



**HAL**  
open science

**Extended phraseological units and literary genres: A  
contrastive analysis of French and English  
lexico-syntactic constructions**

Iva Novakova, Marion Gymnich

► **To cite this version:**

Iva Novakova, Marion Gymnich. Extended phraseological units and literary genres: A contrastive analysis of French and English lexico-syntactic constructions. *Lexique*, 2021, 28. hal-03208828

**HAL Id: hal-03208828**

<https://hal.univ-grenoble-alpes.fr/hal-03208828v1>

Submitted on 18 Nov 2021

**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.

# Extended phraseological units and literary genres: A contrastive analysis of French and English lexico-syntactic constructions

Iva Novakova

Université Grenoble Alpes / Laboratoire LIDILEM (France)

[iva.novakova@univ-grenoble-alpes.fr](mailto:iva.novakova@univ-grenoble-alpes.fr)

Marion Gymnich

University of Bonn (Germany)

[mgymnich@uni-bonn.de](mailto:mgymnich@uni-bonn.de)

## Abstract

The present paper is based on the assumption that the language of the novel as well as that of its various subgenres is characterized by a statistically relevant overrepresentation of certain linguistic phenomena (e.g., lexemes, key words, collocations and colligations, Siepmann, 2015, 2016). Applying state-of-the-art lexicometric methods to extract recurring polylexical units in two large corpora of contemporary French and English novels, we explore the role of phraseological motifs in distinguishing literary subgenres. Unlike traditional corpus-stylistic analyses, which frequently focus on the style of a single author, our corpus-driven approach identifies features of literary (prose) genres on the basis of automatically extracted lexico-syntactic constructions (LSCs) that are statistically specific to a certain subgenre.

**Keywords:** phraseological motif, phraseological units, discursive functions, lexico-syntactic constructions, literary subgenres.

## Résumé

Notre article part de l'hypothèse selon laquelle la langue du roman et de ses différents sous-genres se caractérise par une surreprésentation statistiquement pertinente de certains phénomènes linguistiques (par exemple, les lexèmes, les mots clés, les collocations et les colligations, Siepmann, 2015, 2016). En appliquant des méthodes lexicométriques de pointe permettant d'extraire des unités polylexicales récurrentes dans deux grands corpus de romans français et anglais contemporains, nous explorons le rôle des motifs phraséologiques pour distinguer les sous-genres littéraires. Contrairement aux analyses de la stylistique classique, qui se concentrent souvent sur le style d'un seul auteur, notre approche *data-driven* ('guidée par les données') identifie les caractéristiques des genres littéraires (prose) sur la base de constructions lexico-syntaxiques (CLS) statistiquement spécifiques à tel ou tel sous-genre.

**Mots-clés :** motif phraséologique, unités phraséologiques, fonctions discursives, constructions lexico-syntaxiques, sous-genres littéraires.

## 1. Introduction

The present contrastive study is based on the assumption that a statistically relevant over-representation of certain linguistic phenomena (e.g., lexemes, keywords, collocations, colligations, phraseological units) is characteristic of literary language in general as well as of specific literary (sub)genres. This idea, which opens up innovative approaches to genre theory, was advocated by Siepmann in two seminal articles (2015, 2016). First steps towards checking the validity of this hypothesis had been undertaken in pioneering works in the 1990s–2000s (e.g., Stubbs & Barth, 2003). These studies were typically limited by the comparatively small size of their (exclusively English) corpora as well as by the fact that they did not include extended units of meaning, such as *His thoughts were interrupted by* | *Il en était là de ses réflexions quand* (see Siepmann, 2015, pp. 381-382;<sup>1</sup> 2016, p. 21), which readers are likely to recognise intuitively as features of literary texts. Biber *et al.* (1999) identify some grammatical constructions that are characteristic of literary texts (e.g., demonstratives of the type *that bloody car of mine* or expletive constructions including *there*); yet, they neither take keywords into consideration, nor do they examine the specificity of recurrent elements to subgenres, which promises to be a particularly interesting approach for bridging the gap between (corpus) linguistics and literary studies, as we hope to show in the following.

The case studies in Section 5 of this article go significantly further than Biber *et al.* (1999) by drawing upon large French and English literary corpora and by applying an innovative approach based on lexicometric methods. We specifically seek to explore the role of what Legallois and Tutin (2013) refer to as “*phraséologie étendue*” (‘extended phraseology’) in distinguishing between different subgenres of the novel. This particular approach to phraseology does not only focus on purely linguistic (syntactic-semantic) phenomena, but is also open to discursive aspects of the textual entities (see Diwersy & Legallois, 2019) that are analysed, in our case: literary subgenres. Unlike classical approaches within literary studies, we propose defining and describing subgenres not in terms of their reference worlds or their authors’ style, but on the basis of recurrent lexico-syntactic constructions that are statistically specific to a subgenre and that allow us to identify *phraseological motifs* (Legallois, 2012; Longrée & Mellet, 2013; Novakova & Siepmann, 2020a). The concept of the *phraseological* (or *textual*) *motif* thus constitutes an innovative analytical tool that makes it possible to link the analysis of the syntactic and semantic levels of literary texts with their discursive-textual level.

In the following, we will first provide a short overview of previous research on phraseology and stylistic approaches to literary genres (Section 2). In Section 3, we will define the concept of the phraseological motif. Subsequently (Section 4), we will describe our corpora and the methodology informing our study. Section 5 will be dedicated to a contrastive (French-English) linguistic and

---

<sup>1</sup> Expressions like *creased his brow in thought*, *frowned in thought* constitute extended units of meaning in literary texts, which resemble what Sinclair (2004, pp. 131-148) has called “lexical items” (Siepmann, 2015, p. 380).

stylistic analysis of data extracted from corpora of three subgenres of the contemporary novel (general fiction, crime novels, science fiction); in this section, we will present findings that prove to be interesting for comparing (sub)genres as well as the two languages. Our conclusions will be summarized in Section 6.

## 2. A brief overview of previous research

### 2.1. Approaches to phraseology and idiomaticity

The various concepts related to phraseology in the broad sense have given rise to diverse terms, such as “extended units of meaning” (Sinclair, 2004), “collostructions” (Stefanowitsch & Gries, 2003), “collocations” (Hausmann, 1979; Mel’čuk, Clas & Polguère, 1995; Siepmann, 2005; Tutin, 2010), “lexical bundles” (Biber *et al.*, 1999), “sequential patterns” (Quiniou, Cellier, Charnois & Legallois, 2012), “multi-word expressions” (Steyer & Brunner, 2014). As Novakova and Siepmann (2020a, pp. 4-5) point out, there is no consensus among linguists with regard to the labels used to identify idiomaticity, but there is at least a growing convergence among the different approaches that dispenses with the distinction between a grammar composed of rules and a lexicon consisting of words and phrases. Thus, the Neofirthian approach, whose most well-known proponent is certainly Hoey (2005), advocates for a grammatical lexicon containing both grammatical combinations (“colligations”: e.g., NP + to be + about + V-*ing*) and lexical combinations (“collocations”: *clear motorway*). However, his study is based on a small number of English lexemes and needs to be expanded to larger corpora or performed from a cross-linguistic perspective. Construction Grammars (CxG), on the other hand, treat language as an inventory of form/meaning pairs extending across a continuum, from lexicon to grammatical structures via idiomatic sequences (e.g., Goldberg, 1995; Fillmore, Kay & O’Connor, 1988; Croft, 2001). An obvious relationship exists between contextualism and certain construction grammars such as the “collostructions” of Stefanowitsch and Gries (2003). While the former starts from general constructions such as the ditransitive construction to identify its associated lexis, the latter instead start from individual lexemes (e.g., *give*) to arrive at general syntactic patterns. The strong convergence of these different approaches led to a shift in research focus from the fixed sequences of traditional phraseology, such as routine formulas, proverbs and binary collocations (Hausmann, 1979; Mel’čuk *et al.*, 1995), to other particularly promising concepts like “collocational frameworks” (Renouf & Sinclair, 1991) or “motifs” (Legallois, 2006, 2012; Longrée & Mellet, 2013; Novakova & Siepmann, 2020a). We will define the concept of the phraseological motif in Section 3 below.

