenter l'analyse de données et les résultats sur les plans quantitatif et qualitatif.

---

<sup>7</sup> <https://www.lesechos.fr/2016/04/boissons-les-francais-ont-une-consommation-tres-eclectique-228106>

<sup>8</sup> <https://fr.statista.com/infographie/14509/les-boissons-preferees-de-lete/>

## **Partie 3**

-

## **Analyse**

L'objectif de cette partie est la présentation de données de notre recherche dans une approche quantitative et qualitative. Dans le chapitre 7 nous allons montrer de manière exhaustive les données sur les collocatifs statistiquement significatifs des noms de boissons. Le chapitre 8 est consacré à l'analyse de deux constructions spécifiques de notre corpus : *boire du vin et prendre une bière*.

## **Chapitre 7. La combinatoire des noms de boisson : approche quantitative**

Dans ce chapitre 7, nous allons présenter notre recherche dans une approche quantitative. Dans la section 1, nous allons présenter les collocatifs les plus spécifiques de chaque pivot (*vin, champagne, bière, vodka, café, thé*) dans chaque de sous-genre romanesque de notre corpus (FY, HIST, SF, SENT, POL, GEN)<sup>9</sup>. Pour sélectionner les collocatifs les plus spécifiques de chaque pivot dans chaque sous-genre de notre corpus, nous avons appliqué les critères suivants :

- Un LLR supérieur ou égal à 10,83.
- Un critère de dispersion : les ALR sont présents chez au moins 15 auteurs différents.
- Un critère morpho-syntaxique : la présence d'un adjectif, un nom ou un verbe.

Dans la section 2, nous présenterons une synthèse des remarques sur la spécificité des collocations autour des noms de boissons dans chaque sous-genre romanesque. .

### ***1. Les collocatifs des noms de boisson***

#### ***1.1. Les collocatifs du pivot « vin »***

Le pivot *vin* a le plus de collocatifs statistiquement significatives dans GEN (16 collocatifs) et le moins dans SF (2 collocatifs) comme le montre la figure 1 :

---

<sup>9</sup> Les tableaux exhaustifs de tous les collocatifs dans chaque sous-corpus romanesque sont disponibles en annexe.

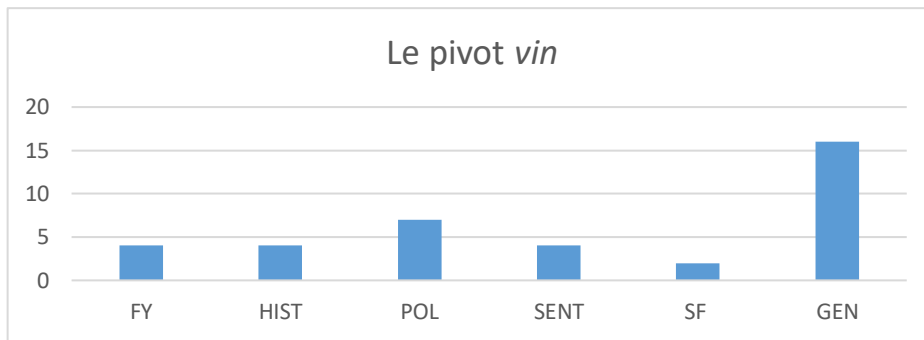


Figure 1 La distribution des collocatifs du pivot *vin* dans les sous-corpus romanesques

Dans le sous-corpus de romans FY, le pivot *vin* apparaît avec 3 collocatifs nominaux (*verre, bouteille, coupe*) et 1 collocatif verbal (*boire*). Le collocatif le plus spécifique du pivot *vin* dans le roman FY est le nom *verre* (LLR 644,6470, voir ex.1).

1) Un **verre de vin** à la main, ils devisaient. (R. Michel, *L'agent des ombres, tome 3 : Sang-pitié*, 2006)

Dans HIST, on observe que le pivot *vin* est présent avec 2 collocatifs nominaux (*verre, bouteille*), 1 collocatif verbal (*boire*) et 1 collocatif adjectival (*blanc*). À la différence de FY, le collocatif le plus spécifique pour le pivot *vin* pour HIST est le verbe *boire* (LLR 649,5558, voir ex.2).

2) Il se **boit beaucoup de vin** et de goutte. (C. Bernard, *Les grands malheurs*, 2004)

Le pivot *vin* dans POL apparaît dans la même proportion, accompagné de 3 collocatifs nominaux (*verre, bouteille, gorgée*) et de 3 collocatifs adjectivaux (*blanc, rouge, bon*) ; il y a aussi 1 collocatif verbal (*boire*). Comme dans FY, le collocatif le plus spécifique pour le pivot *vin* dans le roman policier est le nom *verre* (LLR 811,1349, voir ex.3).

3) Je me sers un **verre de vin** en lui faisant signe de continuer. (A. Brigitte, *Éloge de la phobie*, 2000)

Dans le sous-corpus de romans SENT, le pivot *vin* apparaît avec 2 collocatifs nominaux (*verre, bouteille*), 1 collocatif verbal (*boire*) et 1 collocatif adjectival (*blanc*). Le collocatif le plus spécifique pour le pivot *vin* dans le roman sentimental est encore le nom *verre* (LLR 748,0847, voir ex.4).

4) Antoine prit un autre **verre de vin** rouge. (A. Isabelle, *Tous à mes pieds*, 2008)

Dans les romans SF, on observe exclusivement 2 collocatifs nominaux pour le pivot *vin* (*bouteille, verre*). Le collocatif le plus spécifique est ici le nom *bouteille* (LLR 343,7050, voir ex.5).

5) Ils trinquèrent, finirent leurs assiettes, vidèrent la **bouteille de vin** et commandèrent des cafés. (P. Bordage, *Porteurs d'âmes*, 2007)

Dans GEN, on peut observer une présence dans des proportions presque identiques de collocatifs nominaux (5 collocatifs : *verre, bouteille, gorgée, marchand, pain*), verbaux (5 collocatifs : *boire, verser, goûter, servir, apporter*) et adjectivaux (6 collocatifs : *blanc, rouge, chaud, bon, mauvais, frais*), même si le nombre de collocatifs adjectivaux est légèrement plus élevé que celui des collocatifs nominaux ou verbaux. En revanche, le collocatif le plus spécifique du pivot *vin* dans le sous-corpus de roman général reste également le nom *verre* (LLR 2018,5226, voir ex.6).

6) Elle avait immédiatement répliqué en lui lançant son **verre de vin**. (T. Ben Jelloun, *Le bonheur conjugal*, 2012)

Pour récapituler, nous pouvons constater que le collocatif *verre* est le plus spécifique pour le pivot *vin* dans presque tous les sous-corpus, en excluant HIST et SF. On observe également la prédominance des collocatifs nominaux pour le pivot *vin* dans presque tous les sous-genres romanesques (GEN excepté).

### **1.2. Les collocatifs du pivot « champagne »**

Le pivot *champagne* a seulement des collocatifs statistiquement significatifs dans les sous-corpus GEN et POL (3 collocatifs dans chacun) ainsi que dans SENT (2 collocatifs). Dans les sous-corpus HIST, FY et SF, aucun collocatif spécifique du pivot *champagne* n'a été extrait selon les critères exposés plus haut.

Dans POL, les collocatifs nominaux sont les plus spécifiques pour le pivot *champagne*. On observe 2 collocatifs nominaux (*coupe et bouteille*) et 1 collocatif verbal (*boire*). Le nom *coupe* (LLR 500,0641, voir ex.7) est le collocatif le plus fréquent.

7) Au bureau, Antonin dut reboire une autre **coupe de champagne** et supporter les compliments empoisonnés de ses collègues. (P. Bruckner, *La maison des anges*, 2013)

Dans SENT, on observe la présence exclusive de 2 collocatifs nominaux (*coupe* et *bouteille*). Le nom *coupe* (LLR 1030,4259, voir ex.8) est également le collocatif le plus spécifique .

8) Séraphine, pas convaincue, se ressert une **coupe de champagne**, après en avoir proposé aux autres. (A. Abecassis, *Soirée sushi*, 2010).

Comme dans les autres sous-genres, le pivot *champagne* dans GEN apparaît plutôt avec les collocatifs nominaux. On trouve 2 collocatifs nominaux (*coupe* et *bouteille*) et 1 collocatif verbal (*boire*). Le nom *coupe* (LLR 876,1371, voir ex.9) est encore le collocatif le plus spécifique.

9) Le journaliste buvait, par petites gorgées, une **coupe de champagne**. (P. Combescot, *Les filles du Calvaire*, 1991)

Dans les sous-corpus POL, SENT et GEN, le collocatif *coupe* est le plus spécifique pour le pivot *champagne*. Les collocatifs du pivot *champagne* sont essentiellement nominaux dans ces 3 sous-genres.

### ***1.3. Les collocatifs du pivot « bière »***

Nos critères n'ont pas permis d'extraire des collocatifs spécifiques dans les sous-genres FY, HIST, SENT et SF pour le pivot *bière*. Nous trouvons seulement des collocatifs statistiquement significatifs dans GEN (8 collocatifs) et dans POL (5 collocatifs).

Les collocatifs verbaux sont les plus spécifiques pour le pivot *bière* dans le roman POL dans lequel on trouve 3 collocatifs verbaux (*boire*, *commander*, *prendre*) et 2 collocatifs nominaux (*bouteille* et *gorgée*). Le verbe *boire* est le collocatif le plus fréquent (LLR 713,1845, voir ex.10).

10) Je me levai un moment pour **boire une bière** et me détendre. (D. Daeninckx, *Meurtres pour mémoire*, 1984)

Les collocatifs nominaux sont les plus spécifiques pour le pivot *bière* dans GEN. On y trouve 4 collocatifs nominaux (*canette*, *bouteille*, *gorgée* et *verre*), 3 collocatifs verbaux (*boire*, *commander*, *servir*) et 1 collocatif adjectival (*frais*). Néanmoins, le collocatif le plus spécifique dans ce cas est le verbe *boire* (LLR 950,1346, voir ex.11).

11) Il **buvait de la bière** le soir jusqu'à ce que sa tête tombe dans son assiette. (G. Delacourt, *On ne voyait que le bonheur*, 2014)

Pour récapituler, nous observons une différence dans la nature de collocatifs les plus spécifiques entre ces deux sous-genres romanesques. Dans le sous-corpus POL les collocatifs verbaux sont le plus spécifiques, pendant que dans GEN les plus spécifiques sont les collocatifs nominaux.

#### ***1.4. Les collocatifs du pivot « vodka »***

Nous ne pouvons pas mener l'analyse sur les collocatifs du pivot *vodka* dans le sous-corpus de roman FY, HIST, SENT et SF car les critères de dispersion et morpho-syntaxique ne sont pas remplis. L'analyse est possible seulement sur les sous-corpus POL et GEN. Pour ces deux sous-genres, on observe exclusivement des collocatifs nominaux (1 collocatif nominal dans POL : *bouteille* ; 2 collocatifs nominaux dans GEN : *bouteille, verre*). Le nom *bouteille* est le collocatif le plus spécifique pour POL (LLR 280,8458, ex.12) et GEN (LLR 244,8336, ex.13).

12) La femme turquoise entra, munie d'un plateau chargé d'une **bouteille de vodka** et de sodas. (J.-Ch. Grange, *Le vol de cigognes*, 1994)

13) Ferrandi était au bord de s'écrouler, il avait bu entièrement la **bouteille de vodka**. (Y. Haenel, *Les renards pâles*, 2013)

Nous remarquons le statut particulier du pivot *vodka* qui a le moins de collocatifs spécifiques parmi tous les noms de boissons étudiés.

#### ***1.5. Les collocatifs du pivot « café »***

Les extractions réalisées à partir du pivot *café* dans les sous-corpus de romans FY, HIST et SF montrent que les critères de dispersion et morpho-syntaxique ne sont pas remplis. Par ordre décroissant, on observe le plus grand nombre de collocatifs statistiquement significatifs dans POL (25 collocatifs), un peu moins dans GEN (20 collocatifs) et beaucoup moins dans SENT (5 collocatifs).

Au cours de la recherche sur le pivot *café*, nous avons exclu les collocatifs *coin, entrer, garçon, terrasse*. car ils désignent l'établissement où l'on consomme des boissons et non la boisson elle-même.

Dans le sous-corpus POL, on observe une majorité de collocatifs verbaux pour le pivot *café* : 12 collocatifs verbaux (*boire, préparer, commander, servir, offrir, avaler, siroter, verser, apporter, prendre, proposer, chercher*), 7 collocatifs adjectivaux (*brûlant,*



*noir, chaud, froid, autre, bon petit*) et 6 collocatifs nominaux (*tasse, gorgée, thermos, bol, lait, odeur*). Voici le tableau présentant les collocatifs avec leur LLR :

<b>Collocatif</b>	<b>LLR</b>
<i>tasse</i>	1733,3311
<i>boire</i>	1274,5376
<i>gorgée</i>	500,8689
<i>préparer</i>	280,3858
<i>commander</i>	272,1281
<i>thermos</i>	233,1507
<i>bol</i>	217,4931
<i>lait</i>	184,0082
<i>servir</i>	172,8869
<i>offrir</i>	156,5519
<i>brûlant</i>	151,7094
<i>noir</i>	148,7638
<i>chaud</i>	145,8034
<i>avalier</i>	131,2046
<i>siroter</i>	125,0640
<i>verser</i>	122,6979
<i>froid</i>	97,9720
<i>apporter</i>	97,8613
<i>prendre</i>	97,3587
<i>odeur</i>	72,6861
<i>proposer</i>	30,3585
<i>autre</i>	29,8140

<i>bon</i>	25,8876
<i>petit</i>	20,1379
<i>chercher</i>	10,9050

Tableau 2 Les collocatifs du pivot *café* dans POL avec leur LLR

Le collocatif le plus spécifique est le nom *tasse* ( LLR 1733,3311, ex.14).

14) Mathilde fixe sa **tasse de café**, retenant ses larmes du mieux qu'elle peut. (K. Giebel, *Juste une ombre*, 2012)

Comme dans le sous-corpus POL, on observe une majorité de collocatifs verbaux pour le pivot *café* dans SENT : 3 collocatifs verbaux (*boire, prendre, apporter*) et 2 collocatifs nominaux (*tasse, bol*). Le collocatif le plus spécifique est aussi le nom *tasse* (LLR 980,8135, ex.15).

15) Il se leva pour se resservir une **tasse de café**, regarda l'heure à la montre du four et monta en courant se préparer. (M. Levy, *Mes amis, mes amours*, 2006)

Dans le sous-corpus GEN, le pivot *café* apparaît avec des collocatifs plutôt verbaux : 10 collocatifs verbaux (*boire, préparer, commander, servir, prendre, apporter, avaler, verser, offrir, proposer*), 5 collocatifs nominaux (*tasse, lait, bol, gorgée, odeur*) et 5 collocatifs adjectivaux (*noir, brûlant, chaud, fort, petit*). Le collocatif le plus spécifique pour le pivot *café* dans ce sous-corpus est encore le nom *tasse* (LLR 2017,2237, ex.16)

16) Charles, lui, dort jusque vers deux heures et demie, se réchauffe une **tasse de café**, traînasse un peu au lit avant de se laver et de s'habiller. (G. Perec, *La Vie Mode D'Emploi*, 1978)

Notre analyse montre la surreprésentation de collocatifs statistiquement significatifs verbaux dans les sous-corpus GEN, POL et SENT.

### **1.6. Les collocatifs du pivot « thé »**

L'analyse des collocatifs du pivot *thé* n'est pas possible dans les sous-corpus FY, HIST et SF car les données ne répondent pas au critère de dispersion. Dans GEN, nous trouvons le plus de collocatifs statistiquement significatifs (15 collocatifs). Ce nombre est inférieur dans POL (6 collocatifs) et SENT (4 collocatifs).

Dans le sous-corpus POL, le pivot *thé* apparaît avec les collocatifs plutôt verbaux : 4 collocatifs verbaux (*boire, préparer, servir, prendre*) et 2 collocatifs nominaux (*tasse* et

*gorgée*). Le nom *tasse* est le collocatif le plus spécifique pour le pivot *thé* dans ce sous-corpus (LLR 759,1279, ex.17).

17) Enfin ils se seraient réfugiés dans une brasserie, réchauffant leurs mains autour d'une **tasse de thé**. (Ch. Guillaumot, *Chasses à l'homme*, 2008).

Le pivot *thé* apparaît dans la même mesure avec les collocatifs nominaux et verbaux dans le sous-corpus SENT. Nous trouvons 2 collocatifs verbaux (*boire, prendre*) et 2 collocatifs nominaux (*tasse, salon*). Le collocatif le plus spécifique pour le pivot *thé* dans le sous-corpus SENT est le nom *tasse* (LLR 601,4418, ex.18).

18) Mélie accepta la **tasse de thé** que lui offrit Mozart alors que le Livreur se tassait dans les coussins écrus. (M.-S. Kesteman, *Le livreur*, 2014)

Le pivot *thé* apparaît avec les collocatifs plutôt verbaux dans le sous-corpus GEN. On trouve 7 collocatifs verbaux (*boire, servir, préparer, verser, prendre, apporter, offrir*), 5 collocatifs nominaux (*tasse, menthe, salon, gorgée, verre*) et 2 collocatifs adjectivaux (*vert, chaud*). Le nom *tasse*, comme dans les sous-genres précédents, est le collocatif le plus spécifique pour le pivot *thé* dans le GEN (LLR 1360,9462, ex.19).