## 2.2. Phraseology, stylistics and the theory of literary genres

A considerable number of studies in stylistics (e.g., Barthes, 1966; Greimas, 1972/1982; Leech & Short, 2007), corpus stylistics (Stubbs, 2005; Philippe & Piat, 2009; Fischer-Starcke, 2010; Vaudray-Luigi, 2013; Mahlberg, 2013), and textometry (Guiraud, 1954; Brunet, 1981) have provided sophisticated analyses of stylistic features, including studies of specific lexical and/or grammatical characteristics in works by individual writers like Flaubert, Proust, Dickens and Austen. Other studies have identified stereotypical recurrent patterns in specific genres, such as crime fiction (Todorov, 1978; François, 2009; Lits, 2011), with the goal of developing a typology of literary genres. Despite such efforts, genre theory as a whole has tended to rely first and foremost on content-based features, privileging thematic and macrostructural aspects rather than linguistic criteria in typologies of literary genres (for a discussion of approaches within genre theory, see Duff, 2000; Zymner, 2003; Frow, 2006; Monte & Philippe, 2014).

The thriving field of digital humanities makes it possible to complement predominantly content-based approaches to genre theory by innovative linguistic methods. That means in particular drawing upon the tools provided by corpus linguistics and Natural Language Processing (NLP) as well as relying upon the extraction of huge quantities of linguistic data for stylistic analyses (see e.g., Beauvisage, 2001; Rastier, 2011). The researchers collaborating in the context of the PhraseoRom project<sup>2</sup> have explored new ways of bridging the gap between quantitative, corpus-linguistic approaches and traditional goals of literary studies. They propose detailed analyses of lexico-syntactic constructions that are specific to subgenres of French and English novels. The usefulness of the concept of the phraseological motif for distinguishing between literary subgenres has been proven by various case studies completed within the PhraseoRom project (see Novakova & Siepmann, 2020b).

## 3. The concept of the phraseological motif: a definition

The term “motif” has been used across a range of different disciplines.<sup>3</sup> Here, we define the motif as a lexical-syntactic pattern whose statistical recurrence and specificity are established by means of textometric methods.<sup>4</sup> In concrete terms, phraseological motifs correspond to phraseological sequences that are statistically salient in a corpus, consisting of continuous or discontinuous units, combining lemmas, morphosyntactic categories, function words, and collocations. Linking form and meaning, these “multidimensional units” (Legallois, 2012, p. 45) fulfil pragmatic as well as discursive functions (DFs) in the structure of the text. On the micro-linguistic level, the motif constitutes a “collocational

---

<sup>2</sup> <https://phraseorom.univ-grenoble-alpes.fr>

<sup>3</sup> For an overview of the concept “motif” in disciplines like folklore studies, narratology, bioinformatics, NLP and linguistics, see Legallois & Koch (2020, pp. 17-46).

<sup>4</sup> See also Mellet & Longrée (2012).

frame” which consists of a combination of fixed and variable lexical and grammatical elements (see Longrée & Mellet, 2013, p. 66). These combinations manifest themselves as syntagmatic extensions of the collocational nucleus and as paradigmatic variations on its constituent elements (on the verb and/or on the noun). On the macro-textual level of the plot, these statistically significant recurrent lexico-syntactic patterns tend to play a vital role for the “textual coherence” (Martin, 1983, p. 16) within a literary text and are specific to a particular literary genre.

Phraseological motifs, as we understand them, are distinct from the “constructions” defined by Goldberg (1995). Motifs are structural units that are not defined from the point of view of language but with regard to the text. They have discursive functions, which is not necessarily the case for a construction as defined in Construction Grammar (CxG). For instance, transitive constructions in Balzac’s novels are not necessarily typical of Balzac; they are not always motifs.<sup>5</sup> Unlike “constructions”, motifs are primarily *textual* units. Thus, in our contrastive study, the statistically significant lexical-syntactic patterns that integrate the word *window* and its synonyms into general literature contribute to understanding the function of the window in the narrative (on this topic see also Vidotto & Goossens, 2020). We assume that the concept of the phraseological motif, though not widely known, especially in Anglo-American linguistics and stylistics, yields interesting results for a corpus-driven comparison of subgenres of contemporary novels in French and English.

#### 4. Corpus and methodology

The data for the present study has been extracted from two comparable, syntactically annotated French and English digital corpora of the PhraseoRom project.<sup>6</sup> These corpora consist of approximately 2,000 novels (published from the 1950s to the present) in the two languages and have been partitioned into six sub-corpora (largely on the basis of labels provided by publishing houses, library catalogues and literary awards): general literature (GEN), crime fiction (CRIM), romances (ROM), historical novels (HIST), science fiction (SF), and fantasy (FY). Tables 1 and 2 provide an overview of the composition of the corpora:

---

<sup>5</sup> We are grateful to Dominique Legallois for this example, which he has provided in a private discussion on this topic.

<sup>6</sup> The French and English corpora as well as parallel corpora (French-English and English-French) compiled by researchers in the PhraseoRom project are accessible on <http://phraseotext.univ-grenoble-alpes.fr/phraseobase/index.html>. “English” here is not synonymous with “Anglophone”; the English corpus consists of British and Irish novels, which warrants a more homogeneous corpus than an inclusion of American and Anglophone postcolonial novels would have provided.

Subgenre	Authors	Novels	Tokens
FY	43	109	13,966,566
GEN	172	442	34,593,554
HIST	40	114	14,193,642
CRIM	85	196	18,010,790
ROM	38	110	9,516,633
SF	40	150	13,528,393
<b>Total</b>	<b>418</b>	<b>1,121</b>	<b>103,809,578</b>

Table 1. The number of authors, novels and tokens<sup>7</sup> in the French PhraseoRom corpus.

Subgenre	Authors	Novels	Tokens
FY	57	153	22,531,553
GEN	104	169	19,461,070
HIST	52	128	19,339,040
CRIM	71	163	19,223,169
ROM	39	103	14,440,646
SF	53	133	16,705,651
<b>Total</b>	<b>376</b>	<b>849</b>	<b>111,701,129</b>

Table 2. The number of authors, novels and tokens in the English PhraseoRom corpus.

The corpus-linguistic approach pursued in the present paper is inductive or corpus-driven (Sinclair, 2004; Hoey, 2005; Biber & Conrad, 2009), being based on an automatic extraction of recurrent lexico-syntactic trees or RLTs (see Kraif, 2016), which display lexical units in their syntactic relations of dependency. Figure 1 below illustrates the RLT structure of one of the recurrent expressions specific to French crime fiction – *allumer une nouvelle cigarette* (‘light a new cigarette’):

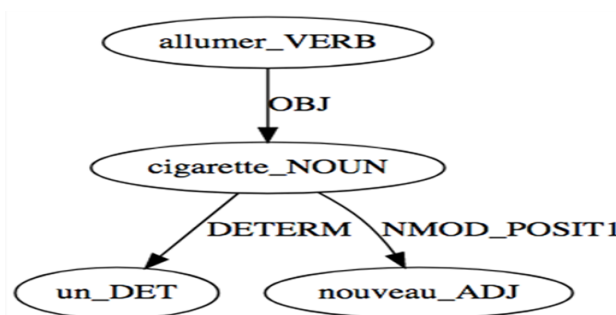


Figure 1. Automatic extraction of the RLT *allumer une nouvelle cigarette* (specific to CRIM).

The specificity of recurrent expressions (around a core word) to one of the subgenres has been determined by taking the entire PhraseoRom corpus into consideration and drawing upon the statistical

<sup>7</sup> Tokens here correspond to the following items in our corpora: words, punctuation marks, dates, and numbers.

index of log-likelihood ratio (LLR, see Dunning, 1993). We decided to use this classical statistical index from textometry for extracting syntactic sub-trees, thus opting for an economical architecture and a simple operation for the PhraseoBase database, which reduces the computation time and avoids heavy machine-processing of textual data, which is currently used in deep learning.