19) Les jambes croisées, nous pourrions en discuter mondainement autour **d'une tasse de thé** puisque tel est l'usage en Angleterre quoique pour tout dire et excusez mon aveu ! (A. Cohen, *Les Valeureux*, 1969)

Pour résumer les résultats de recherche de cette section, nous remarquons que le collocatif nominal *bouteille* est le seul collocatif qui apparaît dans chaque de sous-corpus avec les noms de boissons alcoolisées (*vin, champagne, bière, vodka*). Pour les noms de boissons non alcoolisées (*thé, café*), le collocatif *tasse* apparaît le plus spécifique dans chaque sous-genre. En ce qui concerne les collocatifs verbaux, le collocatif *boire* est le plus spécifique, sauf pour le pivot *champagne* dans SENT et le pivot *vodka* dans POL et GEN. L'analyse des données montre que seulement le pivot *vin* a des collocatifs statistiquement significatives dans tous les sous-corpus. Les pivots *vin* et *café* ont le plus de collocatifs dans les sous-corpus. En revanche, on observe que le pivot *vodka* a le moins de collocatifs dans les sous-corpus.

Le fait que le vin soit la boisson la plus courante dans le roman est étroitement lié à l'identité nationale française. Selon Barthes (1972 : 83) le vin peut être qualifié comme « Boisson totem », une sorte de symbole de la France. Un facteur important est l'atmosphère d'appartenance nationale qui se manifeste par l'acte de boire du vin qui « renforce le

sentiment d'être français et d'appartenir à une communauté régionale et nationale, à un terroir, à une civilisation » (Cornot, 2012: 1423). Nourrisson décrit le lien fort entre la France et la consommation d'alcool comme suit: « une civilisation qui célèbre le vin et l'amour du boire » (Nourrisson, 1988: 503). Nous voyons ainsi à quel point l'esprit français se mêle à l'acte de boire du vin. Nous n'observons pas une telle préférence pour aucun des autres alcools en ce qui concerne la littérature française. Nourrisson remarque une tendance aux XIXe et XXe siècles : « les consommations de vin ne représentent que 60 à 70% du total, tandis que celles de spiritueux figurent pour plus de 20 % » (Nourrisson, 2013b). La différence de quantité d'alcool consommée entre ces deux types peut être significative pour leur présence dans les romans. Cette dépendance peut être liée aussi au très fort mouvement antialcoolique qui a eu lieu en France après la Seconde Guerre mondiale (Nourrisson, 2013a). Ces tendances et le facteur de nationalité et d'identité font du vin l'alcool le plus répandu dans la littérature, tandis que d'autres types d'alcool ont une distribution relativement plus petite.

## ***2. Étude des collocatifs par sous-genre romanesque***

Nous trouvons le plus grand nombre de collocatifs concernant les noms de boissons dans GEN (64 collocatifs). Au contraire, le plus petit nombre de collocatifs statistiquement significatifs pour les noms de boissons se trouve dans SF (2 collocatifs). Dans cette section, nous allons détailler les données sur la répartition des collocatifs dans chacun des sous-corpus.

Dans la première section de ce chapitre nous allons présenter les données concernant la fréquence des collocatifs par sous genre. La première sous-section montre les collocatifs dans FY, SF et HIST- les sous-genres qui comportent un seul pivot avec des collocatifs. La section suivante présente le cas de SENT dans lequel nous trouvons plusieurs pivots avec les collocatifs. Finalement, dans la dernière sous-section nous nous focalisons sur les sous-genres comportant tous les pivots avec des collocatifs – POL et GEN. La deuxième section nous montre la répartition des collocatifs par catégorie grammaticale dans les sous-genres. Dans la dernière section nous traitons des collocatifs spécifiques dans un seul sous-genre.

### ***2.1. La fréquence des collocatifs par sous-genre***

#### **2.1.1. Le cas de FY, SF et HIST : le sous-genre comporte un seul pivot avec des collocatifs**

Dans le sous-corpus FY, on peut distinguer seulement les collocatifs statistiquement significatives du pivot *vin* (4 collocatifs : *verre, bouteille, coupe, boire*). Pour les autres noms de boissons nous ne pouvons pas mener une analyse. Nos critères n'ont pas permis d'extraire des collocatifs spécifiques pour les pivots *champagne, bière, vodka, café* et *thé*.

Comme dans FY, nous trouvons seulement les collocatifs spécifiques pour le pivot *vin* dans HIST (4 collocatifs : *boire, verre, bouteille, blanc*). Nous ne pouvons pas mener notre recherche à défaut de collocatifs statistiquement significatives des pivots *champagne, bière, vodka, café* et *thé*.

Nous ne trouvons que les collocatifs du pivot *vin* (2 collocatifs : *bouteille, verre*) dans SF. Pour d'autres pivots nous ne pouvons pas mener une analyse parce que nos critères n'ont pas permis d'extraire des collocatifs spécifiques.

Notre analyse montre que dans le sous-corpus présentés ci-dessus, un seul pivot de notre liste de boissons a des collocatifs statistiquement significatifs.

#### 2.1.2. Le cas de SENT : le sous-genre comporte plusieurs pivots avec des collocatifs

Dans le sous-corpus SENT nous trouvons plusieurs pivots avec leurs collocatifs comme montre la figure 2 :

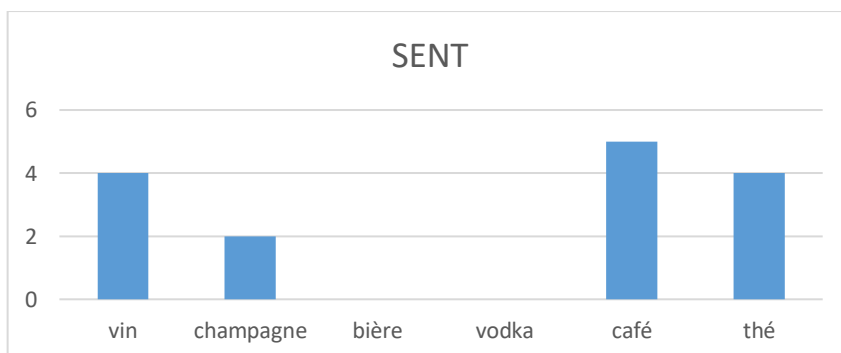


Figure 2 La distribution des collocatifs dans SENT

C'est le pivot *café* qui a le plus de collocatifs dans SENT (5 collocatifs). Comme on peut observer dans la figure 2, SENT présente plus de collocatifs pour les noms de boissons non-alcoolisées. Nous ne trouvons pas de collocatifs statistiquement significatifs pour les pivots *bière* et *vodka*.

#### 2.1.3. Le cas de GEN et POL : le sous-genre comporte tous les pivots avec des collocatifs

La figure 3 montre la distribution des collocatifs de pivots dans POL :

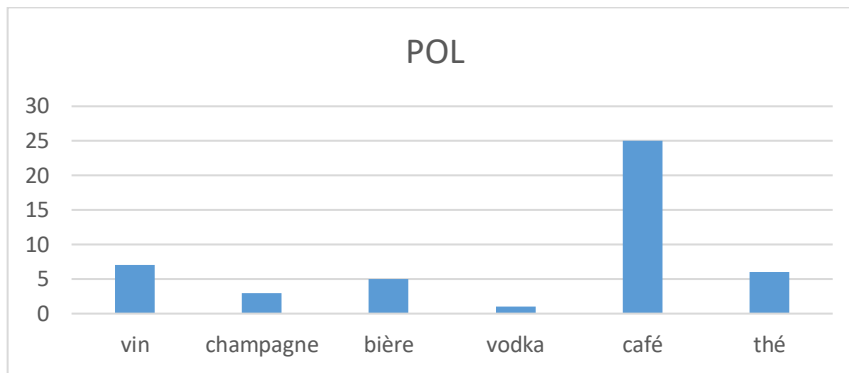


Figure 3 La distribution des collocatifs dans POL

Dans ce sous-corpus POL, nous trouvons des collocatifs pour chaque pivot mais le cas du pivot *café* se dégage nettement par le nombre de collocatifs statistiquement significatifs (25 collocatifs) par rapport aux autres pivots (7 collocatifs du pivot *vin*, 3 collocatifs du pivot *champagne*, 5 collocatifs du pivot *bière*, 1 collocatif du pivot *vodka*, et 6 collocatifs du pivot *thé*).

Nous trouverons également plus de collocatifs de noms de boissons non alcoolisées- 31 collocatifs. Les boissons non-alcoolisées ont seulement 16 collocatifs

La figure 4 présente la distribution des collocatifs dans GEN :

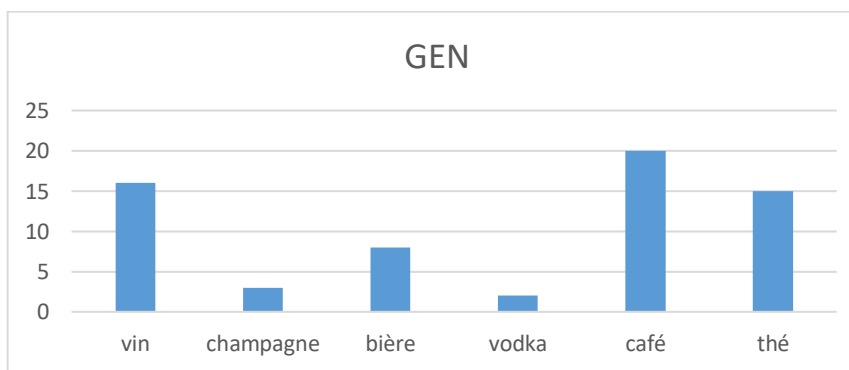


Figure 4 La distribution des collocatifs dans GEN

L'analyse du sous-corpus GEN montre que le pivot *café* apparaît avec le plus grande nombre de collocatifs statistiquement significatifs (20 collocatifs). Comme nous pouvons voir sur la figure 4, dans GEN les collocatifs de noms de boissons non alcoolisées sont plus nombreux- 35 collocatifs. Les boissons non-alcoolisées ont 29 collocatifs.

En observant les données présentées dans la deuxième section, nous pouvons arriver à la conclusion que les collocatifs des noms de boissons sont distribués à un degré différent dans le sous-corpus. Dans le cas des sous-corpus FY, HIST et SF, il n'y a qu'un seul pivot avec des collocatifs statistiquement significatifs. Dans le sous-corpus SENT, on peut trouver des collocatifs statistiquement significatifs pour plusieurs pivots. La situation la plus diversifiée est observée dans les sous-corpus GEN et POL, dans lesquels tous les pivots ont des collocatifs spécifiques statistiquement significatifs.

## 2.2. La répartition des collocatifs par catégorie grammaticale dans les sous-genres

Nous avons aussi mené l'analyse sur les collocatifs selon leur catégorie grammaticale (nominale, verbale ou adjectivale) dans chaque sous-corpus comme le montre la figure 5

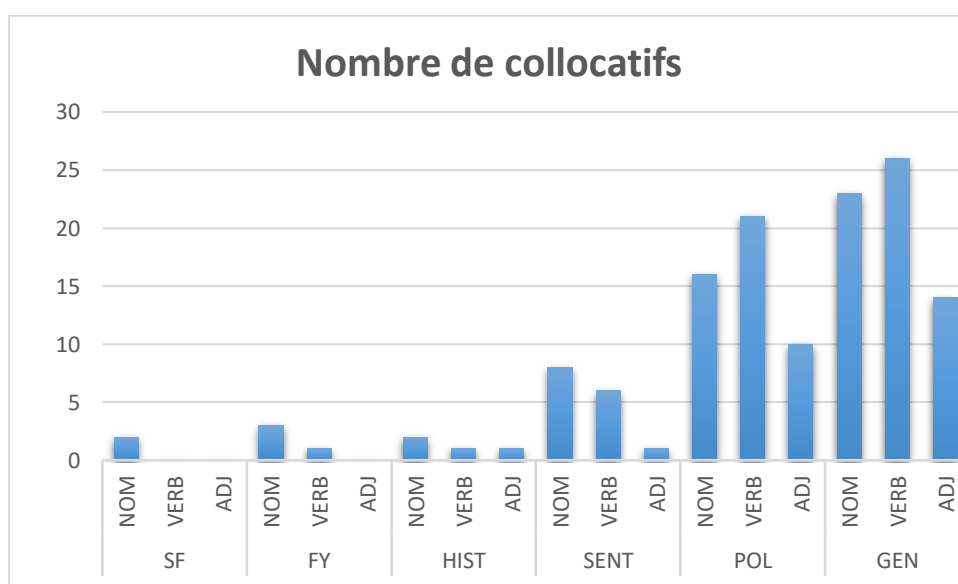


Figure 5 Répartition des collocatifs par catégorie grammaticale dans chaque de sous-corpus

Les données montrent la surreprésentation de collocatifs nominaux dans 4 sous-corpus : SF(2 collocatifs), FY(3 collocatifs), HIST(2 collocatifs) et SENT(8 collocatifs). Les collocatifs verbaux sont les plus présents dans GEN et POL (21 collocatifs dans POL et 26 collocatifs dans GEN) et sont absents de SF. Les collocatifs adjectivaux de noms de boissons sont les moins fréquents dans chaque de sous-genre et sont inexistantes dans SF et dans FY.

### 2.3. Les collocatifs spécifiques dans un seul sous-genre

Nos observations nous permettent de distinguer les collocatifs de pivots qui sont statistiquement significatifs pour un seul sous-corpus de notre travail. Sur la base de nos observations des données (voir les tableaux en annexes), nous avons relevé les collocatifs exclusifs suivants pour seulement deux sous-genres (GEN et POL) :

a) GEN : *vin* + *verser, chaud, goûter, mauvais, marchand, frais, pain, servir, apporter*

*bière* + *canette, verre, frais, servir*

*vodka* + *verre*

*thé* + *menthe, vert, verser, verre, apporter, chaud, offrir*

b) POL : *bière* + *prendre*

*café* + *siroter, froid, autre, chercher, bon*

Notre classification des données nous a permis de distinguer les collocations qui ne sont statistiquement significatives que pour des sous-corpus sélectionnés. Comme on peut le voir, on trouve des collocations statistiquement significatives qui n'apparaissent dans aucun autre sous-corpus sauf les sous-corpus POL et GEN. En tant que plus grand de notre sous-corpus, GEN possède le plus grand nombre de collocations que l'on ne trouve que dans celui-ci.

Ce chapitre 7 avait pour but de présenter une approche quantitative de nos données. Nous avons présenté les données sur les collocatifs nominaux, verbaux et adjectivaux de pivots qui sont les noms de boissons. Nous avons montré les collocatifs les plus spécifiques pour chaque de pivot dans chaque de sous-corpus. Cette partie du travail nous a permis de faire les premières observations sur les collocatifs.

La recherche sur les phraséologismes autour des noms de boissons est quasiment impossible dans le sous-corpus du roman SF, car il n'y a que des collocatifs statistiquement significatifs pour le pivot *vodka*. Nous ne trouvons pas d'autres collocatifs statistiquement significatifs pour les autres noms de boissons dans ce sous-corpus. La situation est similaire avec les sous-corpus HIST et FY, on n'y trouve que des collocatifs pour pivot *vin*. Dans le sous-corpus SENT, notre recherche a révélé que certains de nos pivots sélectionnés ont des collocatifs statistiquement significatifs. Néanmoins, dans les sous-corpus POL et GEN, on remarque la structure de collocations la plus diversifiée autour des noms de boissons - dans ces deux sous-corpus, chacun des pivots possède des collocatifs statistiquement significatifs.



Dans les sous-corpus avec un plus petit nombre de collocatifs statistiquement significatifs, on remarque la surreprésentation des collocatifs nominaux. Dans le même temps, les sous-corpus avec le plus grand nombre de collocatifs de noms de boissons se caractérisent par une surreprésentation de collocatifs verbaux. De plus, dans les sous-corpus GEN et POL, nous remarquons des collocations statistiquement significatives que l'on ne retrouve que dans ces sous-corpus : dans GEN, ce sont des collocations pour 4 pivots, en POL 2 pivots. Sur la base de ces résultats, on constate que les collocations autour des noms de boissons n'apparaissent pas au même degré dans chaque sous-corpus. Les sous-corpus contiennent des collocatifs préférentiels d'une catégorie grammaticale donnée. Le fait que certains sous-corpus soient les seuls à avoir certaines collocations statistiquement significatives prouve qu'une recherche plus approfondie peut contribuer au processus de distinction des genres romanesques.

Dans le chapitre suivant, nous proposerons une approche plus qualitative de nos données. Nous nous focaliserons sur deux expressions spécifiques à un sous-genre par rapport à un autre sous-genre. Nous allons mener notre recherche en analysant le fonctionnement linguistique des deux expressions choisies que nous trouvons dans les sous-corpus de notre recherche.

## Chapitre 8. La combinatoire des noms de boissons : approche qualitative (*boire du vin* et *prendre une bière*)

Dans ce chapitre, notre objectif est d'analyser deux constructions spécifiques, leur fonctionnement linguistique et l'entourage dans lequel elles apparaissent le plus souvent. Sur la base des statistiques présentées dans le chapitre précédent, nous avons sélectionné les deux expressions *boire du vin* et *prendre une bière*. Nous avons choisi ces collocations parce qu'il nous paraît intéressant d'étudier une collocation très fréquente (*boire du vin*) mais aussi une collocation très spécifique (*prendre une bière*) qui peuvent nous aider distinguer les sous-genres romanesques.

### 1. L'analyse de la collocation « *boire du vin* »

Cette partie s'inscrit dans la continuité des recherches menées par Gonon et Sorba (2019) sur les phraséologismes autour du vin et du champagne.