In the framework of the methodology developed within the PhraseoRom project (for more details, see Diwersy *et al.*, 2021), RLTs have to meet the following criteria:<sup>8</sup>

- a THRESHOLD OF STATISTICAL SPECIFICITY (LLR index) of more than 10.83.<sup>9</sup> The extraction of RLTs is iterative. The LLR index is used at each iteration. For example, in a first step, “open + window” is extracted by our Lexicoscope tool (Kraif, 2016) because its specificity threshold exceeds 10.83. Then, the software extracts “open window + the”, and the LLR is computed again for the pair “open + the window”, and so on. As the operation of our tool is based on syntactic dependency relations, there is no predefined limit with regard to the number of words that are “allowed” between the two items *open* and *window*.
- FREQUENCY IN ABSOLUTE TERMS: this criterion refers to the number of times an occurrence appears in a sub-corpus. The number of occurrences of an RLT must be higher than 10.
- DISPERSION: the number of texts (frequency of documents) or authors (frequency of authors) in which the expression – in the form of an RLT – appears. This criterion makes it possible to retain RLTs which are not part of the style of one particular author but which are specific to several authors within the same sub-genre. Thus, we retain RLTs which concern at least 50% of the authors in one of the sub-corpora of the PhraseoBase.
- MORPHO-SYNTACTIC CRITERION: each RLT must contain a verb; this criterion excludes, for instance, purely referential expressions like *à la tombée de la nuit* (‘after nightfall’), *Monsieur le Procureur* (‘Mr. District Attorney’), *les nains de jardin* (‘the garden gnomes’). We have chosen to retain those verbal expressions which present more paradigmatic variation and form “richer” patterns on the syntagmatic level.

By applying these criteria we obtained a total of 8,455 French and 1,915 English RLTs for the six sub-genres. Then, these RLTs were annotated according to a semantic grid developed specifically for the PhraseoRom project (see Goossens *et al.*, 2020). This grid makes use of primary semantic categories

---

<sup>8</sup> For a detailed comparison between the extraction of RLTs and other methods used to extract textual motifs, see Legallois & Koch (2020, pp. 35-36).

<sup>9</sup> Specificity indicates whether the frequency of an occurrence in a sub-corpus deviates from its average frequency in the entire corpus. If the observed relative frequency is much higher than the average relative frequency, the specificity will be high and positive. If it is much lower, it will be high and negative. It is the measure of log-likelihood that makes it possible to calculate it. The threshold of 10.83 allows us to gauge objectively whether the distribution of linguistic units within a corpus is random or not.



(DIMENSIONS), such as *action*, *communication*, *cognition*, as well as secondary semantic categories (VALUES). Table 3 below illustrates the outcome of this two-tier process<sup>10</sup>:

DIMENSIONS	VALUES	Examples
Action	Movement	<i>lever la main</i> ('lift one's hand'), <i>clapped his hands</i>
	Travel	<i>traverser le hall</i> ('cross the hall'), <i>walk along the beach</i>
Setting	Place	<i>l'appartement était petit</i> ('the apartment was small'), <i>was in town</i>
	Time	<i>le temps était gris</i> ('the weather was grey'), <i>spend a week</i>
Communication	Verbal	<i>le capitaine dit</i> ('the captain says'), <i>have a conversation</i>
	Physical	<i>lever les yeux vers lui</i> ('lift one's eyes towards him'), <i>acknowledge with a nod</i>
Cognition		<i>je ne me souviens pas</i> ('I don't remember'), <i>think for a moment</i>
Qualia	Sensation	<i>avoir froid</i> ('feel cold'), <i>my heart is beating</i>
	Affect	<i>se mettre en colère</i> ('to get angry'), <i>I'm shocked</i>

Table 3. A section of the semantic grid developed within the PhraseoRom project.

The semantic annotation and harmonisation of the RLTs was a prerequisite for their automatic clustering,<sup>11</sup> which has led to the emergence of phraseological motifs. In our methodology, the RLTs constitute the starting point for the identification of motifs. In a third step, the RLTs thus classified and clustered were analysed lexico-syntactically across syntagmatic and paradigmatic varieties within the RLTs – e.g., *allumer / éteindre / fumer (nerveusement) une (nouvelle) cigarette* ('light/extinguish/smoke (nervously) a (new) cigarette') –, which are part of our definition of motifs.

<sup>10</sup> This grid has been devised in the tradition of WordNet ontologies (Fellbaum, 1998) and *functional group clusters* (Mahlberg, 2007). The user's manual for the semantic grid is available (in French and English) on <http://phraseotext.univ-grenoble-alpes.fr/phraseobase/ressources-fr.html>.

<sup>11</sup> The automatic clustering of all of the RLTs has been accomplished with the help of Word2vec software. The Word2vec algorithm uses a neural network model to learn word associations from a large text corpus.

Finally, we analysed the discursive functions of the motifs. An elaborate system of annotation for assigning discursive functions (DFs) to phraseological motifs has been devised.<sup>12</sup> The analysis of a large number of motifs extracted from the sub-corpora has led to enriching the list of DFs beyond the two basic and most common ones – the *narrative* and the *descriptive* DFs (see Adam, 2011, p. 267). While the most frequent DFs contribute to structuring a literary text by connecting sequences of actions (*narrative* DF) or are associated with descriptive passages (*descriptive* DF), further DFs include the *affective* DF (= description of affects and emotions), the *pragmatic* DF (= characters' speech acts), and the *cognitive* DF (= presentation of cognitive processes). The *infra-narrative* DF is assigned to motifs that serve to flesh out a scenario (e.g., by introducing a sequence of minor actions) without having any actual consequences for the main action. Similarly, the *infra-descriptive* DF is assigned to motifs that merely provide a descriptive detail. As the exemplary analyses of three motifs that are specific to different subgenres in Section 5 will illustrate, DFs make it possible to pay attention to the textual dimension of RLTs and their structural role. The case studies will also show that there is no one-to-one correspondence between motifs and DFs; instead, one and the same motif can fulfil different DFs.

Thus, in line with functional and contextualist models (Sinclair, 2004; Hoey, 2005), our analysis of motifs includes the four levels of linguistic analysis: the lexical, syntactic, semantic, and discursive levels. The functional analysis including these four parameters allows us to articulate, on the one hand, the micro-linguistic level of the semantic values of the extracted RLTs and their lexico-grammatical variation, observed within the sentence, and, on the other hand, the macro-level of the narrative script through studying discursive functions which can go beyond the sentence level. The identification of discursive functions is based on observations of a wider context (*i.e.*, a few sentences before and after the one in which the motif occurs).

## 5. A contrastive analysis of three motifs

From among the many motifs identified in the categorisation process described above, we have selected three that are specific to different subgenres. The motifs which will be analysed in the following are particularly suitable for a contrastive analysis given that their LLR is high in the French as well as in the English sub-corpora.

### 5.1. The motif of the *window/fenêtre* in general fiction

Several LSCs built around the noun *fenêtre/window* are among the RLTs that are specific to the sub-corpora of general fiction (GEN) in French and English. Still, from a contrastive perspective, there are

---

<sup>12</sup> For a complete list and definitions of the various DFs defined in the PhraseoRom project, see Appendix B in Novakova & Siepmann (2020b, pp. 291-293).

a number of differences between the two sub-corpora. For example, <go to the window> is specific to HIST (LLR 14.08), but virtually non-existent in GEN (LLR 3.49), in contrast to the French corpus, where <aller à/vers la fenêtre> ('to go to/towards the window') is specific to GEN (LLR 11.19). The LLR of the RLTs <ouvrir la fenêtre> ('open the window') | <open the window> and <regarder par la fenêtre> ('look out the window') | <stand at the window>, however, is high in both languages, which suggests that these constructions are particularly interesting candidates for a contrastive analysis:

GEN FR	LLR	GEN EN	LLR
<ouvrir la fenêtre>	67.42	<open the window>	32.94
<regarder par la fenêtre>	27.16	<stand at the window>	20.71

Table 4. RLTs with the nominal core *fenêtre/window* specific to French and English GEN.