Tout d'abord, nous analysons la collocation *boire du vin* qui est très spécifique à tous nos sous-corpus, à l'exception de SF. La figure 6 montre la valeur LLR de la combinaison du pivot *vin* avec le collocatif *boire* dans chaque de sous-corpus. Le LLR de cette collocation est le plus élevé dans GEN tandis que dans SENT on trouve la valeur la moins élevée :

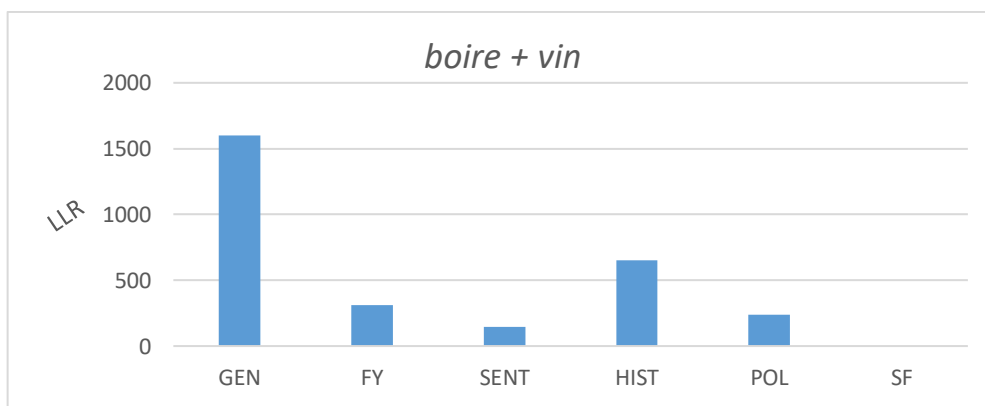


Figure 6 Le calcul de spécificité (LLR) de la collocation *boire+vin* dans les sous-corpus

La collocation *boire+vin* n'est pas également distribuée dans les sous-corpus :

Sous-corpus	Nombre d'occurrences

GEN	302
HIST	125
FY	60
POL	55
SENT	37

Tableau 3 La distribution de la collocation *boire +vin* dans les sous-corpus

Le tableau 3 montre que l'expression *boire+vin* apparaît le plus souvent dans le sous-corpus GEN. Cette collocation est le moins souvent dans SENT, en excluant le sous-corpus SF dans lequel le collocatif *boire* n'est pas statistiquement significatif pour le pivot *vin*. Dans les parties suivantes de ce sous-chapitre nous nous focalisons sur les variations linguistiques de la combinatoire du pivot *vin* avec le collocatif *boire* dans GEN. Néanmoins, dans chaque partie que nous aborderons, nous comparerons les données des autres sous-corpus<sup>10</sup>.

Dans cette section, premièrement nous allons faire le point sur la variation sur le déterminant dans la collocation *boire du vin*. Ensuite, nous présenterons les variations temporelles et modales du verbe *boire*. Finalement, nous nous focaliserons sur les personnes verbales de la construction *boire du vin*.

### 1.1. La variation sur le déterminant

Nous observerons dans cette partie les variations de la collocation *boire+DET+vin* dans le sous-corpus GEN. En ce qui concerne la détermination, nous constatons que dans la grande majorité de cas, le nom *vin* privilégie légèrement l'article partitif *du*. La construction *boire du vin* apparaît 128 fois (48% ) dans le sous-corpus GEN (ex.20).

20) Les gens riaient, parlaient, **buvaient du vin** qu'ils sortaient de leur sac dans le car.(E. Ionesco, *Le Solitaire*, 1973)

La figure 7, qui est une visualisation de l'un des arbres lexico-syntaxiques, correspond à l'expression *boire du vin*.

---

<sup>10</sup> Les diagrammes concernant la répartition des temps verbaux et des personnes grammaticales dans chaque sous-corpus romanesque sont disponibles en annexe.

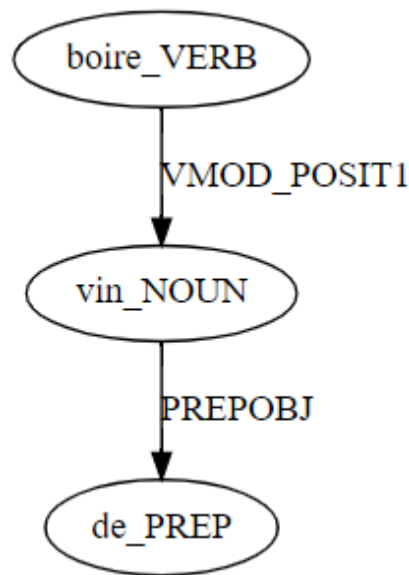


Figure 7 L'arbre lexico-syntaxique de *boire du vin*

Ce déterminant est également le plus courant dans la construction *boire du vin* dans d'autres sous-corpus. Il se classe premier en termes de pourcentage de tous les déterminants utilisés aussi dans le sous-corpus FY (26 occurrences), HIST (59 occurrences) et POL (21 occurrences) dans lesquels il représente 40-50% des occurrences des déterminants. Le déterminant *du* est également le déterminant le plus fréquent de la construction *boire du vin* dans le sous-corpus SENT (9 occurrences). Néanmoins, dans ce sous-corpus donné, il ne constitue que 26% de tous les articles, ce qui diffère de l'occurrence moyenne dans les autres sous-corpus.

L'article partitif apparaît aussi dans la variante *de*. Dans le sous-corpus GEN, nous trouvons 59 d'occurrences où le nom *vin* est déterminé par le déterminant partitif *de*. On peut isoler cette variante d'abord dans les constructions suivantes qui ont toutes une tournure négative :

- *ne boire pas de vin* (8 occurrences) (ex.21)
- *ne boire plus guère de vin* (1 occurrence)
- *ne boire plus de vin* (1 occurrence)
- *ne boire pratiquement jamais de vin* (1 occurrence)
- *n'avoir jamais bu de vin* ( 1 occurrence) (ex.22)

Dans ces cas, nous parlons du déterminant partitif devant un nom massif. Cette structure c'est la construction syntaxique prototypique en français.

21) Il **ne boit pas de vin** à cette heure. (P. Cintas, *Carabin Carabas*, 2016)

22) Elle **n'avait jamais bu de vin**, le roi n'admettant que l'eau à table. (Ch. Rochefort, *Oeuvre romanesque*, 2004)

En analysant les contextes gauche et droit de l'expression *boire de vin*, nous observons que cette construction peut aller avec un autre déterminant. Elle est dans certains cas accompagnée de déterminants complexes. Nous trouvons 8 structures qui accompagnent notre collocation 43 fois. Les déterminants complexes dans ces cas apparaissent dans la combinaison d'un déterminant quantifiant et d'un déterminant référent en indiquant l'intensité de l'acte de boire. Nous distinguons les structures suivantes :

- *un peu de vin* (16 occurrences)
- *trop de vin* (9 occurrences)
- *beaucoup de vin* (8 occurrences)
- *tant de vin* (4 occurrences)
- *autant de vin* (3 occurrences)
- *tellement de vin* (1 occurrence)
- *encore moins de vin* (1 occurrence)
- *très peu de vin* (1 occurrence)

Voici quelques exemples issus du sous-corpus GEN :

23) Je racontais des histoires, je me taisais un instant, je **buvais un peu de vin**, je soulevais l'énorme bouteille pour en verser à son tour dans le verre de Béatrice. (J. D'Ormesson, *Un amour pour rien*, 1960)

24) Elle **but beaucoup de vin** mandarin et fut très joyeuse (A. Nothomb, *Biographie de la Faim*, 2004)

25) Ça fait que vous êtes obligé de **boire trop de vin**, et vous finirez par mourir à la peine. (M. Pagnol, *L'eau des collines 1. Jean de Florette*, 1963)

Les variations sur le déterminant partitif et complexe *de* sont également largement présentes dans les autres sous-corpus. Nous trouvons la construction *boire de vin* dans FY (13 occurrences, 23%), POL (15 occurrences, 29%), HIST (15 occurrences, 13%) SENT (8 occurrences, 23%). Nous pouvons remarquer que le sous-corpus HIST a statistiquement le moins de constructions *boire de vin*, moins de 20%. Dans le sous-corpus SENT, cependant, les déterminants partitif *de* et complexe avec *de* sont distribués presque au même niveau que le déterminant partitif *du* (seulement 3% de différence).

Nous pouvons remarquer que dans certains cas le pivot *vin* choisi aussi l'article défini *le* avec le collocatif *boire*. Dans notre sous-corpus, nous observons 40 occurrences (14%) de la structure *boire le vin*. La construction *boire+DET+vin* se rencontre aussi avec l'article indéfini *un / des*. Dans notre sous-corpus GEN, nous trouvons 18 (5,96%) occurrences de la collocation *boire un vin* et 5 occurrences de la structure *boire des vins* (1,65%). En observant les données des corpus, nous avons constaté que la construction *boire un vin* est très souvent suivi d'un adjectif postposé indiquant son type ou son trait caractéristique. Les expressions sont accompagnées de 8 adjectifs différents (ex.26 et 28) : *inépuisable, blanc, coupé, aromatisé, chaud, nouveau, rouge, léger*. On distingue également 2 variations avec un groupe adverbe-adjectif : *très doux* et *jamais bu*. Nous notons aussi cette construction suivie de la préposition *de* avec le lieu d'origine du vin (ex.27) : *d'Espagne, de Provence, du Languedoc*, et aussi 1 occurrence de la construction dans laquelle adjectif est suivi d'un nom de lieu d'origine : *rouge de Savoie*.

26) Pendant ce temps, Renata nous versait à **boire un vin rouge**, des plus épais et *robotatifs*. (D. Fernandez, *Pise 1951*, 2011)

27) Nous **boirons un vin de Provence**. (P. Cintas, *Rendez-vous des fées*, 2016)

28) Il y avait également les autres qui, poussés par le froid, y avaient trouvé refuge et qui, **en buvant un vin chaud**, attendaient patiemment que la neige se calme pour rentrer chez eux. (P. Combescot, *Les filles du Calvaire*, 1991)

Comme pour les autres sous-corpus, nous notons une variation de la construction *boire du vin* avec un déterminant indéfini *un/des* dans les sous-corpus FY (2 occurrences, 4%), POL (4 occurrences, 8%), SENT (4 occurrences, 12%), HIST (11 occurrences, 9%). Nous pouvons constater que la plus petite part des déterminants indéfinis est caractéristique pour le sous-corpus FY. C'est dans le sous-corpus SENT que nous trouvons le pourcentage

le plus élevé de déterminants indéfinis dans la distribution de tous. Dans le cas d'un déterminant défini *le*, la répartition de la présence dans les autres sous-corpus est la suivante : FY (7 occurrences, 13%), POL (7 occurrences, 13%), SENT (7 occurrences, 21%), HIST (23 occurrences, 19%). Ces résultats montrent que les sous-corpus SENT et HIST ont le pourcentage le plus élevé de ce déterminant par rapport aux autres sous-corpus.

Les données issues du sous-corpus GEN montrent aussi une variation non négligeable mais qui porte sur un petit nombre d'occurrences concernant le déterminant possessif utilisé dans la construction *boire+vin*. Notre analyse nous permet de repérer le nombre d'occurrences suivant : *son vin* (14 occurrences- 4,63%, ex.29), *leur vin* (3 occurrences, 0,99%, ex.30), *mon vin* (2 occurrences, 0,66%), *notre vin* (2 occurrences, 0,66%), *votre vin* (1 occurrence, 0,33%). Comme nous pouvons observer, cette variation apparaît plutôt avec la 3<sup>e</sup> personne.

29) Il **but son vin**, alluma une cigarette tirée du paquet que M. Robin lui avait laissé, puis il regarda le réveil. (B. Clavel, *La grande patience. Les fruits de l'hiver*, 1968)

30) Les hommes **buvaient leur vin** aussitôt servi, un devoir, et ils s'en allaient chez eux, pressés. (M. Duras, *Moderato cantabile*, 1958)

Les variantes de la construction *boire du vin* sont aussi observables avec les déterminants possessifs dans les autres sous-corpus : FY (7 occurrences, 12%), POL (2 occurrences, 4%), SENT (5 occurrences, 15%), HIST (10 occurrences, 8%). Notre analyse montre que dans le sous-corpus POL nous trouvons le moins de pronoms possessifs, tandis que dans le sous-corpus SENT ils constituent un pourcentage important.

En menant l'analyse sur notre sous-corpus, nous avons remarqué la variation de la construction *boire+vin* avec un déterminant démonstratif au singulier et au pluriel (4% d'occurrences). Nous trouvons 9 occurrences de la variation *boire ce vin* (ex.31) et 1 occurrence de la variation *boire ces vins*. Voici un exemple illustrant cette construction :

31) Ils **ont bu ce vin** clair et tout frais monté de la cave dans un pot couvert de buée. (B. Clavel, *Pirates du Rhône*, 1957.)

D'autres sous-corpus contiennent également une variation de construction *boire du vin* avec un déterminant démonstratif, mais ce sont des quantités insignifiantes. Ils

apparaissent dans les proportions suivantes: FY (1 occurrence, 2%), POL (3 occurrences, 6%), SENT (1 occurrence, 3%), HIST (2 occurrences, 2%). Comme on peut le voir, le déterminant démonstratif constitue un pourcentage relativement faible de tous les déterminants de sous-corpus étudiés. Le sous-corpus POL a la part de pourcentage la plus élevée, mais elle n'est que de 6%, ce qui n'est pas un nombre important d'occurrences.

Dans cette section, nous avons analysé les variations de l'article dans la construction *boire du vin* dans GEN en comparant avec les autres sous-corpus.. Nos résultats indiquent que cette construction montre une forte préférence pour l'article partitif *du* (128 occurrences dans GEN). Néanmoins, dans 58 occurrences de GEN, nous trouvons une variation avec l'article partitif *de*. On retrouve également une variante avec un article complexe (43 occurrences dans GEN). L'article défini *le* se trouve dans les 40 occurrences de la construction *boire du vin* dans GEN. Le déterminant possessif apparaît dans 8 occurrences, tandis que le déterminant démonstratif dans 10 occurrences dans ce sous-corpus.

Il faut noter que la variation *boire du vin* est la plus courante dans tous les sous-corpus. Cependant, la répartition des différentes variations de déterminants dans la construction *boire du vin* n'est pas au même niveau dans tous les sous-corpus. Nous avons remarqué certains faits à mentionner dans notre analyse. Le premier est que seulement dans le sous-corpus HIST le déterminant partitif n'occupe pas la deuxième place par rapport à son occurrence dans la structure étudiée - dans ce cas, le deuxième article le plus fréquemment utilisé est le déterminant défini *le*. Un phénomène à noter est une structure de distribution des déterminants complètement différente dans SENT que dans les autres sous-corpus. Alors que dans d'autres sous-corpus, la distribution des différents types de déterminants est assez comparable, il y a plus de 40% de la quantité du déterminant *du*, et les autres représentations des déterminants à un niveau similaire, la distribution dans SENT semble assez différente. Les trois variations principales et les plus courantes se produisent dans les montants suivants: *boire du vin* (26%), *boire de vin* (23%), *boire le vin* (21%). On peut voir ici une distribution assez homogène du nombre d'occurrences de déterminants analysés dans SENT qui interviennent à peu près dans les mêmes proportions. Les différences caractéristiques dans la distribution des déterminants dans les sous-corpus peuvent, dans une certaine mesure, être l'une des manières d'approcher le problème de la distinction de genres romanesques.



## 1.2. Les variations temporelles et modales du verbe « boire »

En analysant les variations temporelles du verbe *boire* dans la construction *boire du vin* dans GEN, nous pouvons remarquer quelques préférences concernant l'emploi des temps grammaticaux (voir figure 8).

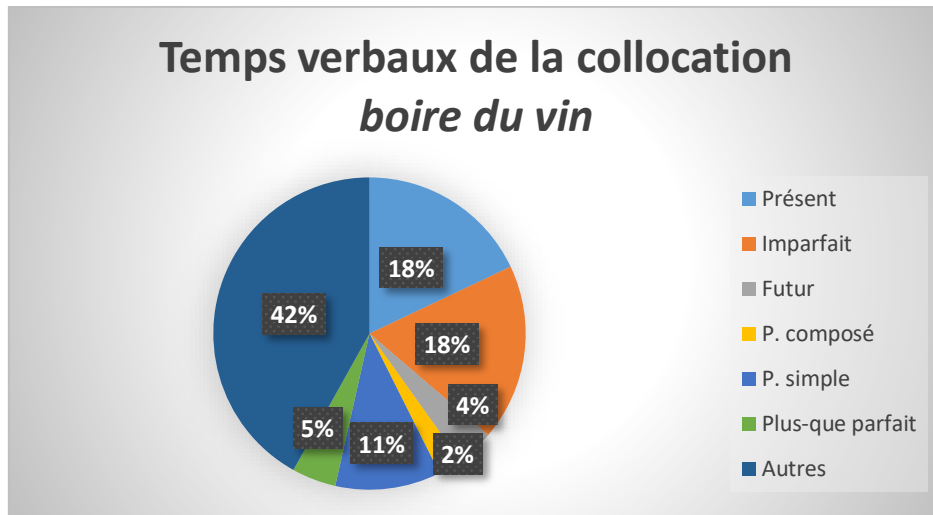


Figure 8 Répartition des temps verbaux pour la construction boire du vin dans GEN

Nous constatons que *boire du vin* privilégie légèrement le présent (18%) et l'imparfait (18%). Cette utilisation fréquente de ces deux temps est conditionnée par leurs valeurs dans le texte.

En effet, l'imparfait est considéré comme un temps imperfectif. Il montre le procès de l'intérieur de l'action, dans son déroulement. Par conséquent, l'imparfait n'assigne pas de bornes temporelles évidentes de l'acte de boire du vin. L'imparfait signale que le procès est en cours au moment choisi comme point de repère. Sans l'indication du moment d'achèvement de l'action, nous considérons l'imparfait comme le temps non limitatif. L'usage narratif de l'imparfait indique une valeur d'arrière-plan.

Voici quelques exemples issus de sous-corpus GEN :

32) Au village, on jouait au yoyo et **on buvait du vin** dans les cafés au lieu de cidre.

(A. Ernaux, *La place*, 1983)

33) Hélène **buvait du vin** blanc et les regardait danser ; après, Thomas mit une musique lente, et elle m'invita à danser avec elle. (J. Littell, *Les Bienveillantes*, 2006)

34) Et moi, le soir, je m'enfermais dans ma chambre, je **buvais du vin**, j'abreuvais la bête.  
(G. Delacourt, *On ne voyait que le bonheur*, 2014)

Comme le montrent les exemples(ex. 32-34), le fait de boire du vin n'est pas la seule activité des personnages. Il apparaît généralement en combinaison avec d'autres activités. En règle générale, le fait de boire du vin n'est pas la seule activité exercée par les personnages

En ce qui concerne le présent, le plus souvent il apparaît comme le présent de narration. Un narrateur qui veut recréer le plus fidèlement possible l'immédiateté de l'expérience sur la pratique du boire du vin (ex.36). Nous pouvons considérer ce temps verbal comme une alternative aux narrations construites au passé simple. Comme pour l'imparfait, l'acte de boire du vin dans le présent est souvent empilé avec une autre activité en cours (ex.37).

Voici les exemples :

35) Et nous **buvons le vin** millénaire en nous inclinant l'un vers l'autre, cérémonieusement.  
(A. Langfus, *Les bagages de sable*, 1962)

36) Puis il reprend sa place, et **boit le vin** à petites gorgées tremblantes. (Y.Berger, *Le sud*, 1962)

37) Ils **boivent du mauvais vin**, disent n'importe quoi puis oublient. (T. Ben Jelloun, *Amours sorcières*, 2003)

Ainsi, l'utilisation de l'imparfait et du temps présent implique une activité en cours, généralement accompagnée d'une autre activité. L'acte de boire du vin exprimé dans ces deux temps n'a pas de délai, il est indéfini dans le temps, on peut donc dire que c'est un élément du récit qui décrit le contexte des événements.

En comparant GEN à d'autres sous-corpus, nous pouvons voir que dans certains cas, la structure de la distribution des temps grammaticaux semble être similaire à celle de GEN. La structure la plus proche apparaît dans HIST, où, comme dans le cas de GEN, les temps personnels, le présent (11%) et l'imparfait (14%) sont les plus présents dans la construction *boire du vin*. Néanmoins, le passé simple est également présent (9%), mais il ne représente pas un pourcentage inférieur de la distribution de temps. Dans le cas de SENT également, les temps présent (25%) et imparfait (19%) prévalent. La présence de la construction *boire*

*du vin* dans SENT au passé composé (11%) et au passé simple (11%) est plus petite. Dans le sous-corpus POL, on observe une préférence pour le passé composé (17%). Il est intéressant de noter que le présent (12%), l'imparfait (12%) et le passé simple (12%) sont répartis dans la même mesure. Le sous-corpus FY montre une structure préférée différente - le temps personnel le plus observé est l'imparfait (18%), le temps présent ne représente que 3% ce qui est une différence significative en comparant avec les autres sous-corpus. Les temps passés simple (10%) et composé (12%) ne sont pas une petite partie de la présence de temps dans la construction *boire du vin* dans FY.

Comme nous l'avons montré au-dessus, la collocation *boire du vin* apparaît majoritairement à un mode personnel. Néanmoins, la rubrique « Autres » sur la figure 8 ci-dessus fait référence au mode impersonnel (42%) : l'infinitif *boire* et la forme en *-ant* (*buvant*). La forme *en buvant* représente 8,5% des emplois de la construction *boire du vin*. L'utilisation du gérondif, comme dans le cas de l'imparfait, implique l'accompagnement de l'action se déroulant au premier plan. Dans la section suivante nous allons se focaliser sur les emplois du gérondif *en buvant* avec les autres verbes auxquels le circonstant *en buvant du vin* apporte des précisions.

Le verbe *manger* est un collocatif privilégié (LLR 240,6203, 2 occurrences) du gérondif *en buvant* (ex.40). Un autre verbe lié à l'alimentation est utilisé : *goûter* (LLR56,2822, 1 occurrence). Les verbes *déguster* (1 occurrence) et *dévorer* (1 occurrence) sont aussi présents mais ils ne sont pas statistiquement significatifs pour cette structure. Néanmoins, l'utilisation de ce mode se produit également avec l'utilisation d'autres verbes impliquant une activité individuelle liée à la cognition, comme par exemple *attendre* (ex.39, 2 occurrences), *penser* (1 occurrence), *passer* (1 occurrence), *lire le quotidien* (1 occurrence) etc. En plus, nous pouvons observer l'apparition de verbes relatifs aux activités sociales comme *discuter*, *écouter*, *parler* (ex.38), *s'entretenir* etc (1 occurrence de chaque collocatif).

38) **Ils avaient parlé**, tous les deux, **en buvant un peu de vin**, de leurs guerres et de l'empereur. (J. D'Ormesson, *Histoire du juif errant*, 1990)

39) Assise à l'angle du foyer, **je l'attends en buvant du vin** chaud que Danièle m'a préparé. (F. Tristan. *Le dieu des mouches*, 1959)

40) Attilio avait deux bicyclettes, deux fusils de chasse, un bateau pour aller à la pêche, et à table, **on avait mangé** un gigot de mouton entier, **en buvant du vin** cacheté! (M. Pagnol, *L'eau des collines*. Jean de Florette, 1963)

En analysant les exemples extraits du corpus, nous avons remarqué l'emploi fréquent de la forme infinitive *boire du vin* (43 occurrences). Nous avons noté des extensions aspectuelles sur l'axe syntagmatique dans les constructions. Dans la plupart des cas, les verbes qui accompagnent l'infinitif *boire* sont utilisés à l'imparfait. Nous remarquons les constructions introductrices suivantes de ces temps : *conseiller de, empêcher de, contraindre à, convoquer pour, ne pas devoir, falloir, pouvoir, (ne pas) vouloir, ne pas être l'habitude de, essayer de, se mettre à, rester à, faire*. L'utilisation de l'imparfait contribue au fait que l'action est perçue à l'intérieur de son déroulement. L'imparfait dans ces cas implique un aspect sécant qui donne à la narration un certain flou et l'absence de la limite finale de l'action. D'autres constructions qui ont le même aspect sont *continuer à, ne pas avoir l'âge de*. Ils ont été appliqués au présent. Les constructions *être obligé de, passer la nuit/son temps à, ne pas cesser, devoir*, utilisées au plus-que-parfait, ont plutôt un aspect global dans lequel l'action est perçue de l'extérieur et dans sa globalité. Parmi les verbes listés ci-dessus, on peut également distinguer les semi-auxiliaires lesquels perdent une partie du sens pour exprimer le sens de la construction dans lequel ils apparaissent. Nous trouvons les semi-auxiliaires suivants : *devoir, pouvoir, vouloir, faire*. Certaines des constructions introductrices du verbe *boire* expriment une contrainte, mais dans le contexte ces expressions ne sont pas si fortes, elles sont utilisées plutôt amicalement.

41) **Qu'il faut** se réjouir, faire la fête, se livrer à l'amour, boire du **vin**. (A. Nothomb, *Barbe bleue*, 2012)

42) Il prit ces deux jeunes Maghrébins à parti, et se mit en tête de leur **faire boire du vin** rouge. (M. Tournier, *La goutte d'or*, 1985)

Comme dans GEN, dans d'autres sous-corpus, on peut également observer une forte préférence pour les modes impersonnels, principalement l'infinitif et le gérondif. La distribution de la construction *boire du vin* en mode impersonnel, la plus proche du GEN, est indiquée dans POL (43%) et FY (47%). Il faut noter que HIST (56%) se distingue par le plus grand nombre de la structure *boire du vin* présente en mode impersonnel. Dans les trois sous-corpus mentionnés ci-dessus, l'infinitif est la grande majorité des structures impersonnelles. La situation est différente dans le cas de SENT. Bien que dans ce sous-

corpus les constructions impersonnelles soient également les plus souvent observées, elles ne constituent que 28%, ce qui par rapport aux autres sous-corpus ne représente qu'une petite partie de la distribution de la structure. Il faut également souligner la particularité de la distribution dans le cas de SENT qui est assez homogène en termes de temps et de modes les plus courants. Le mode impersonnel (28%), le présent (25%) et l'imparfait (19%) sont présents à un degré très similaire dans SENT. En le comparant avec d'autres sous-corpus, nous pouvons remarquer que dans leur cas, nous pouvons observer un avantage significatif de la présence du mode impersonnel, alors que dans SENT cet avantage est assez négligeable.

Dans cette section, nous avons présenté la distribution des temps et des modes utilisés dans la construction *boire du vin* présente dans le sous-corpus de nos recherches. Nous avons effectué une analyse détaillée des cooccurrences de la structure *boire du vin* dans GEN. Nous avons ensuite comparé les données de tous les sous-corpus. Nous avons remarqué des différences significatives dans les réalisations de la construction *boire du vin* dans différents sous-corpus. D'après les observations les plus importantes, il convient de noter que les sous-corpus GEN, SENT et HIST montrent une structure similaire de la pour les temps personnels, alors que dans le cas de FY et POL, d'autres structures de la distribution des temps personnels apparaissent. Un autre fait à noter est la surreprésentation significative du mode impersonnel dans HIST par rapport aux autres sous-corpus. Il faut également mentionner une distribution assez caractéristique des temps et des modes dans SENT, qui diffère complètement du modèle observé dans d'autres sous-corpus.

### ***1.3. Les personnes verbales de la construction « boire du vin »***

Un aspect qui mérite d'être étudié en ce qui concerne la consommation de vin dans le corpus GEN est l'utilisation des personnes grammaticales avec la structure *boire du vin*. La répartition des personnes verbales dans ce cas, en d'autres termes, l'identité des buveurs de vin, est présentée dans la figure 9 :

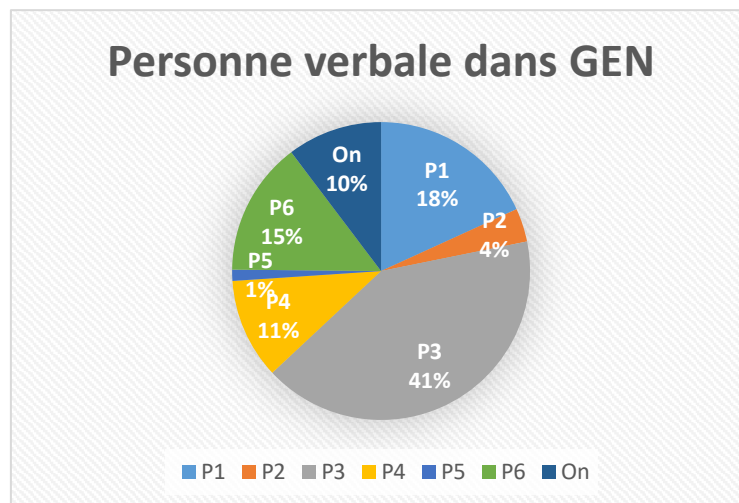


Figure 9 Répartition des personnes verbales pour la construction *boire du vin* dans GEN

En analysant les réalisations de l'ALR *boire du vin* dans GEN, nous remarquons que les personnes du singulier P1, P2 et P3 constituent la grande majorité d'occurrences (63%). Cette représentation pointe vers la consommation plutôt individuelle du vin dans GEN. Cependant, nous devons également faire attention aux personnes P4 (11%), P6 (15%) et *on* (10%) qui introduisent un mode de consommation collectif.

Voici les exemples issus de sous-corpus GEN :

43) **Je bois du vin**, de la bière, je mange surtout des légumes, des pâtes, très peu de viande. (E. Ajar, *Gros calin*, 1974)

44) C'est sans doute l'alcool, **il a bu trop de vin** au repas, l'imbécile. (M. Houellebecq, *Extension du domaine de la lutte*, 1994)

45) **Nous boirons du vin** blanc de l'île bien frais, et du Cheval Blanc 1955 comme s'il en pleuvait. (K. Haedens, *L'été finit sous les tilleuls*, 1966)

Les résultats de nos recherches sur ce point s'alignent largement sur l'analyse de Gonon & Sorba (2019) sur les phraséologismes autour de la consommation du vin et du champagne.

Nous avons également mené une analyse sur l'identité des buveurs de vin désignés par P3. Sur 68 occurrences de la P3, le pronom *il* apparaît dans 48 occurrences (70,6%) pendant que le pronom *elle* seulement dans 20 occurrences (29,4%). On peut donc affirmer que dans le sous-corpus GEN ce sont plutôt les hommes qui boivent du vin :

46) **Il boit** par petites gorgées **le vin** que le gros homme lui sert. (P. Claudel, *La petite fille de Monsieur Linh*, 2005)

47) **Il y buvait le vin** de sa tranquillité, ce qui n'allait pas sans lui causer de menus problèmes, depuis les acidités qui le faisaient claquer des dents jusqu'aux rougeurs qui le démangeaient après le repas, sous les yeux qu'il avait un peu glauques. (P. Cintas, *Coq à l'âne Cocaïne*, 2016)

48) **Elle boira du vin** avec lui, elle lui racontera sa journée. (L. Slimani, *Dans le jardin de l'ogre*, 2014)

Le sous-corpus FY a une prédominance du singulier, comme le GEN, ce qui indique le caractère plutôt individuel de l'acte de boire du vin. Les personnes P1 (12%), P2 (6%) et P3 (55%) ensemble, ils représentent jusqu'à 73% de toutes les représentations de personnes grammaticales dans la construction *boire du vin* dans FY. On observe également le caractère collectif de l'acte de boire du vin dans les P4 (3%), P6 (21%) et *on* (3%) qui ne représente que 27% de la représentation totale des personnes grammaticales dans le cas étudié. En analysant P3 en termes de distribution identitaire des buveurs de vin, nous arrivons à la conclusion que dans la plupart des cas (89%), les hommes sont les personnes qui boivent du vin en FY.

En ce qui concerne la distribution des personnes grammaticales dans POL, nous pouvons remarquer un schéma similaire à celui de GEN et FY. Les personnes P1 (13%), P2 (3%) et P3 (55%) constituent la majorité (71%) de la mise en œuvre de la collocation *boire du vin* sous forme personnelle, ce qui indique un caractère plutôt individuel de cette activité, comme dans le cas de GEN et FY. Néanmoins, les P4 (3%), P5 (3%) et P6 (23%), utilisés dans la construction du *boire du vin*, figurent également dans POL et nous en concluons que l'acte de boire du vin est également collectif et peut être réalisé par beaucoup.

Dans le cas de SENT, on observe une structure de distribution des personnes verbales différente de celle observée dans le sous-corpus précédent. On observe une forte préférence pour P4 (14%), P5 (7%), P6 (21%) et *on* (18%), ce qui indique un caractère plutôt collectif de l'acte de boire du vin dans SENT. Bien que P1 (18%) et P3 (22%) représentent 40% de la mise en œuvre des personnes verbales dans SENT en indiquant un caractère individuel, le vin dans SENT est plutôt consommé par de nombreuses personnes.

Dans HIST, comme dans le cas de SENT, l'analyse a montré que ce sous-corpus privilégie le caractère collectif de l'acte de boire du vin en mettant en œuvre P4 (14%), P5 (3%), P6 (32%) et *on* (10%) qui représentent ensemble 59% de l'ensemble des exemples. De plus, il faut souligner que HIST possède également des réalisations P1 (11%), P2 (3%) et P3 (27%), grâce auxquelles on peut également observer le caractère individuel de la consommation de vin.

Dans cette partie, nous avons traité des variations linguistiques de la collocation *boire du vin* dans GEN en comparant avec les autres sous-corpus. Nos recherches nous ont permis d'observer des différences significatives résultant dans la distribution des variations linguistiques de la collocation *boire du vin* dans les six sous-corpus observés.

Dans un premier temps, nous avons effectué une analyse de la distribution des variations de la construction *boire du vin* avec différents déterminants. Nos recherches ont montré, surtout, que SENT se distingue par une structure assez homogène de la distribution des principales variantes pendant que dans l'autre sous-corpus la variation *boire du vin* constitue la grande majorité de réalisation de cette collocation. Il convient également de noter que HIST se distingue par le fait que la deuxième variante la plus courante est la construction *boire le vin* contrairement aux autres sous-corpus dans lequel c'est la construction *boire de vin*.

L'analyse des temps et des modes grammaticaux nous a également permis de remarquer plusieurs traits pour distinguer les genres. Le premier est la préférence du passé composé présent uniquement en POL. Les sous-corpus restants se distinguent par une préférence pour les temps personnels présents et imparfait. La seule exception est FY où la distribution du temps présent est négligeable. Bien que les temps personnels constituent une grande partie de la distribution de la construction *boire du vin*, les modes impersonnels sont principalement utilisés. Nous notons ici la surreprésentation du mode impersonnel dans HIST par rapport aux autres sous-corpus et un très petit nombre d'implémentations de ces structures dans SENT.