The following examples illustrate the syntactic distribution of <ouvrir la fenêtre> and <open the window> :

- (1) Il fut tout de même étonné de se retrouver vivant le lendemain ; étonné, mais aussi un brin émerveillé. **Il ouvrit la fenêtre** et huma avec avidité l'odeur ambiante de papaye et de fleurs sauvages. (Monénémo T., *Le Roi de Kahel*, 2008)  
'He was nevertheless surprised to find himself living on the following day; surprised, but also a little delighted. He opened the window and inhaled avidly the predominant scent of papayas and wild flowers.'
- (2) Il s'est levé, il **a ouvert la fenêtre**, s'est penché, il n'y avait personne dans le passage. (Cintas P., *Carabin Carabas*, 2016)  
'He stood up, he opened the window, leant out, there was nobody in the passage.'
- (3) Je leur dis bonsoir, **ouvris la fenêtre**, respirai une seconde l'air délicieusement frais de la campagne, la nuit, et me précipitai vers mon lit. (Sagan F., *Un profil perdu*, 1974)  
'I said good night to them, opened the window, inhaled for a second the deliciously fresh country air, the night, and I threw myself on my bed.'
- (4) Amin flipped up the blind, **opened the window** and looked out. (Newby P.H., *Something to Answer For*, 1969)
- (5) Mary **opened the window** and stepped out. (McEwan I., *The Comfort of Strangers*, 1981)
- (6) They **opened the window** and walked out on to the balcony. (Thomas D.M., *The White Hotel*, 1981)

An examination of their syntactic distribution reveals that both <ouvrir la fenêtre> and <open the window> occur especially in juxtaposition in French (2-3) and in coordination in English (4-6), with a higher specificity index in English than in French for the latter type of construction: the LLR for the coordinating conjunction *and* is 63.21. The conjunction *et* ('and'), by contrast, does not even appear in the lexicogram (the table of contingency) of <ouvrir la fenêtre>.

As the examples above demonstrate, both <ouvrir la fenêtre> and <open the window> typically occur within (more or less extensive) sequences of mundane actions, which usually have no impact on major events of the narrative: *se lever et se pencher* (2), *dire bonsoir* (3), *look out* (4), *step out* (5), *walked out on to the balcony* (6). The RLT <open the window> is most often followed by the verb

*look* (LLR 10.22) and, somewhat less frequently, by *sit* (LLR 7.24). The data also shows that in French the verbs co-occurring most often with <ouvrir la fenêtre> are *s'accouder* ('to lean on one's elbows', LLR 23.9), *aller* ('go', LLR 20.39), *se pencher* (to lean out, LLR 20.24) and *se lever* ('to stand up', LLR 19.53). Instead of fulfilling a *narrative* DF by marking steps in the progression of the narrative, the motif is much more likely to be part of a portrayal of everyday life. It serves to flesh out a situation and render it more plausible due to the attention to detail created by references to (small-scale) actions which do not have any direct impact on the main plot (= *infra-narrative* DF). These conclusions about the main function of <ouvrir la fenêtre> and <open the window> resemble findings on the motif <regarder par la fenêtre> ('to look out of the window') by Vidotto and Goossens (2020), who observe that <regarder par la fenêtre> also tends to be linked with an *infra-narrative* DF in GEN.

In the wake of Romanticism, the motif of the (open) window was associated with the *affektive* DF in French and English literature as well as in visual arts. In a study of 19<sup>th</sup>-century painting, Rewald (2011, p. 3) argues that "the Romantics found a potent symbol for the experience of standing on the threshold between an interior and the outside world" in the motif of the open window. Quite similar overtones appear in British novels from the 19<sup>th</sup> century, such as E. Brontë's *Wuthering Heights* (1847), where a delirious female protagonist implores her servant to "Open the window again wide" (163) to allow her to see the moors: "I wish I were out of doors – I wish I were a girl again, half savage and hardy, and free..." (163). Here, the motif is primarily associated with an *affektive* DF (expression of emotional states). In a similarly emotionally intense scene, the title character in Ch. Brontë's *Jane Eyre* (1847) opens a window and compares her life to a "prison-ground" (117), yearning for a "varied field of hopes and fears, of sensations and excitements" (116). Vidotto and Goossens (2020, p. 50) argue that the motif of the window has also played an important role in French literary history, drawing upon the authority of Baudelaire, who claims in "Fenêtres": "il n'est pas d'objet plus profond, plus mystérieux, plus fécond, plus ténébreux, plus éblouissant qu'une fenêtre éclairée d'une chandelle" ('There is no object deeper, more mysterious, more fertile, darker, more dazzling than a window lit by a candle').<sup>13</sup>

What also proves to be interesting in terms of DFs is a syntactic distribution of the motif that is specific to French GEN in the PhraseoRom corpus, namely the syntagmatic extension of the RLT <ouvrir la fenêtre> by an infinitive construction introduced by the preposition *pour* (LLR 20.41) + Vinf, which expresses the aim of the action: *pour aérer* ('to ventilate' LLR 56.6), *pour réfléchir* ('to think'), *pour faire entrer de l'air frais* ('to let in fresh air'):

- (7) **J'avais ouvert les deux fenêtres pour faire entrer un peu d'air frais**, mais cela n'avait servi à rien. (Modiano P., *Dimanches d'août*, 1986)  
'I had opened the two windows to let in a little fresh air, but that had not helped.'

<sup>13</sup> Baudelaire, Charles. "Fenêtres", *Le Spleen de Paris*. Paris : Michel Lévy, 1869, pp. 109-110.

- (8) Elle **ouvrit la fenêtre** pour revenir à sa réflexion de tout à l'heure, tandis qu'ils évoquaient des souvenirs. (Cintas P., *Carabin Carabas*, 2016)  
 'She opened the window to return to her previous meditations, when they had evoked memories.'

In (8), the act of opening a window is linked with reflexion and remembering, which means that the *infra-narrative* DF is combined with a *cognitive* DF in these instances. Opening the window is also associated with perception (e.g., (7) *pour faire entrer un peu d'air frais*). Such constructions are frequently linked with information on the characters' emotions, for example pleasure or even delight (ex. (1): *Il ouvrit la fenêtre et huma avec avidité l'odeur ambiante de papaye et de fleurs sauvages* or ex. (3): *je respirai une seconde l'air délicieusement frais de la campagne*). In these cases, the *infra-narrative* DF of the motif is augmented by an *affective* DF. On the whole, the enrichment of the *infra-narrative* DF by either a *cognitive* or an *affective* DF appears to be much more common in the case of French <ouvrir la fenêtre> than with its English equivalent.

In contrast to the motif <open the window>, which typically appears in sequences of trivial, small-scale actions and thus primarily fulfils an *infra-narrative* DF (even more so than its French equivalent <ouvrir la fenêtre>), the motif <stand at the window>, associated with the semantic categories of state and location (see the semantic grid in Table 3), is primarily associated with a *descriptive* DF, given that it typically initiates a sequence detailing what a character perceives. The affinity of <stand at the window> with a *descriptive* DF coincides with the fact that the verb *look* (used in the *ing*-form, *looking down, up*) is among the most specific lexemes co-occurring with the RLT <stand at the window>, but the RLT is also combined with other verbs of visual perception, such as *seeing, watching, gazing out (at him)*.

- (9) **I shall stand** for five minutes **at the window** of the rôtisserie, watching the chickens on the spit (Harris J., *Chocolat*, 1999)  
 (10) Tomáš **stands at the windows**, smoking and looking at the view. (Mawer S., *The Glass Room*, 2009)

Although the motif frequently occurs in its minimal form, there is also some syntagmatic variation, which may enhance the *descriptive* DF. Besides modifiers (adjectival and prepositional phrases (PPs)) of the noun *window* (*little, big, rear, kitchen, passenger, tall, of her hotel suite, of the glass room, of my flat, of the house*), the adverbial adjuncts between *open* and *window* (*for a long time, for five minutes, for a minute*) are among the common syntagmatic extensions of the RLT *open the window*. Many of these extensions provide the reader with a clearer sense of what the setting looks like.