L'observation de la réalisation de personnes verbales dans les sous-corpus a montré une préférence pour le singulier dans GEN, POL et FY, tandis que HIST et SENT se caractérisent par une préférence pour le pluriel. L'analyse a également montré que pour tous les sous-corpus, les hommes consomment du vin beaucoup plus souvent que les femmes.

Les résultats de nos observations sur les variations linguistiques de la construction *boire du vin* montrent des différences significatives entre les sous-corpus sur le plan syntaxique.



Dans la section suivante, nous allons analyser l'expression *prendre une bière*.

## 2. L'analyse de la collocation *prendre une bière*

Nous avons décidé de mener une analyse de la collocation *prendre une bière* qui est apparue comme spécifique au sous-corpus POL. Même si le collocatif *prendre* est aussi statistiquement significatif pour les pivots *café et thé*, il apparaît comme plus spécifique pour le pivot *bière*. Notre analyse se fonde sur 45 occurrences de la collocation dans POL. La figure 10, qui est une visualisation de l'arbre lexico-syntaxique, correspond à l'expression *prendre une bière* :

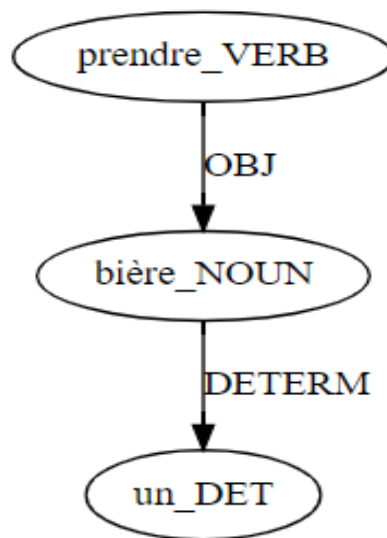


Figure 10 Arbre lexico-syntaxique récurrent automatiquement extrait (*prendre une bière*)

Dans cette section nous allons nous focaliser sur l'analyse linguistique de cette collocation, en observant les variations de la construction.

### 2.1. La variation sur le déterminant

En analysant la détermination dans la collocation *prendre une bière*, nous constatons que dans presque tous les cas, le nom *bière* et le verbe *prendre* privilégient légèrement l'article indéfini *une*- 36 occurrences (ex.49). Seulement 2 fois, nous remarquons l'utilisation de l'article défini *la* (ex.50). Nous notons aussi 1 occurrence de la variation du déterminant avec le possessif *sa*. Les autres exemples (3 occurrences) de la variation du déterminant sont les suivants : *des bières*, *deux bières* et *trois bières*. Voici les exemples issus du corpus :

49) Il sortit de la pièce, referma la porte, alla **prendre une bière** et la but lentement, assis au salon. (P. Senécal, *Les sept jours du talion*, 2002)

50) **Je pris la bière** qui restait, après quoi il fut vide, vides aussi les placards, rien à manger, si, quelques tranches de vieux pain, rien à manger rien à boire, pouvais-je recevoir une invitée dans cet appartement vide et lui offrir pour toute boisson l'eau tiède et infectée du Rhône que crachotaient mes robinets jaunes, pour toute nourriture l'air épaissi et parcheminé par la chaleur. (R.Belletto, *L'enfer*, 1986)

Notre analyse montre que le pivot *bière* dans la collocation *prendre une bière* choisit préférentiellement l'article indéfini *une*. Cette très forte préférence présente peu d'exceptions sous forme de variations de collocation.

## 2.2. Les temps verbaux du verbe « prendre »

En analysant les variations temporelles du verbe *prendre* dans la construction *prendre une bière* nous pouvons remarquer quelques préférences concernant la sélection des temps grammaticaux :

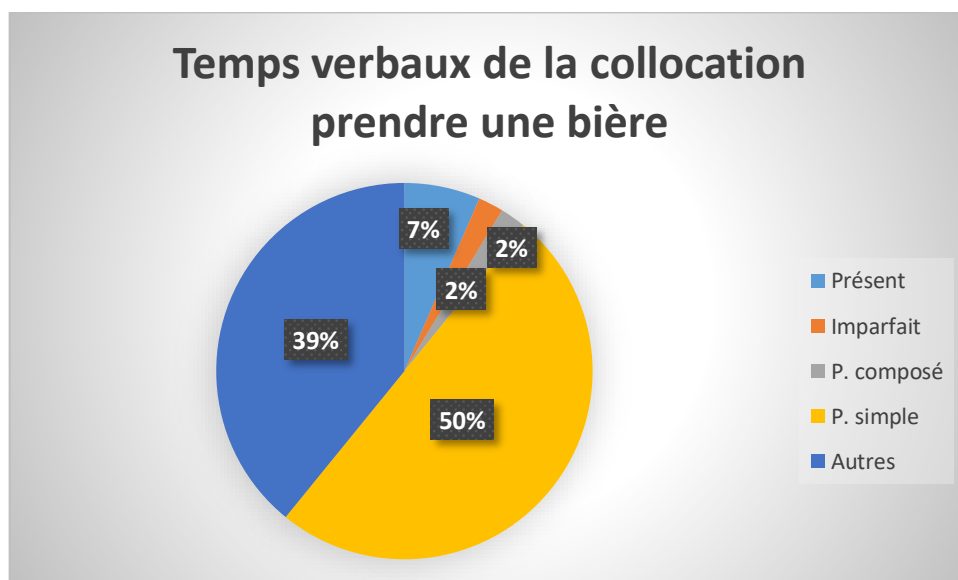


Figure 11 Répartition des temps verbaux pour la construction prendre une bière

Les données de la figure 11 indiquent un emploi majoritaire de la construction *prendre une bière* au passé simple. Seulement 7% d'occurrences sont au présent et aucune au futur. Le passé simple est un temps *perfectif* et limitatif. Il implique la narration d'un point apparu à un moment donné : L'utilisation du passé simple indique l'aspect global dans lequel on voit le procès dans sa globalité. L'action est envisagée avec les limites, en ayant un début et une fin. Dans certains cas (ex.51) nous observons un aspect de la succession d'actions brèves au passé simple.

51) Le médecin **prit une bière** au réfrigérateur et en but une longue gorgée, appuyé contre le comptoir. (P. Senécal, *Les sept jours du talion*, 2002)

52) Devant la gare d'Hazebrouck, ils **prirent une bière** en observant les lourds nuages gris qui s'accumulaient dans le ciel flamand. (J.-B. Pouy, *Larchmutz 5632*, 1999)

Nous remarquons que le passé simple est le plus courant pour la structure *prendre une bière*. L'utilisation de ce temps permet d'exprimer des actions accomplies et ponctuelles, un fait bref. Nous pouvons certainement envisager l'acte de *prendre une bière* comme une telle activité. Le fait que cette collocation soit si peu utilisée au présent ou imparfait prouve qu'il s'agit d'une activité qui a un aspect global, décrivant une activité de l'extérieur.

### 2.3. *Les modes verbaux de la collocation « prendre une bière »*

Dans le graphique présenté ci-dessus (figure 11), « Autres » (39%) regroupe les formes de l'infinitif (16 occurrences) et du conditionnel présent (2 occurrences). Les verbes introducteurs utilisés avec la construction *prendre une bière* prennent souvent la forme de verbes semi-auxiliaires. Nous remarquons, entre autres, les semi-auxiliaires temporels : *aller* (5 occurrences) qui exprime le futur proche et *venir de* (1 occurrence) qui fait référence au passé récent. Le verbe *être en train de* (1 occurrence) apparaît comme un semi-auxiliaire aspectuel duratif (ex.53). Les autres occurrences sont introduites par les verbes suivants: *se rendre de*, *vouloir*, *se lever*, *penser de*, *aller pour* (ex. 54).

53) La Japonaise **devait être en train de prendre sa bière** du soir dans le derrière (R. Belletto, *L'enfer*, 1986)

54) Il m'explique qu'il est revenu de New York et **qu'il voudrait bien prendre une bière** avec un « vrai Québécois qui est capable de boire sans vouloir absolument se battre » (P. Senécal, *Le Passager*, 1995)

Nos résultats montrent une forte préférence pour l'utilisation de la collocation *prendre une bière* dans la construction infinitive. Il est généralement utilisé avec des verbes semi-auxiliaires ou des verbes exprimant un désir ou une intention de *prendre une bière*.

### 2.4. *Les personnes verbales de la construction « prendre une bière »*

Les données issues du sous-corpus POL montrent que l'identité des buveurs est un aspect à mentionner. La répartition des personnes verbales- les buveurs de la bière est présentée dans la figure 12 :

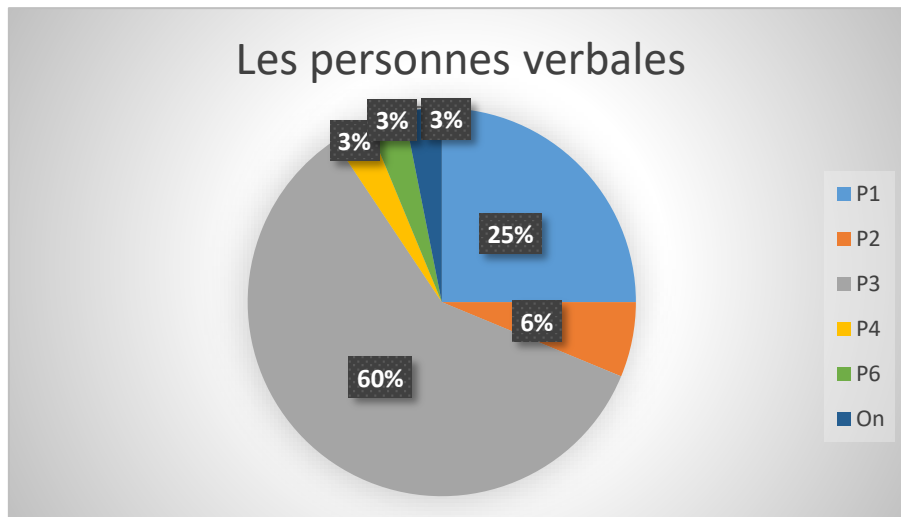


Figure 12 Répartition des personnes verbales pour la construction *prendre une bière*

En analysant les réalisations de l'ALR *prendre une bière*, nous notons que les personnes du singulier P1, P2 et P3 constituent la grande majorité d'occurrences (91%, ex.53 et 54). Cet état de fait est une indication claire d'un aspect de la consommation plutôt individuelle de la bière dans le sous-genre POL. Les personnes P4, P6 et *on* constituent seulement 9% d'occurrences ce qui indique que *prendre une bière* est un acte réalisé rarement à plusieurs.

53) **Il prit une bière** dans le frigidaire, la déboucha au décapsuleur vissé au mur et la vida d'un trait en réprimant une envie de pleurer. (B. Daeninckx, *Le géant inachevé*, 1984)

54) Finalement, **je prendrais bien une bière**, je lançai à Mme Orsini quand elle rouvrit sa porte. (J.-C. Izzo, *La trilogie marseillaise*. Solea, 1998)

En ce qui concerne l'analyse de la P3, seul 2 occurrences utilisent le pronom féminin *elle*. Dans les 17 cas restants, le buveur est identifié par le pronom *il*. Ce sont donc plutôt les hommes qui prennent une bière dans POL.

### 2.5. La variation avec l'adverbe et l'adjectif

Les données du corpus montrent que la structure *prendre une bière* est assez stable et nous trouvons peu de variations en son sein (20% d'occurrences). Elle peut subir une extension sur l'axe syntagmatique et être accompagnée d'adverbes, mais ces cas sont assez rares. Dans le sous-copis POL, nous distinguons seulement 4 occurrences d'une extension de la collocation *prendre une bière* par l'adverbe. On trouve 3 occurrences de la construction *prendre une autre bière* (ex.56) et 1 occurrence de la construction *prendre une bonne grosse bière* (ex.55) :

55) Qu'est-ce que tu penserais de **prendre une bonne grosse bière** bien fraîche, hé ?  
(J. Vautrin, *La Canicule*, 1982)

56) Il baissa le feu, couvrit le chaudron et **prit une autre bière** dans le réfrigérateur.  
(Ch. Brouillet, *Maud Graham. C'est Pour Mieux T'aimer Mon Enfant*, 1996)

## 2.6. La comparaison des collocations « boire du vin » et « prendre une bière »

Notre analyse de la structure *boire du vin* dans tous les sous-corpus et *prendre une bière* dans POL a révélé des différences significatives dans le comportement entre ces deux collocations. Bien que les deux constructions soient des collocations verbales contenant des noms de boissons, leurs variations linguistiques présentent de nombreuses différences.

La première différence significative entre ces constructions est la distribution des variantes des déterminants dans des collocations données. La construction *boire du vin* est caractérisée par un grand nombre de variantes des déterminants. Nous pouvons observer de nombreuses occurrences des variantes avec un déterminant défini, indéfini, partitif et peu d'occurrences avec par exemple un déterminant possessif ou démonstratif. La situation est complètement différente avec la collocation *prendre une bière*. Cette structure est plutôt stable et montre peu de variantes, autant que 87% de toutes les occurrences sont des variantes avec un déterminant indéfini *une*.

Il faut également noter un aspect intéressant qui est l'utilisation de ces deux collocations dans les temps verbaux et dans les modes. Quant au mode personnel, la construction *boire du vin*, selon le sous-corpus, est généralement utilisée au présent ou à l'imparfait, moins au passé simple et passe composé. Néanmoins, cette collocation est utilisée dans le mode impersonnel, en particulier dans l'infinitif. Dans le cas de la collocation *prendre une bière*, la répartition du temps et des variantes modales est assez différente. On observe ici un avantage significatif de l'utilisation du passé simple (jusqu'à 50 %). Il faut également noter que malgré la part importante du mode impersonnel, cette construction n'est utilisée qu'à l'infinitif, aucune occurrence de cette collocation en gérondif n'apparaît du tout.

En analysant la distribution des personnes verbales de ces deux collocations, on peut remarquer que la construction *prendre une bière* préfère dans une large mesure le singulier (91%). Quand il s'agit de la collocation *boire du vin*, selon le sous-corpus, elle apparaît plus souvent au singulier ou au pluriel, mais aucun sous-corpus n'a une telle différence entre ces personnes que la construction *prendre une bière* en POL.

Nous arrivons à la conclusion que le fonctionnement linguistique des collocations particulières autour des noms de boissons peut différer considérablement, ce qui peut contribuer au processus de distinction entre les genres romanesques.

Le chapitre 8 de notre travail était consacré à l'analyse linguistique de la collocation *boire du vin* dans GEN en comparant avec les autres sous-corpus et *prendre une bière* dans POL. Nous concluons que notre analyse a révélé des différences significatives dans le fonctionnement de la collocation dans les sous-corpus.

Dans le cas de la collocation *boire du vin*, nous avons constaté de nombreux écarts par rapport à la variante de base en termes de déterminants. L'analyse a permis de distinguer deux cas de sous-corpus dans lesquels la collocation se comporte différemment des autres sous-corpus. La distribution des trois variantes de base de la collocation (avec les déterminants partitifs *de* et *du*, le déterminant défini *le*) dans SENT est assez équilibrée, tandis que dans le sous-corpus restant on voit un avantage significatif des variations avec le déterminant partitif *du*. HIST se distingue par le fait que sa deuxième variante la plus courante d'un déterminant dans la construction *boire du vin* est un déterminant défini *le*, et non un déterminant du partitif *de*, comme dans d'autres cas.

En ce qui concerne les temps verbaux, POL est le seul sous-corpus caractérisé par une surreprésentation du passé composé. Comparé à d'autres sous-corpus, FY montre très peu de préférence pour le présent. Il convient également de noter la part importante du mode impersonnel dans HIST et, en comparaison, la très petite part de ce mode dans SENT.

Les observations de personnes verbales ont montré que GEN, FY et POL sont caractérisés par une préférence pour singulier, tandis que dans le cas de HIST et SENT nous rencontrons une surreprésentation du pluriel.

Dans la seconde partie de notre analyse, nous nous sommes concentrées sur la collocation *prendre une bière* dans POL. D'après nos observations, nous pouvons constater qu'il s'agit d'une construction plus stable que la collocation *boire du vin*. Elle montre peu de variations en ce qui concerne le déterminant, la majorité des occurrences apparaît dans la variante avec le déterminant indéfini *une*.

L'étude de l'utilisation des temps verbaux a montré que la collocation *prendre une bière* dans POL montre une préférence pour le passé simple. Ce fait linguistique est contraire au comportement de la collocation *boire du vin* dans les sous-corpus pour lequel les constructions impersonnelles représentaient le plus grand pourcentage en comparant avec les temps verbaux individuels.

La collocation *prendre une bière* est utilisée dans POL dans presque tous les cas au singulier. Dans le cas de la construction *boire du vin*, la préférence dépend du sous-corpus, mais les différences entre le singulier et le pluriel ne sont pas significatives du tout.



## Conclusion

Nous avons analysé, à travers ce travail de recherche deux expressions contenant les noms de boissons : *boire du vin* dans tous les sous-corpus et *prendre une bière* dans POL. Le but de notre étude était d'étudier si les phraséologismes liés aux noms de boissons peuvent être un moyen de distinguer les genres romanesques.

Tout d'abord, grâce à l'outil le *Lexicoscope* nous avons eu la possibilité de créer les tableaux avec des données détaillées sur les pivots (les noms de boissons) avec leurs collocatifs dans six sous-corpus du roman contemporain. Sur la base des tableaux que nous avons créés, nous avons effectué une analyse quantitative de données, qui nous a permis de constater les premières différences entre les sous-corpus. Ensuite, nous avons choisi deux collocations à tester dans l'approche qualitative. Nous avons observé leur fonctionnement linguistique et les variantes possibles dans les sous-corpus. Nous avons principalement analysé quatre aspects: les variantes des déterminants, les temps et les modes verbaux ainsi que les personnes verbales.