All in all, we can conclude that the two motifs examined above – <ouvrir la fenêtre> / <open the window> and <stand at the window> – fulfil primarily an *infra-narrative* and a *descriptive* DF, respectively. A combination with an additional DF, especially a *cognitive* or *affective* DF, appears to be significantly more common in the French than in the English corpus. Given the tradition of linking these motifs with the depiction of emotional states and/or contemplation, which is very prominent in French and British literature in the 19<sup>th</sup> century, the widespread lack of *cognitive* or *affective* DFs in

contemporary English GEN might be interpreted as an attempt to avoid what could be perceived as a cliché due to having been used extensively in the past. At any rate, the motifs seem to have been largely deprived of their former association with emotions and contemplation in English GEN, being in effect trivialised, which is reflected in <open the window> being associated predominantly with an *infra-narrative* DF.

## 5.2. The motif *découvrir le corps* / *find the body* in crime fiction

Crime fiction has the reputation of being one of the most formulaic, “highly codified” (Ciocia, 2015, p. 109) genres. Its formulaic nature is apparent in a tendency to fall back on stereotypical character types, recurrent character constellations and plot patterns. Yet, this repetition has not been detrimental to the success of the genre; instead, a balance between familiar ingredients and a certain degree of variation seems to be what readers of crime fiction expect. Even in the more experimental types of crime writing, readers typically find a few basic building blocks, including, first and foremost, an investigator, a perpetrator and a victim (see Todorov, 1978, p. 9–19). Since crime fiction, which initially often told stories about theft or blackmail, has come to focus almost exclusively on murder since the 1920s (see Worthington, 2011, p. 117), there is usually a corpse.

Textometric corpus analysis offers tools that promise to identify links between patterns on the macro-level of the plot and the micro-level of phraseological motifs. Expressions referring to the discovery of a murder victim are among the statistically most salient RLTs in the French and English crime fiction (CRIM) sub-corpora, which correlates with the genre’s preoccupation with murder and suggests that the discovery of the victim is a central component of the macro-level of the plot:

CRIM FR	LLR	CRIM EN	LLR
< découvrir corps/cadavre >	159.71/40.64	< find body >	105.79
< trouver corps/cadavre >	34.59/35.72	< discover body >	66.65
< retrouver corps/cadavre >	47.57/25.86	< body was found >	42.29
< voir corps/cadavre >	39.85/10.37		
< examiner corps >	50.43		

Table 5. RLTs with the nominal core *corps/cadavre/body* specific to French and English CRIM.

While the English texts clearly favour the (more euphemistic) term *body*, there is some paradigmatic variation between the lexemes *corps* (‘body’) and *cadavre* (‘corpse’) in French. As a comparison of the LLR indices reveals, the number of RLTs built around the nominal core *corps/cadavre* in the French sub-corpus is higher than that of RLTs around *body* in the English sub-corpus. What the motifs

in both sub-corpora have in common, however, is that the RLTs often occur in their minimal form and are to be found frequently in direct speech and specifically in interrogative sentences:

- (11) Trois autres soldats lourdement armés apparurent dans le couloir et entrèrent dans la pièce. Frewin enchaîna: – Qui **a découvert le corps**? (Chattam M., *Le cycle de l'homme 2 Prédateurs*, 2007)  
‘Three other heavily armed soldiers appeared in the hall and entered the room. Frewin continued: – Who discovered the body?’
- (12) – Mr Main, who **discovered the body**? (Forbes C., *The Main Chance*, 2005)
- (13) I found him, you know, she said. I **found the body**. (Beaton M.C., *Agatha Raisin and the Quiche of Death*, 1993)

The RLTs listed above reflect the fact that the discovery of a (murder) victim is among the recurrent plot elements in crime fiction; in other words, the motif constitutes a part of what might be called an “investigation script”.<sup>14</sup> Since the motif refers to a crucial event in the plot and is an important trigger for further actions (the investigation), a *narrative* DF can be assigned to this motif.

In the French examples, there is a tendency for the RLT to occur in cleft sentences. Given that this type of construction serves as a “focusing device” (Fischer, 2009, p. 171), the syntactic distribution pattern may highlight the significance of a certain character (namely the one who found the body) in the investigation script:

- (14) Qui êtes-vous au juste ? **C’est moi qui ai découvert le corps d’Eduardo**. (Grangé J.-C., *La Forêt des Mânes*, 2009)  
‘Who are you actually? It’s me who discovered Eduardo’s body.’
- (15) **C’est la femme de ménage qui a trouvé le corps**, elle vient une fois par semaine, le mardi matin. Pas de chance, dit Adamsberg. (Vargas F., *Pars vite et reviens tard*, 2001)  
‘It is the cleaning woman who found the body; she comes once per week, on Tuesday mornings. No luck, said Adamsberg.’

Similar cleft sentences are extremely rare in the English sub-corpus:

- (16) There isn’t any – I mean – well – **there was a girl who discovered the body**. (Christie A., *The Clocks*, 1963)

---

<sup>14</sup> The notion of the “script” used here draws inspiration from Baroni (2007, p. 175), who employs the term “script” to refer to trivial, everyday actions that need no further explanation, such as lighting a cigarette or drinking coffee. While the investigation of a crime *as such* may not be an ordinary action, much of this procedure can certainly be classified as “routine” from the point of view of seasoned readers (and authors) of crime fiction due to the formulaic nature of the genre. Thus, we suggest expanding the concept of the “script” to refer to recurrent, stereotypical scenarios in formulaic fiction.

Given that cleft sentences as such are not particularly unusual in English (see Fischer, 2009),<sup>15</sup> the reason for the lack of cleft sentences stressing the role of the one who discovered the body might be a stylistic one, as we will suggest below.

There is usually more information about the victim's identity than about the one who finds the corpse. This is how we can account for the frequent syntagmatic extensions of the RLT, which are used in simple sentences and introduce details about victims (their identity, age, state, etc.). Other syntagmatic extensions specify the time and/or location of the murder:

- (17) The very-respectable Colonel and Mrs Bantry have awakened to **discover the body** of a young woman in their library. (Christie A., *A Murder is Announced*, 1950)
- (18) A fisherman **found the body** of a small, female child on a beach on Cat Island. (Bagley D., *Wyatt's Hurricane & Bahama Crisis*, 1982)
- (19) The gardener **discovered the body**, but not until this afternoon. (Forsyth F., *The Day of the Jackal*, 1972)

In (17) and (18), the identity of the victim (the body *of a young woman, of a small, female child*) as well as the place where the body has been found (*in their library, on Cat Island*) are mentioned. In (19) the motif is used essentially in its minimal realisation; the only detail that appears here is the temporal adverbial (*not until this afternoon*). In English, the motif is more likely to occur in subordinate clauses than in French:

- (20) He was still in the park when the kids **found the burning body**. (May P., *The Firemaker*, 1992)
- (21) I'm wondering whether she was taking care to ensure that there was a witness with her when she found the body (James P.D., *The Private Patient*, 2008)

These subordinate clauses are introduced most often by the conjunction *when*, which even turns out to be the most specific element co-occurring with the motif (with an LLR of 51.65). Thus, in narrative sequences, the motif tends to indicate the moment when a corpse is found.

Among the French examples, elaborate temporal and spatial information as well as details on the identity and state of the victim are twice as frequent as in the English examples:

- (22) Un flic baignait dans son sang à l'intérieur de la voiture banalisée, la carotide tranchée. Puis **ils avaient découvert le corps** de son collègue tué par balle dans son propre appartement. (Molay F., *La 7e femme*, 2006)  
'A cop was bathing in his own blood inside the civilian car, with his carotid artery cut through. Then they had found the body of his colleague who had been killed by a bullet in his own apartment.'
- (23) Tendresse **découvrit le cadavre** de sa grand-mère dans la pièce principale, couché devant la télévision allumée. Le décès était récent. (Besson P., *Mais le fleuve tuera l'homme blanc*, 2009)

---

<sup>15</sup> In his contrastive study (English-German), Fischer observes: "Clefts are used naturally in spoken, and in written, English texts: in journalistic articles, political speeches, narrative and academic writing alike." (2009, p. 175)



‘Tendresse found the corpse of her grandmother in the main room, lying in front of the television, which was still on. The death was recent.’