Notre recherche nous a permis de mettre en évidence des différences significatives dans le comportement et la distribution des variantes des collocations autour de noms de boissons étudiées. Les quatre points de comparaison distingués au cours de notre recherche nous ont permis de constater que les collocations avec les noms de boissons fonctionnent de manière plus ou moins stable en apparaissant dans les différentes variations avec le degré relatif d'occurrence selon chaque de sous-corpus.

L'étude que nous avons effectuée nous a permis de constater que même si la collocation *boire du vin* apparaît dans presque tous les sous-corpus et il semblerait qu'il soit impossible de l'utiliser pour distinguer les genres littéraires, on peut observer une petite quantité de différences significatives dans le fonctionnement de cette collocation dans les sous-corpus individuels. Comme un exemple on peut mentionner la structure caractéristique de la distribution des déterminants dans SENT ou la préférence pour le passé composé dans POL.

La collocation *prendre une bière* est moins courante et est statistiquement significative seulement pour POL. Néanmoins, l'analyse de cette collocation et la comparaison avec celle du *boire du vin* nous a permis de constater que les phraséologismes autour de la boisson peuvent fonctionner différemment et avoir une distribution de variantes différente dans les quatre catégories étudiées par nous.

Le travail sur ce sujet nous a permis d'explorer les questions comme les rapports entre le genre romanesque et la syntaxe ainsi que la phraséologie. Nous trouvons pourtant que cette recherche et introduction de la phraséologie pourraient constituer un complément d'une approche stylistique et un approfondissement de la recherche sur la caractérisation et distinction des genres littéraires.

Ainsi, les résultats de nos analyses nous permettent de confirmer que ces expressions lexico-syntaxiques peuvent être considérées comme le moyen pour distinguer les genres romanesques.

Nous arrivons à la conclusion que notre étude pourrait être prolongée aussi bien au niveau stylistique et l'analyse d'usage sociale qui permettraient une compréhension plus large du contexte et du fonctionnement de collocations autour de la consommation d'alcool dans divers genres romanesques. Nous sommes conscients que notre analyse qualitative ne s'est basée que sur deux collocations, ce qui constitue un champ de la recherche très limité. Nous pensons qu'il serait intéressant de procéder à l'analyse sur d'autres collocations pour avoir une image plus complète du fonctionnement de ce type de collocation dans les genres romanesques.

## Bibliographie

- Adam, J.-M. (2001). Discours et interdisciplinarité : Benveniste lecteur de Saussure. *Cahiers Ferdinand de Saussure*, (54), 201-218.
- Baker, M. (1995). Corpora in translation studies : An overview and some suggestions for future research. *Target*, v. 7(2), 223-243.
- Bakhtine, M. (1984). *Esthétique de la création verbale*. Paris: Gallimard.
- Bakhtine, M. (1986). *The problem of speech genres. In Speech genres and other late essays*. Austin: University of Texas Press.
- Bakhtine, M., Todorov, T., & Aucouturier, A. (1984). *Esthétique de la création verbale*. Paris: Gallimard.
- Bally, C. (1909). *Traité de stylistique française*. Heidelberg: C. Winter.
- Barthes, R. (1972). *Mythologies*. New York: Farrar, Straus and Giroux.
- Béjoint, H. (2007). Informatique et lexicographie de corpus : Les nouveaux dictionnaires. *Revue française de linguistique appliquée*, Vol. XII(1), 7-23.
- Bernard, C. (1996). *Le passé recomposé. Le roman historique français du dix-neuvième siècle*. Paris: Hachette.
- Bilger, M. (2000). *Corpus : Méthodologie et applications linguistiques*. Paris: H. Champion.
- Blumenthal, P. (2007). Sciences de l'Homme vs sciences exactes : Combinatoire des mots dans la vulgarisation scientifique. *Revue française de linguistique appliquée*, Vol. XII(2), 15-28.
- Bolly, C. (2010). Flou phraséologique, quasi-grammaticalisation et pseudo marqueurs de discours : Un no man's land entre syntaxe et discours ? *Linx. Revue des linguistes de l'université Paris X Nanterre*, (62-63), 11-38. <https://doi.org/10.4000/linx.1356>
- Bommier-Pincemin, B. (1999a). Construire et utiliser un corpus : Le point de vue d'une sémantique textuelle interprétative. In A. Condamines (Éd.), *Corpus et Traitement Automatique des Langues : Pour une réflexion méthodologique (TALN'99)* (p. 81-100).

- Bommier-Pincemin, B. (1999b). *Diffusion ciblée automatique d'informations : Conception et mise en oeuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents. (Mémoire de doctorat)*. L'UNIVERSITÉ PARIS IV, Paris.
- Bouquet, S. (2004). Linguistique générale et linguistique des genres (introduction au numéro). *Langages*, (153), 3-14.
- Boutevin, C., & Principalli, P. (2008). *Dictionnaire de la littérature de jeunesse : À l'usage des professeurs des écoles / Christine Boutevin, Patricia Richard-Principalli*. Paris: Vuibert.
- Bozzetto, R. (2007). *La science-fiction*. Paris: Armand Colin.
- Branca-Rosoff, S. (1999). Types, modes et genres : Entre langue et discours. *Langage & société*, 87(1), 5-24. <https://doi.org/10.3406/lsoc.1999.2851>
- Bronckart, J.-P. (1996). Genres de textes, types de discours et opérations psycholinguistiques. *Voies livres*, (78), 1-20.
- Bronckart, J.-P. (2004). Les genres de textes et leur contribution au développement psychologique. *Langages*, n° 153(1), 98-108.
- Bronckart, J.-P. (2008). Genres de textes, types de discours et « degrés » de langue. *Texte !*, XIII(1), 96.
- Castex, P. G. (1994). *Le Conte Fantastique en France : De Nodier à Maupassant* (7th éd.). Paris: José Corti.
- Charaudeau, P., & Maingueneau, D. (2002). *Dictionnaire d'Analyse du Discours*. Seuil.
- Condamines, A. (2005). Linguistique de corpus et terminologie. *Langages*, 157(1), 36-47. Cairn.info. <https://doi.org/10.3917/lang.157.0036>
- Constans, E. (1999). *Parlez-moi d'amour : Le roman sentimental, des romans grecs aux collections de l'an 2000*. Limoges: PULIM.
- Cornot, D. (2012). Vin et femme. In J.-P. Poulain (Éd.), *Dictionnaire des cultures alimentaires: Vol. I*. Paris: Presses Universitaires de France.
- Cruse, D. A. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.

- Delais-Roussarie, E., & Durand, J. (2003). *Corpus et Variation en Phonologie du Français : Méthodes et Analyses*. Toulouse: Presses Universitaires du Mirail.
- Dubois, J. (2005). *Le roman policier ou la modernité*. Paris: Armand Colin.
- Duby, D. (1996). L'art, l'écriture et l'histoire. *Le Débat*, 92(5), 174-192.
- Eshkol-Taravella, I., & Lefeuve-Halftermeyer, A. (2017). Linguistique de corpus : Vues sur la constitution, l'analyse et l'outillage. *Corela. Cognition, représentation, langage*, (HS-21).  
<https://doi.org/10.4000/corela.4800>
- Firth, J. R. (1957). *Papers in Linguistics 1934-1951*. London: Oxford University Press.
- Francis, G. (1991). Nominal group heads and clause structure. *WORD*, 42(2), 145-156.  
<https://doi.org/10.1080/00437956.1991.11435836>
- Freyermuth, S. (2011). Un genre peut en cacher un autre : Une histoire de détournement. *Linx. Revue des linguistes de l'université Paris X Nanterre*, (64-65), 173-187.  
<https://doi.org/10.4000/linx.1411>
- Gengembre, G. (2006). *Le Roman historique. "50 questions"*. Paris: Klincksieck.
- Gledhill, C., & Frath, P. (2007). Collocation, phrasème, dénomination : Vers une théorie de la créativité phraséologique. *La linguistique*, Vol. 43(1), 63-88.
- Gonon, Laetitia, & Sorba, J. (2019). Phraséologismes autour du vin et du champagne dans le roman contemporain. *Terminologies et discours gastronomiques et œnologiques. Le vin et les autres produits*. Présenté à Innsbrück, Austria.
- González Rey, I. (2002). *La phraséologie du français*. Toulouse: Presses Universitaires du Mirail.
- González Rey, I. (2008). *La didactique du français idiomatique*. Fernelmont: E.M.E.
- González Rey, I., & Fernández González, V. (2006). La didactique des expressions figées en langue étrangère. In *Análisis del discurso : Lengua, cultura, valores; actas del I Congreso Internacional (Universidad de Navarra, Pamplona, noviembre de 2002)* (p. 1015-1029). Madrid: Arco Libros.

- Grossmann, F., Gymnich, M., & Siepmann, D. (2019). Alcohol and Tobacco Consumption in English and French Novels since the 1950s : A Corpus-stylistic Analysis. In I. Novakova & D. Siepmann (Éds.), *Phraseology and style in sub-genres of the novel : A synthesis of corpus and literary perspectives*. PALGRAVE, MACMILLAN. Consulté à l'adresse <https://hal.archives-ouvertes.fr/hal-02013409>
- Habert, B., Nazarenko, A., & Salem, Gérard. (1997). *Les linguistiques de corpus*. Paris: Armand Colin.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Hausmann, F. J., & Blumenthal, P. (2006). Présentation : Collocations, corpus, dictionnaires. *Langue française, n° 150(2)*, 3-13.
- Hausman. (1989). Le dictionnaire des collocations. In *Wörterbücher/Dictionaries. Ein internationales Handbuch zur Lexikographie/An International Encyclopedia of Lexicography* (p. 1010-1019). Berlin/New-York : de Gruyter.
- Heid, U. (1994). On Ways Words Work Together. Topics in Lexical Combinatorics. *EURALEX'94 Proceedings*, p. 226-257.
- Hoey, M. (2005). *Lexical Priming : A New Theory of Words and Language*. London ; New York: Routledge.
- Huemer, W., Gibson, J., & Pocci, L. (Éds.). (2007). Why Read Literature? The cognitive function of form. In *A Sense of the World : Essays on Fiction, Narrative, and Knowledge*. New York: Routledge. Consulté à l'adresse [https://www.academia.edu/3700797/Why\\_Read\\_Literature\\_The\\_cognitive\\_function\\_of\\_form](https://www.academia.edu/3700797/Why_Read_Literature_The_cognitive_function_of_form)
- Jacques, M.-P. (2005). Pourquoi une linguistique de corpus. In G. Williams (Éd.), *Linguistique de corpus*. Rennes: Presses Universitaires de Rennes.
- Kraif, O., & Diwersy, S. (2012). Le Lexicoscope : Un outil pour l'étude de profils combinatoires et l'extraction de constructions lexico-syntaxiques (The Lexicoscope : an integrated tool for

- combinatoric profiles observation and lexico-syntactic constructs extraction) [in French].  
*Proceedings of the Joint Conference JEP-TALN-RECITAL 2012, volume 2: TALN*, 399–406.  
 Grenoble, France: ATALA/AFCP. Consulté à l'adresse  
<https://www.aclweb.org/anthology/F12-2033>
- Kraif, O., Novakova, I., & Sorba, J. (2016). Constructions lexico-syntaxiques spécifiques dans le roman policier et la science-fiction. *LIDIL - Revue de linguistique et de didactique des langues*, 53, 143-159. <https://doi.org/10.4000/lidil.3976>
- Larsson, F. (1991). Approche du roman sentimental durassien. In *Le roman sentimental: Vol. II* (p. 77-94). Limoges: PULIM.
- Leech, G. (1992). *Directions in Corpus Linguistics : Proceedings of Nobel Symposium 82; Stockholm, 4 - 8 August 1991*. Berlin: Mouton de Gruyter.
- Leech, G. (2007). New resources, or just better old ones? The Holy Grail of representativeness. In M. Hundt, N. Nesselhauf, & C. Biewer (Éds.), *Corpus Linguistics and the Web* (p. 133-149). Brill | Rodopi. [https://doi.org/10.1163/9789401203791\\_009](https://doi.org/10.1163/9789401203791_009)
- Legallois, D. (2012). La colligation : Autre nom de la collocation grammaticale ou autre logique de la relation mutuelle entre syntaxe et sémantique ? *Corpus*, (11). Consulté à l'adresse <http://journals.openedition.org/corpus/2202>
- Legallois, D., Charnois, T., & Poibeau, T. (2016). Repérer les clichés dans les romans sentimentaux grâce à la méthode des « motifs ». *Lidil. Revue de linguistique et de didactique des langues*, (53), 95-117. <https://doi.org/10.4000/lidil.3950>
- Legallois, D., & Tutin, A. (2013). Présentation : Vers une extension du domaine de la phraséologie. *Langages*, n° 189(1), 3-25.
- Lengert, J. (2001). Phraseologie / Phraséologie. In *Geschichte des Faches Romanistik. Methodologie (Das Sprachsystem): Vol. I/1* (Reprint 2011, p. 802-853). Berlin, Boston: De Gruyter. <https://doi.org/10.1515/9783110938388-034>

- Lits, M. (1999). *Le roman policier : Introduction à la théorie et à l'histoire d'un genre littéraire*. Liège: Cefal.
- Lukács, G. (2000). *Le roman historique*. Paris: Payot, coll. « Petite bibliothèque Payot ».
- Mayaffre, D. (2005). Rôle et place des corpus en linguistique : Réflexions introductives. *Texto!*, X(4). Consulté à l'adresse [http://www.revue-texto.net/Reperes/Themes/Mayaffre\\_Corpus.html](http://www.revue-texto.net/Reperes/Themes/Mayaffre_Corpus.html)
- Mel'čuk, I. (1984). *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques*. Montréal: Presses de l'Université de Montréal.
- Mel'čuk, I. (1988). *Dependency Syntax : Theory and Practice*. State University of New York Press.
- Montémont, V. (2009). Queneau, Perec, Duras : Trois manières de boire dans le roman français. *CONTEXTES. Revue de sociologie de la littérature*, (6). <https://doi.org/10.4000/contextes.4525>
- Nourrisson, D. (1988). Aux origines de l'antialcoolisme. *Histoire, économie & société*, 7(4), 491-506. <https://doi.org/10.3406/hes.1988.2392>
- Nourrisson, D. (2013a). 8—Le buveur de la seconde moitié du XXe siècle. In *Crus et cuites. Histoire du buveur* (p. 263-292). Paris: Éditions Perrin. Consulté à l'adresse <https://www.cairn.info/crus-et-cuites-histoire-du-buveur--9782262042837-page-263.htm>
- Nourrisson, D. (2013b). Présentation. In *Crus et cuites. Histoire du buveur* (p. 109-114). Paris: Éditions Perrin. Consulté à l'adresse <https://www.cairn.info/crus-et-cuites-histoire-du-buveur--9782262042837-page-109.htm>
- Novakova, I., & Tutin, A. (Éds.). (2009). *Le lexique des émotions*. Grenoble: ELLUG.
- Palmer, H. E. (1933). *Second interim report on English collocations*. Tokyo: Kaitakusha.
- Pierrel, J.-M. (2005). Linguistique de corpus et Traitement Automatique de la Langue. *Communication et connaissances : supports et médiations à l'âge de l'information / Jean-Gabriel Ganascia coordinateur*. Présenté à Communication et connaissances : supports et



- médiations à l'âge de l'information / Jean-Gabriel Ganascia coordinateur. Consulté à l'adresse <https://halshs.archives-ouvertes.fr/halshs-00005041>
- Poslaniec, C., & Houyel, C. (2001). *Activités de lecture à partir de la littérature policière*. Hachette Éducation.
- Rastier, F. (2001). *Arts et sciences du texte*. Presses Universitaires de France.  
<https://doi.org/10.3917/puf.rast.2001.01>
- Rastier, F. (2005). Enjeux épistémologiques de la linguistique de corpus ». In *La linguistique de corpus* (p. 31-45). Rennes: Presses Universitaires de Rennes.
- Roa, M., & Santiago, R. (2011). *Le policier mobile et les mobiles du policier : Un « mauvais genre » sous la loupe didactique ( Mémoire du master)* (Université de Genève). Université de Genève, Genève. Consulté à l'adresse <https://archive-ouverte.unige.ch/unige:17584>
- Sartre, J.-P. (1947). *'Aminadab' ou du fantastique considéré comme un langage, Situations I*. Paris: Gallimard.
- Schaeffer, J.-M. (1989). *Qu'est-ce qu'un genre littéraire ?* Paris: Le Seuil.
- Schiff, B. (2012). The Function of Narrative : Toward a Narrative Psychology of Meaning. *Narrative Works*, 2(1).
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- Sinclair, J. (1996). Preliminary recommendations on corpus typology. *EAGLES, Document EAG-TCWG-CTYP/P*.
- Sinclair, J. (2003). *Reading Concordances. An Introduction*. London: Longman.
- Todorov, T. (1970). *Introduction à la littérature fantastique*. Paris: Seuil.
- Todorov, T. (1971). *Poétique de la prose*. Paris: Seuil.
- Tutin, A. (2013). Les collocations lexicales : Une relation essentiellement binaire définie par la relation prédicat-argument. *Langages*, n° 189(1), 47-63.
- Tutin, A., & Grossmann, F. (2002). Collocations régulières et irrégulières : Esquisse de typologie du phénomène collocatif. *Revue française de linguistique appliquée, Vol. VII(1)*, 7-25.