- (24) **On découvre le corps** du capitaine Duvale le lundi à l’aube, dans son bureau, une balle logée entre les deux yeux. (Azincourt J., *Nuit de rage*, 2001)

‘They discovered the body of Captain Duvale on Monday at the break of dawn, in his office, with a bullet lodged between his eyes.’

In the syntagmatically extended versions of the motif <découvrir corps>, there are prepositional phrases specifying the victim’s identity (*de son collègue, de sa grand-mère, du capitaine Duvale*) as well as temporal (*lundi, à l’aube*) and spatial adverbials (*dans son propre appartement, dans la pièce principale, dans son bureau*). There are also adjectival and nominal detached constructions describing the modalities of the murder or the state of the corpse (*tué par balle, couché devant la télévision allumée, une balle logée entre les deux yeux*).

Another feature that distinguishes the English and French data is the frequent usage of the passive construction *the body was found* (LLR 42.29), which, again, occurs most often in subordinate clauses (25), though not exclusively (26):

- (25) Their shoes matched the marks on the ground where **the body was found**, and his blood was on their clothes. (Mina D., *Field of Blood*, 2005)

- (26) **The** badly decomposed **body** of a man in his 50s **was found** by council workers at a block of flats in Briarstone yesterday. (Haynes E., *Human Remains*, 2013)

In these examples, the passive construction directs attention to the discovery of a victim, often without focusing on the one who found it. The main purpose of the construction <the body was found> is conveying information on a place, which is typically expressed in a prepositional phrase following the construction. As the data from the English-French parallel corpus shows, this construction tends to be translated into active constructions in French, which indicates a tendency of French to avoid the passive:

- (27) a. But if Kenamoun’s **body had not been discovered**, Ineny would have no idea that Huy was aware of his treachery. (Gill A., *City of the Dead*, 1993)

- (27) b. Mais si **l’on n’avait pas encore découvert le cadavre** de Kenamoun, le secrétaire n’aurait aucune idée que Huy était informé de sa trahison.

This difference is the consequence of stylistic preferences in the two languages, but has no consequences for the information that is conveyed, since the French indefinite pronoun *on* does not specify an agent, either.

As our analysis has shown, we can assign a *narrative* DF to all of the motifs with the nominal core *body/corps* in the two sub-corpora. The motifs mark progress in an investigation and tend to be combined with the presentation of evidence, since both the place where a victim is found and the state the victim is in, *i.e.*, injuries indicating how the murder was committed, may provide important clues.

Our linguistic conclusions corroborate Saint-Gelais' comments on the two different approaches to reading crime fiction:

On conçoit dès lors que le roman policier soit traversé par une tension entre deux angles de lecture: d'une part, un angle cognitiviste, qui amène à négliger la matérialité textuelle au profit des opérations inférentielles effectuées sur la base des données fictives ; d'autre part, un angle discursif, qui prend en compte les agencements du texte et leur incidence propre sur la lecture<sup>16</sup>. (Saint-Gelais, 1997, p. 797)

The varied linguistic manifestations of the motifs with the nominal core *corps/body* correlate with a progression of the investigation (plot level) drawing upon evidence provided by the body of the victim, which is in line with the cognitive appeal of the genre. A predominantly *narrative* DF is apparent in the French and English sub-corpora.

Despite this similarity, there are differences between the sub-corpora, e.g., more syntagmatic variation and a tendency to identify the person(s) who discovered the victim in the French examples. Leaving out the information who found the body in the English examples (often by means of the construction <the body was found>) suggests that familiarity with the underlying investigation script makes it possible for readers to cope with logical gaps in the textual realisation of the cognitive script. Providing information on the person(s) who found the corpse, let alone their reaction to the discovery, may have little value as evidence in a clue puzzle. Still, the decision to elaborate on the discovery of the corpse (or not) may have far-reaching aesthetic consequences, which are exemplified by the contrast between forensic crime fiction with its interest in decomposing bodies and detective novels in the British Golden Age tradition, which tend to shy away from gore and prefer what Worthington (2011, p. 120) refers to as “muted violence”.

### 5.3. The motif of *the screen/l'écran* in science fiction

Given the range of themes characteristic of science fiction, it seems hardly surprising that analyses of lexical fields have revealed a clear bias towards a “scientific” vocabulary in this genre (see Klein, 2018). The preference for (pseudo-)scientific lexemes goes hand in hand with a striking lexical creativity that serves to introduce new concepts and contributes to the strong world-building component inherent in SF (see Stockwell, 2000, pp. 115-138; Wolf, 2012, pp. 97-106). As Gonon and Kraif (2020, p. 152) have shown in their quantitative study, “neologisms are among science fiction's (SF's) defining features; these refer to realities unfamiliar to readers, such as futuristic inventions, advanced technology, anthropomorphic species from other planets, extraterrestrial fauna and flora, that

---

<sup>16</sup> “It is therefore conceivable that the crime novel is subject to a tension between two reading angles: on the one hand, a cognitivist angle, which leads to the neglect of textual materiality in favour of inferential operations carried out on the basis of fictitious data; on the other hand, a discursive angle, which takes into account the arrangements of the text and their own impact on reading.” [*The translation is ours.*]

have no referents in reality”. While neologisms like *tricomputer* or *vodoceur* (Gonon & Kraif, 2020, p. 164) stand out as world-building strategies, recurrent LSCs, which have remained largely unexplored in SF so far, may likewise establish a “scientific” vocabulary.

Our corpus-driven study has revealed that several constructions with the nominal core *l'écran/the screen* are statistically relevant to the French and English SF sub-corpora:

SF FR	LLR	SF EN	LLR
< apparaître sur l'écran >	110.66	< appear on the screen >	43.25
< s'inscrire sur l'écran >	32.68	< look the screen > <sup>17</sup>	53.00
< défiler sur l'écran >	16.90	< stare the screen >	27.72
< l'écran montre >	97.76	< the screen shows >	62.16
< l'écran s'éteint >	81.19		
< l'écran s'allume >	52.08		

Table 6. RLTs with the nominal core *écran/screen* specific to French and English SF.

The LLR index of the RLT < apparaître sur l'écran > ('appear on the screen') is very high, and the LLR of its paradigmatic variants < s'inscrire sur l'écran > ('register on the screen') and < défiler sur l'écran > ('scroll on the screen'), which are part of the same motif, are likewise well above the threshold of 10.83. The RLT < appear on the screen > is specific to English SF, albeit with a lower LLR (43.25). There are two further RLTs with the nominal core *écran/screen* that are highly specific to SF in both languages: < l'écran montre > (LLR 97.76) and its English equivalent < the screen shows > (LLR 62.16). In English, < look the screen > and < stare the screen > are also specific to SF, whereas their French equivalents < regarder l'écran > ('look at the screen') and < voir l'écran > ('see the screen') do not pass the threshold of 10.83; < l'écran s'éteint > ('the screen goes blank') and < l'écran s'allume > ('the screen lights up') have no statistically significant counterparts in the English corpus.

Computer and communication screens were of course much more futuristic for readers in the 1950s than they are today, in the wake of the digital revolution. Still, screens seem to have become a standard ingredient of SF. In some cases, the overall impression of a “scientific” fictional world is reinforced by syntagmatic extensions that specify the type of screen (*de l'holoviseur, de contrôle, du radar, téléviseur, sphérique, digital, video, of the communicator*). In the following, we will focus on parallels between the two sub-corpora and compare the functions of the RLTs < apparaître sur

<sup>17</sup> At first sight, the lack of prepositions may be surprising in the case of < look the screen > and < stare the screen >, but these are the RLTs identified by the methodology which we explained above. While these expressions are often associated with the preposition *at*, other prepositions may also occur (*on, from, through, into*), which means that the preposition is *not* statistically relevant here, unlike in < appear on the screen >.

l'écran> / <appear on the screen> and <l'écran montre> / <the screen shows> as well as the motifs which these RLTs establish in the two languages.

A comparison of the RLTs <apparaître sur l'écran> and <appear on the screen> reveals striking similarities with respect to their features and functions. The syntactic distribution of the RLTs turns out to be almost identical in the two corpora; they appear either in simple or complex sentences (primarily in paratactic ones) or in sentences introducing direct speech. In simple and complex sentences, the apparition of an entity (e.g., a face, a person, an object, a message, numbers) on a screen typically marks the beginning of a new narrative sequence and thus drives the plot onward, in accordance with the inchoative value of the verb *apparaître/appear*. The motif thus has primarily a *narrative* DF. The following examples from the French and the English corpora illustrate the prime functions of the motif:

- (28) La pièce était vide, mais un écran qui surmontait le dispositif d'autoguidage devint lumineux. Le ronronnement d'un mécanisme se fit entendre et un être monstrueux **apparut sur l'écran**. (Guieu J., *L'invasion de la terre*, 1952)  
 'The room was empty, but a screen positioned above the self-guiding device became bright. The humming of a mechanism could be heard and a monstrous being appeared on the screen.'
- (29) Une phrase **s'inscrivit** peu à peu **sur l'écran** de l'ordinateur. Simultanément, un synthétiseur vocal la prononça. Tous entendirent la réponse de la fourmi. (Werber B., *2 La trilogie des fourmis 2 Le jour des fourmis*, 1992)  
 'A sentence gradually inscribed itself on the computer screen. Simultaneously, a voice synthesizer pronounced it. Everyone heard the answer of the ant.'
- (30) The room was quiet for almost a minute as images of Mars at different resolutions **appeared on the giant screen** on the wall. (Clarke A.C., *3 Rama 03 – The Garden of Rama*, 1991)
- (31) Then a pale blue light **appeared on the screen**, and he almost dropped it in his surprise. (Gibson G., *Against Gravity*, 2005)

Most of the motifs have been realised in their minimal form (NP + V), though there is some paradigmatic variation with regard to the verb in French: *apparaître/défiler/s'inscrire sur l'écran*. The sight of someone or something that is unknown, spectacular, perhaps threatening (*un être monstrueux*) or sublime (*images of Mars*) on a screen is apt to contribute to the "sense of wonder" (Sawyer, 2015) as well as to the suspense deemed typical of SF: the apparition of *un être monstrueux* (28), of *images of Mars* (30), or even of *a pale blue light* (31) on a screen seems likely to make the reader expect exciting events to unfold. Being associated with a *narrative* DF, the motif <apparaître sur l'écran> | <appear on the screen> has a significant impact on the plot, frequently setting a whole sequence of events in motion, e.g., in (29): *une phrase s'inscrivit sur l'écran* → *un synthétiseur vocal la prononça* → *tous entendirent la réponse de la fourmi*. As the analysis so far already suggests, constructions with the nominal core *écran/screen* are eminently suitable for telling the kind of stories readers expect from SF.

The second shared recurrent distribution pattern of the motifs <apparaître sur l'écran> and <appear on the screen> is their function of preparing the path for direct speech/dialogue: the appearance of someone on a screen often initiates the act of speaking or even a dialogue sequence. Thus, *visage* (LLR 107.97) and *face* LLR 46.54 in English, which are – besides *image* (LLR 63.66) in French – among the most specific subject NPs of the sentence combined with *apparaître sur l'écran* (*appear on the screen*), seem apt to introduce an act of communication. Typically, the one whose face appears is about to turn into a speaker:

- (32) Un Mokranien casqué **apparut sur l'écran**.  
 – Pourquoi nous avoir stoppés ? s'indigna-t-il.  
 – Depuis quand, Boukoniev, discutes-tu les ordres du Chef Letvinof ? (Guieu J., *Hantise sur le monde*, 1953)  
 'A helmeted Mokranian appeared on the screen. – Why did you stop us? he said indignantly. – Since when, Bukoniev, do you discuss the orders of Chief Letvinof?'
- (33) A bleary-eyed Karl, who had obviously not slept since their last conversation, **appeared on his screen**.  
 "Here it is", he said, exhaustion and triumph competing in his voice. (Clarke A.C., *Imperial Earth*, 1976)

Examples (32) and (33) suggest that the function of introducing direct speech/a dialogue tends to be distinct from that of being linked to a sense of wonder, which was discussed before.

The functions described so far differ from those that can be identified for the motif <l'écran montre> | <the screen shows>, which displays some paradigmatic variation in French – *montrer* ('show'), *se maculer* ('smear'), *s'afficher* ('display') – but not in English. This motif typically introduces descriptive sequences specifying what can be seen on the screen, be it a portrait or an action, which means the motif is primarily associated with a *descriptive* DF:

- (34) **L'écran** de contrôle du thorax **montrait** la foule des passagers groupés devant la terrasse surmontée de l'écran double où l'on voyait la Terre sur une moitié et l'espace à l'avant du Papillon sur l'autre. (Werber B., *Le papillon des étoiles*, 2006)  
 'The thorax control screen showed the crowd of passengers grouped in front of the terrace topped by the double screen where one could see Earth on one half and the space in front of the Papillon on the other.'
- (35) **L'écran montrait** maintenant le ciel constellé d'astres. La monstrueuse comète Yahoun obstruait une partie de l'horizon et plongeait la capitale nergalienne dans un bain de terrifiantes vapeurs toxiques. (Guieu J., *Nous les Martiens*, 1954)  
 'The screen now showed the sky studded with stars. The monstrous comet Yahoun blocked part of the horizon and plunged the Nergalian capital into a bath of terrifying toxic fumes.'

Especially in the French examples, the descriptive sequences tend to be quite detailed, elaborating on what can be seen on the screen (e.g., in (34): *la foule de passagers groupés devant la terrasse*, and in (35): *le ciel constellé d'astres*). The motif <l'écran montre> includes some variation with respect to the verb (*l'écran montre, révèle, s'allume, frémit, s'illumine*):

- (36) Tout à coup, **l'écran** vert du radar **montra** un point brillant qui apparaissait et disparaissait alternativement – filant en ligne droite – au fur et à mesure que tournait le repère partant du centre de l'écran. (Guieu J., *L'homme de l'espace*, 1954)  
 'Suddenly, the green screen of the radar showed a bright dot that kept appearing and disappearing – spinning in a straight line – progressing with the rotating mark originating from the center of the screen.'
- (37) Le commandant demeurait toujours invisible. Rien n'indiquait que ses intentions demeuraient conformes aux instructions belliqueuses du conseil des Six. Brusquement, dans la cabine radio, **l'écran du téléviseur s'alluma** et **révéla** la chambre de Jacques Dureur où s'entassaient pêle-mêle, des bouteilles, des fourrures, des pipes, des aliments. Les lèvres pâles et serrées, il ordonna : – Mettez-moi en communication avec le lieutenant Cardeau, promptement. (Curval P., *Le ressac de l'espace*, 1962)  
 'The commander still remained hidden. Nothing indicated that his intentions were still in accordance with the warlike instructions of the Council of the Six. Suddenly, in the radio cabin, the television screen lit up and revealed Jacques Dureur's room, where bottles, furs, pipes, food were piled up in a jumble. Through pale, tight lips, he ordered: – Put me immediately through to Lieutenant Cardeau.'

When the RLT is accompanied by an adverbial of time (*tout à coup*, *brusquement*) in coordinated and juxtaposed sentences, the entity appearing on the screen may contribute to advancing the action, while the descriptive component simultaneously fades somewhat into the background. In other words, there is a hybrid DF: *narrative* (here: something new is introduced into the story, which advances the plot) + *infra-descriptive* (= the newly introduced object or place is described). This pattern, which is quite common in French, may also introduce direct speech (as (37) illustrates).