Williams, G., Collectif, Jacques, M.-P., Rastier, F., & Blanche-Benveniste, C. (2005). *La linguistique de corpus*. Rennes: Presses Universitaires de Rennes.

Zaharieva, R., & Kaldieva-Zaharieva, S. (2017). Des principaux termes dans le domaine de la phraséologie. In F. Grossmann, *La phraséologie : Sémantique, syntaxe, discours* (p. 15-37). Paris: Champion.

## **Sigles et abréviations utilisés**

ALR : arbres lexico-syntaxique récurrent

CLS : construction lexico-syntaxique

LLR : log-Likelihood Ratio

FY : le sous-corpus de roman fantasy

GEN le sous-corpus de roman général

HIST : le sous-corpus de roman historique

SENT : le sous-corpus de roman sentimental

SF : le sous-corpus de roman de science-fiction

POL : le sous-corpus de roman policier

## Table des illustrations

Tableau 1 La composition du corpus romanesque PhraseoRom partitionné en sous-genres .....	37
Tableau 2 Les collocatifs du pivot <i>café</i> dans POL avec leur LLR .....	54
Tableau 4 Les collocatifs du pivot <i>vin</i> dans FY .....	103
Tableau 5 Les collocatifs du pivot <i>vin</i> dans GEN.....	105
Tableau 6 Les collocatifs du pivot <i>champagne</i> dans POL .....	106
Tableau 7 Les collocatifs du pivot <i>champagne</i> dans SENT .....	106
Tableau 8 Les collocatifs du pivot <i>champagne</i> dans GEN.....	106
Tableau 9 Les collocatifs du pivot <i>bière</i> dans POL .....	107
Tableau 10 Les collocatifs du pivot <i>bière</i> dans GEN .....	107
Tableau 11 Les collocatifs du pivot <i>vodka</i> dans POL.....	108
Tableau 12 Les collocatifs du pivot <i>vodka</i> dans GEN .....	108
Tableau 13 Les collocatifs du pivot <i>café</i> dans POL.....	110
Tableau 14 Les collocatifs du pivot <i>café</i> dans SENT .....	110
Tableau 15 Les collocatifs du pivot <i>café</i> dans GEN.....	111
Tableau 16 Les collocatifs du pivot <i>thé</i> dans POL .....	112
Tableau 17 Les collocatifs du pivot <i>thé</i> dans SENT .....	112
Tableau 18 Les collocatifs du pivot <i>thé</i> dans GEN.....	113
Figure 1 La distribution des collocatifs du pivot <i>vin</i> dans les sous-corpus romanesques .....	49
Figure 2 La distribution des collocatifs dans SENT .....	57
Figure 3 La distribution des collocatifs dans POL.....	58
Figure 4 La distribution des collocatifs dans GEN .....	58
Figure 5 Répartition des collocatifs par catégorie grammaticale dans chaque de sous-corpus .....	59
Figure 6 Le calcul de spécificité (LLR) de la collocation <i>boire+vin</i> dans les sous-corpus .....	62
Figure 7 L'arbre lexico-syntaxique de <i>boire du vin</i> .....	64
Figure 8 Répartition des temps verbaux pour la construction <i>boire du vin</i> dans GEN .....	69
Figure 9 Répartition des personnes verbales pour la construction <i>boire du vin</i> dans GEN .....	74
Figure 10 Arbre lexico-syntaxique récurrent automatiquement extrait ( <i>prendre une bière</i> ) .....	78
Figure 11 Répartition des temps verbaux pour la construction <i>prendre une bière</i> .....	79
Figure 12 Répartition des personnes verbales pour la construction <i>prendre une bière</i> .....	81

## Table des annexes

Annexe 1 La liste des auteurs de roman policier .....	98
Annexe 2 La liste des auteurs de roman fantastique .....	99
Annexe 3 La liste des auteurs de roman science-fiction .....	100
Annexe 4 La liste des auteurs de roman historique .....	101
Annexe 5 La liste des auteurs de roman sentimental .....	102
Annexe 7 Les collocatifs du pivot <i>vin</i> dans les sous-corpus .....	103
Annexe 8 Les collocatifs du pivot <i>champagne</i> dans les sous-corpus .....	106
Annexe 9 Les collocatifs du pivot <i>bière</i> dans les sous-corpus .....	107
Annexe 10 Les collocatifs du pivot <i>vodka</i> dans les sous-corpus .....	108
Annexe 11 Les collocatifs du pivot <i>café</i> dans les sous-corpus.....	109
Annexe 12 Les collocatifs du pivot <i>thé</i> dans les sous-corpus .....	112
Annexe 13 Les diagrammes concernant la collocation <i>boire du vin</i> dans FY.....	114
Annexe 14 Les diagrammes concernant la collocation <i>boire du vin</i> dans HIST .....	115
Annexe 15 Les diagrammes concernant la collocation <i>boire du vin dans</i> SENT.....	116
Annexe 16 Les diagrammes concernant la collocation <i>boire du vin dans</i> POL.....	117

## **Annexe 1**

### **La liste des auteurs de roman policier**

André Arnaud, Brigitte Aubert, Alain Audin, José Azincourt, Cédric Bannel, Baptiste-Marrey (Jean-Claude Marrey), Fred Belin, René Belletto, Patrick Besson, Philippe Besson, Boileau-Narcejac (Thomas Narcejac), Pierre Borromée, Chrystine Brouillet, Pascal Bruckner, Serge Brussolo, Jérôme Camut (en collaboration avec Nathalie Hug), Antoine Chainas, Patrick Chamoiseau, Maxime Chattam, Philippe Claudel, Sandrine Collette, Michel Crespy, Didier Daeninckx, Maurice G. Dantec, Frédéric Dard, Jérôme Delafosse, André Delauré, Gérard Delteil, Gérard de Villiers, Joël Dicker, DOA (Hervé Albertazzi), Sébastien Fritsch, Max Gallo, Olivier Gay, Karine Giébel, Franz-Olivier Giesbert, Jean-Christophe Grange, Jules Grasset, Christophe Guillaumot, Marek Halter, Jean-Claude Izzo, Andrea H. Japp, Sébastien Japrisot, Jérôme Jarrige, Serge Joncour, Thierry Jonquet, Yasmina Khadra, André Klopmann, Marie Laberge, Denis Langlois, Pierre Lemaitre, Bernard Lenteric, Daniel Macouin, Pierre Magnan, Léo Malet, Jean-Patrick Manchette, Charlène Mauwls, Martin Michaud, Frédérique Molay, Camille Morillon, Guillaume Musso, Jérôme Noirez, Christian Olivaux, Jean-Hugues Opper, Jean-François Parot, Daniel Pennac, Georges Perec, Jean-Bernard Pouy, Claude Ragon, Jean-Christophe Rufin, Patrick Senécal, Jean-Claude Starosta, Dominique Sylvain, Danielle Thiéry, Franck Thilliez, Didier van Cauwelaert, Fred Vargas, Jean Vautrin, Boris Vian, Marc Villard, Eric Vincent.

## **Annexe 2**

### **La liste des auteurs de roman fantastique**

Raphaël Albert, Ange ( Anne Guéro et Gérard Guéro), Francis Berthelot, Éric Boisset, Pierre Bottero, Charlotte Bousquet, Édouard Brasey, Flavia Bujor, Marie Caillet, Maxime Chattam, Fabien Clavel, Alain Damasio, Alcide Dario, Thomas Day ( et avec Ugo Bellagamba), Chris Debien, Denis Duclos, Catherine Dufour, Jean-Louis Fetjaine, Mathieu Gaborit, Laurent Gidon, Régis Goddyn, Pierre Grimbert , Johan Heliot, Jean-Philippe Jaworski, Hervé Jubert, Gabriel Katz, Laurent Kloetzer, Henri Lœvenbruck, Colin Marchika, Justine Niogret, Jérôme Noirez, Olivier Peru, Pierre Pevel, Michel Robert, Anne Robillard, Magali Ségura, Bénédicte Taffin, John Thomas, Élodie Tirel, Adrien Tomas, Christian Vilà.

### **Annexe 3**

#### **La liste des auteurs de roman science-fiction**

Yal Ayerdhal, Jacques Barbéri, Pierre Barbet, René Barjavel, Pierre Bordage, Pierre, Boule, B. R. Bruss, Serge Brussolo, Richard Canal, Francis Carsac, Philippe Curval, Maurice G. Dantec, Michel Demuth, Thierry Di Rollo, Drode Daniel, Catherine Dufour, Jean-Pierre Fontana, Laurent Genefort, Jimmy Guieu, Johan Heliot, Nathalie Henneberg, Paul-Jean Héroult, Michel Jeury, Gérard Klein, Nicolas Lefebvre, Serge Lehman, Jean-Marc Ligny ( et avec D. Goult), Robert Merle, Pierre Pelot, Christiane Rochefort, Jean-Christophe Rufin, Didier Van Cauwelaert, Antoine Volodine, Élisabeth Vonarburg, Roland Wagner, Daniel Walther, Bernard Werber, Joëlle Wintrebert



## **Annexe 4**

### **La liste des auteurs de roman historique**

Laurent Binet, Lucien Bodard, Serge Brussolo, Albert Camus,  
Françoise Chandernagor, Bernard Clavel, Pierre Combescot, Jean d'Aillon,  
Christophe Donner, Maurice Druon, Clara Dupont-Monod, Frédéric H. Fajardie, Éric Faye,  
Max Gallo, Laurent Gaudé, Patrick Girard, Patrick Grainville, Johan Heliot, Christian Jacq,  
Pascal Lainé, Pierre Lemaitre, Andreï Makine, Robert Merle, Pierre Michon,  
Céline Minard, Louis Noir, Erik Orsenna, Patrick Rambaud, Jean-Christophe Rufin,  
Françoise Sagan, Eric-Emmanuel Schmitt, Christian Signol, François Taillandier, Chantal  
Thomas, Michel Tournier, Henri Troyat, Jean Vautrin, Marguerite Yourcenar

## **Annexe 5**

### **La liste des auteurs de roman sentimental**

Agnès Abécassis, Isabelle Alexis, Daniel Arsand, Emmanuèle Bernheim, Janine Boissard, Françoise Bourdin ; Adèle Bréau, Patrick Cauvin, Noëlle Châtelet, Maxime Chattam, Nicole de Buron, Grégoire Delacourt, Audrey Falk, Claudie Gallay, Anna Gavalda, Serge et Anne Golon, Sophie Henrionnet, Magalie Inguibert, Alexandre Jardin, Sophie Jomain, Gaëlle Josse, Brigitte Kernel, Marie-Sophie Kesteman, Marie-Sisi Labrèche, Sophie Lambert, Agnès Ledig, Gilles Legardinier, Frédéric Lenoir et Simonetta Greggio, Marc Levy, Marianne Levy, Guillaume Musso, Véronique Olmi, Jonathan Pacula, Catherine Pancol, Hélène Simart, Alain Teulié, Laura Trompette, Didier Van Cauwelaert, Marie Vareille

**Annexe 7**  
**Les collocatifs du pivot *vin* dans les sous-corpus**

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>verre</i>	nom	17	644,6470
<i>bouteille</i>	nom	18	446,0244
<i>coupe</i>	nom	15	427,7249
<i>boire</i>	verbe	19	370,8604

Tableau 3 Les collocatifs du pivot *vin* dans FY

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>boire</i>	verbe	20	649,5558
<i>verre</i>	nom	19	429,3436
<i>bouteille</i>	nom	15	428,9810
<i>blanc</i>	adjectif	15	405,4834

Tableau 3 Les collocatifs du pivot *vin* dans HIST

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>verre</i>	nom	37	811,1349
<i>blanc</i>	adjectif	28	796,5733
<i>bouteille</i>	nom	42	782,1366
<i>rouge</i>	adjectif	24	294,2725
<i>boire</i>	verbe	22	237,0997
<i>gorgée</i>	nom	18	223,7001

<i>bon</i>	adjectif	18	86,0948
------------	----------	----	---------

Tableau 4 Les collocatifs du pivot *vin* dans POL

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>verre</i>	nom	28	748,0847
<i>bouteille</i>	nom	23	505,0578
<i>blanc</i>	adjectif	16	435,6294
<i>boire</i>	verbe	16	144,9442

Tableau 5 Les collocatifs du pivot *vin* dans SENT

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>bouteille</i>	nom	15	343,7050
<i>verre</i>	nom	19	309,0661

Tableau 6 Les collocatifs du pivot *vin* dans SF

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>verre</i>	nom	83	2018,5226
<i>bouteille</i>	nom	81	1760,7624
<i>boire</i>	verbe	91	1601,3497
<i>blanc</i>	adjectif	78	1382,6211
<i>rouge</i>	adjectif	64	1052,1526
<i>verser</i>	verbe	24	175,1847
<i>chaud</i>	adjectif	18	172,6900
<i>gorgée</i>	nom	17	120,0713
<i>bon</i>	adjectif	37	117,0713

<i>goûter</i>	verbe	18	100,2743
<i>mauvais</i>	adjectif	16	80,9115
<i>marchand</i>	nom	15	77,2318
<i>frais</i>	adjectif	16	77,0092
<i>pain</i>	nom	18	75,5650
<i>servir</i>	verbe	27	67,6217
<i>apporter</i>	verbe	23	47,7495

Tableau 4 Les collocatifs du pivot *vin* dans GEN

**Annexe 8**  
**Les collocatifs du pivot *champagne* dans les sous-corpus**

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>coupe</i>	nom	24	500,0641
<i>bouteille</i>	nom	24	311,2611
<i>boire</i>	verbe	21	133,2788

Tableau 5 Les collocatifs du pivot *champagne* dans POL

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>coupe</i>	nom	22	1030,4259
<i>bouteille</i>	nom	23	692,8237

Tableau 6 Les collocatifs du pivot *champagne* dans SENT

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>coupe</i>	nom	47	876,1371
<i>bouteille</i>	nom	47	791,2424
<i>boire</i>	verbe	41	568,1158

Tableau 7 Les collocatifs du pivot *champagne* dans GEN

**Annexe 9**  
**Les collocatifs du pivot *bière* dans les sous-corpus**

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>boire</i>	verbe	35	713,1845
<i>bouteille</i>	nom	24	288,1791
<i>gorgée</i>	nom	16	204,7591
<i>commander</i>	verbe	19	178,5021
<i>prendre</i>	verbe	17	29,6939

Tableau 8 Les collocatifs du pivot *bière* dans POL

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>boire</i>	verbe	68	950,1346
<i>canette</i>	nom	31	671,0113
<i>bouteille</i>	nom	35	278,3050
<i>commander</i>	verbe	26	246,7599
<i>gorgée</i>	nom	24	235,7974
<i>verre</i>	nom	34	186,0987
<i>frais</i>	adjectif	16	134,2760
<i>servir</i>	verbe	19	33,6650

Tableau 9 Les collocatifs du pivot *bière* dans GEN

**Annexe 10**  
**Les collocatifs du pivot *vodka* dans les sous-corpus**

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>bouteille</i>	nom	15	280,8458

Tableau 10 Les collocatifs du pivot *vodka* dans POL

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>bouteille</i>	nom	19	244,8336
<i>verre</i>	nom	15	165,0457

Tableau 11 Les collocatifs du pivot *vodka* dans GEN



**Annexe 11**  
**Les collocatifs du pivot *café* dans les sous-corpus**

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>tasse</i>	nom	52	1733,3311
<i>boire</i>	verbe	49	1274,5376
<i>gorgée</i>	nom	33	500,8689
<i>préparer</i>	verbe	29	280,3858
<i>commander</i>	verbe	33	272,1281
<i>thermos</i>	nom	16	233,1507
<i>bol</i>	nom	21	217,4931
<i>lait</i>	nom	16	184,0082
<i>servir</i>	verbe	29	172,8869
<i>offrir</i>	verbe	24	156,5519
<i>brûlant</i>	adjectif	17	151,7094
<i>noir</i>	adjectif	27	148,7638
<i>chaud</i>	adjectif	21	145,8034
<i>aval</i>	verbe	18	131,2046
<i>siroter</i>	verbe	17	125,0640
<i>verser</i>	verbe	21	122,6979
<i>froid</i>	adjectif	16	97,9720
<i>apporter</i>	verbe	22	97,8613
<i>prendre</i>	verbe	42	97,3587
<i>odeur</i>	nom	21	72,6861
<i>proposer</i>	verbe	15	30,3585

<i>autre</i>	adjectif	21	29,8140
<i>bon</i>	adjectif	21	25,8876
<i>petit</i>	adjectif	15	20,1379
<i>chercher</i>	verbe	19	10,9050

Tableau 12 Les collocatifs du pivot *café* dans POL

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>tasse</i>	nom	18	980,8135
<i>boire</i>	verbe	25	762,0541
<i>bol</i>	nom	15	161,6575
<i>prendre</i>	verbe	21	135,6228
<i>apporter</i>	verbe	18	71,6732

Tableau 13 Les collocatifs du pivot *café* dans SENT

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>tasse</i>	nom	77	2017,2237
<i>boire</i>	verbe	104	1973,8685
<i>lait</i>	nom	39	615,4999
<i>bol</i>	nom	29	386,3190
<i>préparer</i>	verbe	44	309,1450
<i>commander</i>	verbe	33	202,6784
<i>servir</i>	verbe	49	188,8710
<i>noir</i>	adjectif	30	149,3671
<i>prendre</i>	verbe	79	142,9835
<i>apporter</i>	verbe	39	133,2828

<i>aval</i>	verbe	28	125,6545
<i>gorgée</i>	nom	20	99,8613
<i>brûlant</i>	adjectif	15	97,0700
<i>chaud</i>	adjectif	21	91,7481
<i>odeur</i>	nom	31	72,8228
<i>fort</i>	adjectif	15	57,5321
<i>verser</i>	verbe	16	50,3659
<i>offrir</i>	verbe	24	28,9713
<i>petit</i>	adjectif	27	27,0232
<i>proposer</i>	verbe	17	26,3074

Tableau 14 Les collocatifs du pivot *café* dans GEN

**Annexe 12**  
**Les collocatifs du pivot *thé* dans les sous-corpus**

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>tasse</i>	nom	31	759,1279
<i>boire</i>	verbe	30	392,3465
<i>gorgée</i>	nom	20	246,8535
<i>préparer</i>	verbe	18	169,2749
<i>servir</i>	verbe	17	75,4689
<i>prendre</i>	verbe	21	16,6809

Tableau 15 Les collocatifs du pivot *thé* dans POL

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>tasse</i>	nom	21	601,4418
<i>salon</i>	nom	16	284,3828
<i>boire</i>	verbe	16	221,2842
<i>prendre</i>	verbe	15	46,1511

Tableau 16 Les collocatifs du pivot *thé* dans SENT

<b>Collocatif</b>	<b>Catégorie</b>	<b>Dispersion</b>	<b>LLR</b>
<i>tasse</i>	nom	57	1360,9462
<i>boire</i>	verbe	76	997,0058
<i>menthe</i>	nom	20	219,4797
<i>salon</i>	nom	31	216,0655

<i>vert</i>	adjectif	21	185,5794
<i>servir</i>	verbe	46	182,8370
<i>préparer</i>	verbe	34	178,8923
<i>gorgée</i>	nom	24	176,2615
<i>verser</i>	verbe	28	172,4955
<i>verre</i>	nom	17	104,1510
<i>prendre</i>	verbe	58	78,3906
<i>apporter</i>	verbe	26	75,9444
<i>chaud</i>	adjectif	17	68,1002
<i>offrir</i>	verbe	17	18,5442

Tableau 17 Les collocatifs du pivot *thé* dans GEN

**Annexe 13**  
**Les diagrammes concernant la collocation *boire du vin* dans FY**

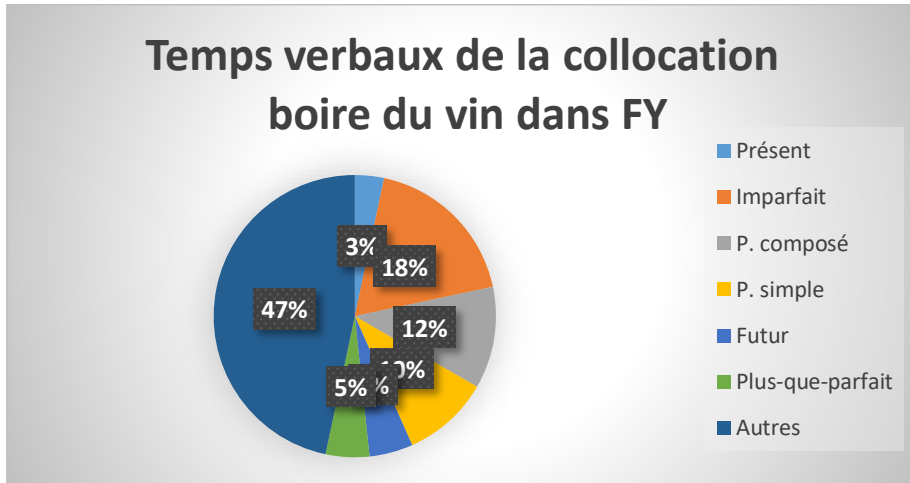


Figure 1 La distribution des temps verbaux de la collocation *boire du vin* dans FY

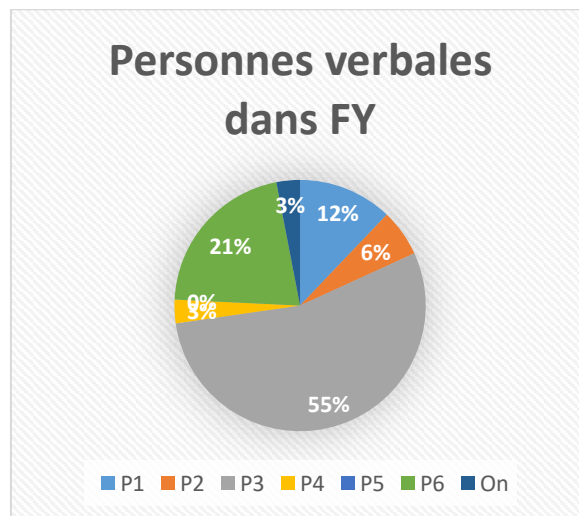


Figure 2 La distribution des personnes verbales de la collocation *boire du vin* dans FY

## Annexe 14

### Les diagrammes concernant la collocation *boire du vin* dans HIST

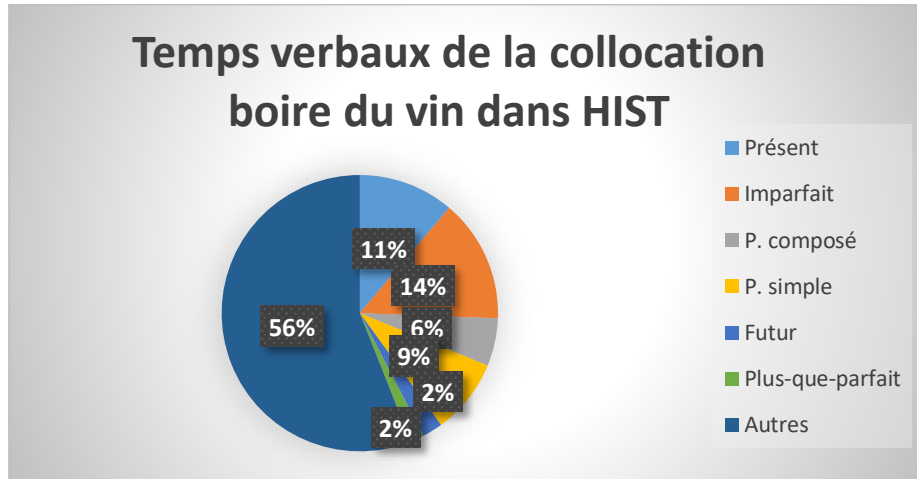


Figure 3 La distribution des temps verbaux de la collocation *boire du vin* dans HIST

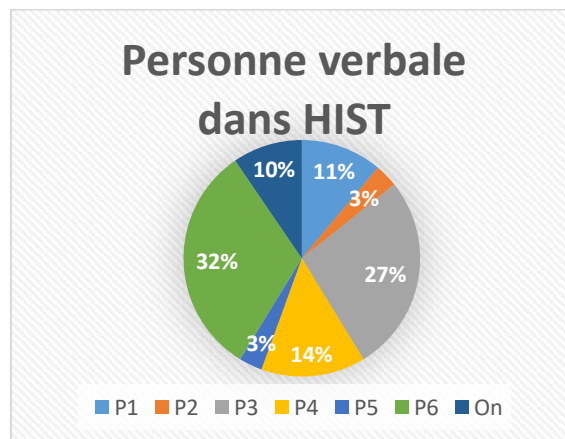


Figure 4 La distribution des personnes verbales de la collocation *boire du vin* dans HIST

**Annexe 15**  
**Les diagrammes concernant la collocation *boire du vin* dans SENT**

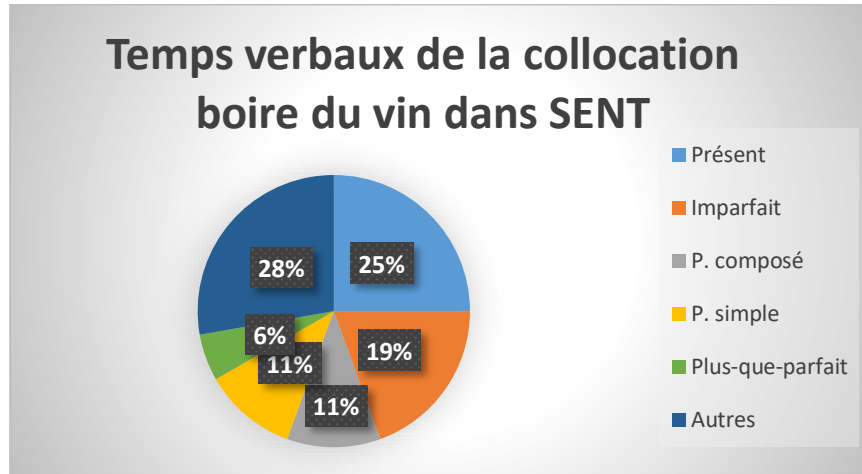


Figure 5 La distribution des temps verbaux de la collocation *boire du vin* dans SENT

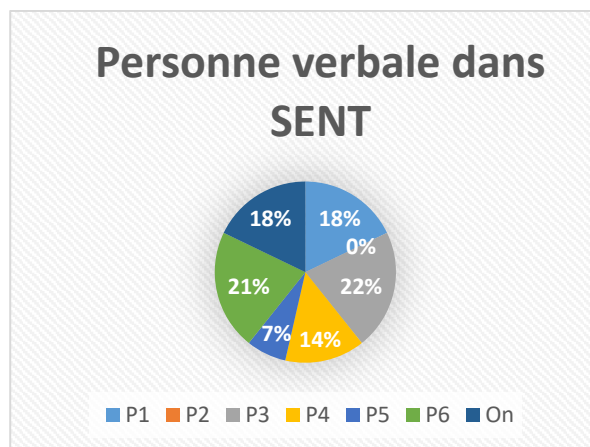


Figure 6 La distribution des personnes verbales de la collocation *boire du vin* dans SENT



## Annexe 16

### Les diagrammes concernant la collocation *boire du vin* dans POL

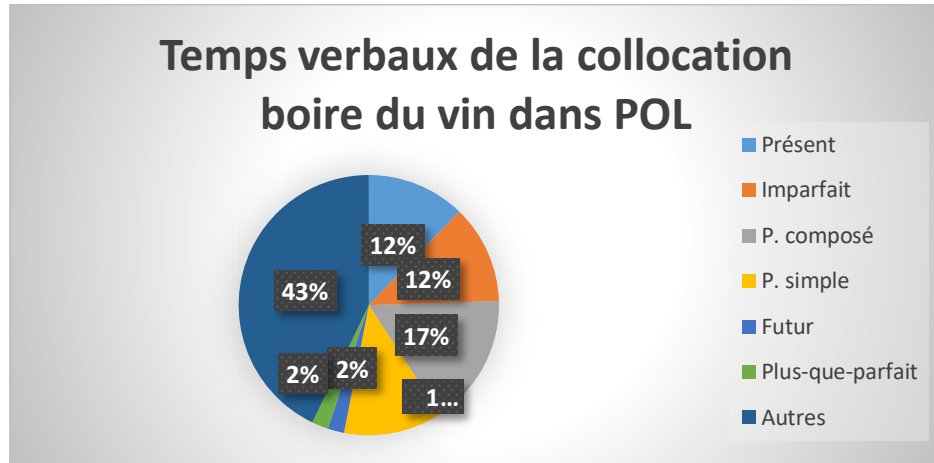


Figure 7 La distribution des temps verbaux de la collocation *boire du vin* dans POL

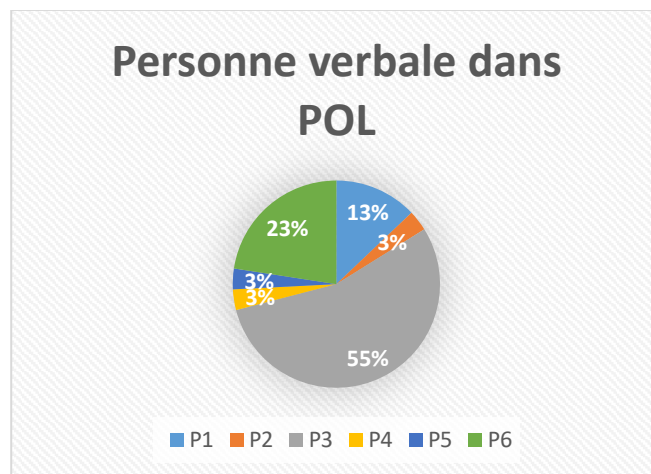


Figure 8 La distribution des personnes verbales de la collocation *boire du vin* dans POL

# Table des matières

Remerciements.....	3
Introduction.....	7
<b>PARTIE 1 - CADRE THEORIQUE.....</b>	<b>10</b>
CHAPITRE 1. LA LINGUISTIQUE DE CORPUS.....	11
1. La notion de linguistique de corpus.....	12
2. La notion de corpus.....	13
3. Le Traitement Automatique des Langues (TAL).....	15
CHAPITRE 2. LA PHRASEOLOGIE.....	18
1. Le domaine de la phraséologie.....	18
2. La collocation.....	19
3. La collocation grammaticale et la colligation.....	23
CHAPITRE 3. LA QUESTION DU GENRE TEXTUEL.....	26
1. La notion de genre et le genre romanesque.....	26
2. Les sous-genres textuels.....	28
2.1. Le roman policier.....	28
2.2. Le roman fantastique.....	30
2.3. Le roman de science-fiction.....	31
2.4. Le roman historique.....	32
2.5. Le roman sentimental.....	33
<b>PARTIE 2 - METHODOLOGIE.....</b>	<b>35</b>
CHAPITRE 4. LE LEXICOSCOPE.....	36
1. Présentation de l'outil et du corpus.....	36
2. Concordance et cooccurrence.....	38
3. L'arbre lexico-syntaxique récurrent (ALR).....	41
<b>PARTIE 3 - ANALYSE.....</b>	<b>47</b>
CHAPITRE 7. LA COMBINATOIRE DES NOMS DE BOISSON : APPROCHE QUANTITATIVE.....	48
1. Les collocatifs des noms de boisson.....	48
1.1. Les collocatifs du pivot vin.....	48
1.2. Les collocatifs du pivot champagne.....	50
1.3. Les collocatifs du pivot bière.....	51
1.4. Les collocatifs du pivot vodka.....	52
1.5. Les collocatifs du pivot café.....	52
1.6. Les collocatifs du pivot thé.....	54
2. Étude des collocatifs par sous-genre romanesque.....	56
2.1. La fréquence des collocatifs par sous-genre.....	56
2.1.1. Le cas de FY, SF et HIST : le sous-genre comporte un seul pivot avec des collocatifs.....	56
2.1.2. Le cas de SENT : le sous-genre comporte plusieurs pivots avec des collocatifs.....	57
2.1.3. Le cas de GEN et POL : le sous-genre comporte tous les pivots avec des collocatifs.....	57
2.2. La répartition des collocatifs par catégorie grammaticale dans les sous-genres.....	59
2.3. Les collocatifs spécifiques dans un seul sous-genre.....	60
CHAPITRE 8. LA COMBINATOIRE DES NOMS DE BOISSONS : APPROCHE QUALITATIVE ( <i>BOIRE DU VIN ET PRENDRE UNE BIERE</i> ).....	62
1. L'analyse de la collocation boire du vin.....	62
1.1. La variation sur le déterminant.....	63
1.2. Les variations temporelles et modales du verbe boire.....	69
1.3. Les personnes verbales de la construction boire du vin.....	73
2. L'analyse de la collocation prendre une bière.....	78
2.1. La variation sur le déterminant.....	78
2.2. Les temps verbaux du verbe prendre.....	79
2.3. Les modes verbaux de la collocation prendre une bière.....	80
2.4. Les personnes verbales de la construction prendre une bière.....	80

2.5. La variation avec l’adverbe et l’adjectif .....	81
2.6. La comparaison des collocations boire du vin et prendre une bière.....	82
Conclusion .....	85
Bibliographie .....	87
Sigles et abréviations utilisés .....	95
Table des annexes.....	97
Table des matières.....	118

**MOTS-CLÉS** : phraséologisme, construction lexico-syntaxique, collocation, genre textuel, linguistique de corpus

### **RÉSUMÉ**

L'objectif de ce mémoire est d'analyser sur le plan quantitatif et qualitatif les expressions contenant les noms de boissons dans les six sous-corpus de roman contemporain français : policiers, fantastiques, de science-fiction, historiques, sentimental et de littérature générale ou blanche. Le but est de mener une analyse syntaxique se focalisant sur le fonctionnement linguistique et l'entourage de constructions étudiées afin de relever si les collocations repérées peuvent servir à distinguer les sous-genres romanesques. Le travail est fondé sur le corpus littéraires du projet PhraséoRom- le *Lexicoscope*. Premièrement, les collocations statistiquement les plus spécifiques contenant les noms de boissons sont repérées. Nous proposons ensuite une analyse afin d'identifier les différentes variations des éléments qui accompagnent la construction analysée.

**KEYWORDS** : phraseological unit, lexico-syntactic structure, collocation, textual genre, corpus linguistics

### **ABSTRACT**

The objective of this work is to analyze quantitatively and qualitatively the expressions containing the names of beverages in six sub-corpora of contemporary French novels : detective, fantastic, science-fiction, historical, sentimental and general fiction (white). The main goal of this work is to carry out a syntactic analysis focusing on the linguistic functioning and the surroundings of studied constructions in order to note whether the identified collocations can be used to distinguish the fictional subgenres. The work is based on the literary corpus of the PhraséoRom project-Lexicoscope. Once the statistically most specific collocations containing the names of beverages are identified, an analysis is proposed in order to identify the different variations of the elements that accompany the constructions analyzed.