There are clear parallels between <l'écran montre> and <the screen shows> in terms of the functions of the motif, since the English data also reveals that the motif is linked with a *descriptive* DF. The motif often forms the core of an independent sentence that describes what can be seen on the screen(s) – a location, an object, etc. – and may initiate an enumeration or even a longer descriptive passage:

- (38) **Other screens showed** ships, dozens of dark ships, dropping from the sky. They were gas-capable Mercatoria spacecraft, some as little as fifty metres long, others three or four times that size; soot-black ellipsoids with thick wings and sleek but rudimentary tailplanes and engine pods. (Banks I.M., *The Algebraist*, 2004)
- (39) **One of the five screens** on the wall opposite his desk **showed** a real-time weather-satellite image of Amarisk and the ocean to the west, both covered by spiral arms of cloud. (Hamilton P.F., *Reality Dysfunction – Emergence & Expansion*, 1996)

All in all, the motifs with the nominal core *écran/screen* show strong similarities in the two languages in terms of their functions. The DF of <apparaître sur l'écran> | <appear on the screen> is first and foremost *narrative*, while the motif <l'écran montre> | <the screen shows> is primarily associated with a *descriptive* DF (and occasionally a hybrid DF: *narrative* + *infra-descriptive*). These tendencies, which may be found in French and English SF alike, can presumably be attributed to narrative conventions of the genre rather than to linguistic differences.

## 6. Conclusion

As our case studies (Sections 5.1-5.3.) have shown, the concept of the phraseological motif proves to be a useful tool for linking the micro-level of recurrent lexico-syntactic constructions with the macro-level of the plot in narrative texts. In short, due to its dual function as a unit that is both structuring and distinctive (see Longrée & Mellet, 2013), the phraseological motif is a lexico-syntactic sequence that turns out to be useful for distinguishing subgenres of the novel:

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

Moreover, a scrutiny of extended phraseology makes it possible to approach the definition and description of literary genres on the basis of an innovative, quantitative methodology. More specifically, our case studies allow us to formulate a number of conclusions, which of course need to be validated in further analyses: (1) The similarities between English and French literature appear to be more pronounced in crime fiction and science fiction than in general fiction. This might point towards a greater convergence among popular genres as far as stylistic and structural features are concerned, irrespective of whether they were written by French or British/Irish writers. (2) The French motifs on the whole tend to display a higher degree of paradigmatic and syntagmatic variation than their English counterparts, which may be attributed to variation being a preeminent ideal of French literary language (see Philippe, 2016). (3) Our study also revealed a wide variety of DFs of the motifs, which enrich the list of the two most common ones – the *narrative* and the *descriptive* DFs. The detailed analysis of these DFs in English and French led us to include some additional DFs like the *affective* and *cognitive* ones. The functions of the motifs often overlap and show the significance of this phraseological sequence for the organisation and construction of the plot in novels belonging to the different subgenres.

## Bibliography

- Adam, J.-M. (2011 [1992]). *Les textes : types et prototypes*. Paris : Armand Colin.
- Baroni, R. (2007). *La tension narrative : suspense, curiosité et surprise*. Paris : Seuil.
- Barthes, R. (1966). Introduction à l'analyse structurale des récits. *Communications*, 8, 1-27.
- Beauvisage, T. (2001). Exploiter des données morphosyntaxiques pour l'étude statistique des genres : application au roman policier. *TAL*, 43. <http://www.revue-texto.net/Inedits/Beauvisage/index.html>.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Biber, D., & Conrad, S. (2009). *Register, Genre, and Style*. Cambridge: Cambridge University Press.

- Brontë, Ch. (1985 [1847]). *Jane Eyre*. Harmondsworth: Penguin.
- Brontë, E. (1985 [1847]). *Wuthering Heights*. Harmondsworth: Penguin.
- Brunet, E. (1981). *Le vocabulaire français de 1789 à nos jours*. Geneva / Paris : Slatkine et Champion.
- Ciocia, S. (2015). Rules are meant to be broken: Twentieth- and twenty-first-century crime writing. In C. Berberich (Ed.), *The Bloomsbury Introduction to Popular Fiction* (pp. 108-128). London: Bloomsbury.
- Croft, W. (2001). *Radical Construction Grammar*. Oxford: Oxford University Press.
- Diwersy, S., & Legallois, D. (2019). L'apport de la méthode des motifs aux analyses phraséologiques en discours. In M. Kauffer & Y. Keromnes (Eds), *Theorie und Empirie in der Phraseologie* (pp. 115-132). Tübingen: Stauffenburg.
- Diwersy, S., Gonon, L., Goossens, V., Kraif, O., Novakova, I., Sorba, J., & Vidotto, I. (2021). La phraséologie du roman contemporain dans les corpus et les applications de la PhraseoBase. *Corpus*, 22. <https://doi.org/10.4000/corpus.6101>
- Duff, D. (2000). *Modern Genre Theory*. London: Longman.
- Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence: computational linguistics. *Computational Linguistics*, 19(1), 61-74.
- Fellbaum, C. (1998). A semantic network of English: the mother of all WordNets. *Computers and the Humanities*, 32, 209-222.
- Fillmore, C., Kay, P., & O'Connor, C. (1988). Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language*, 64, 501-538.
- Fischer, K. (2009). Cleft sentences: form, function, and translation. *Journal of Germanic Linguistics*, 21(2), 167-191.
- Fischer-Starcke, B. (2010). *Corpus Linguistics in Literary Analysis: Jane Austen and her Contemporaries*. London: Continuum.
- François, M. (2009). Le stéréotype dans le roman policier. *Cahiers de narratologie*, 17. <http://narratologie.revues.org/1095>.
- Frow, J. (2005). *Genre*. London / New York: Routledge.
- Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Constructions*. Chicago: The University of Chicago Press.
- Gonon, L., & Kraif, O. (2020). French and American science fiction during the nineties: a contrastive study of fiction words and phraseology. In I. Novakova & D. Siepmann (Eds), *Phraseology and Style in Subgenres of the Novel: A Synthesis of Corpus and Literary Perspectives* (pp. 151-188). Cham: Palgrave Macmillan.
- Goossens, V., Jacquot, C., & Dyka, S. (2020). Science fiction versus fantasy: A semantic categorization and its contribution to distinguishing two literary genres. In I. Novakova & D. Siepmann (Eds),



- Phraseology and Style in Subgenres of the Novel: A Synthesis of Corpus and Literary Perspectives* (pp. 189-221). Cham: Palgrave Macmillan.
- Greimas, A.J. (1982 [1972]). *Essais de sémiotique poétique*. Paris : Larousse.
- Guiraud, P. (1954). *Les caractères statistiques du vocabulaire*. Paris : P.U.F.
- Hausmann, F.J. (1979). Un dictionnaire des collocations est-il possible ? *Travaux de littérature et de linguistique de l'université de Strasbourg*, 17(1), 187-195.
- Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. London/New York: Routledge.
- Klein, G. (2018). Préface. In G. Costes & J. Altairac, *RétrofictionS : Encyclopédie de la Conjecture romanesque rationnelle francophone, de F. Rabelais à R. Barjavel, 1532–1951* (pp. 5-8). Paris : Les Belles Lettres.
- Kraif, O. (2016). Le lexicoscope : un outil d'extraction des séquences phraséologiques basé sur des corpus arborés. *Cahiers de lexicologie*, 108, 91-106.
- Leech, G., & Short, M. (2007). *Style in Fiction: A Linguistic Introduction to English Fictional Prose*. London: Pearson.
- Legallois, D. (2006). Quand le texte signale sa structure : la fonction textuelle des noms sous-spécifiés. *Corela 5*. <https://journals.openedition.org/corela/pdf/1465>.
- Legallois, D. (2012). La colligation : autre nom de la collocation grammaticale ou autre logique de la relation mutuelle entre syntaxe et sémantique ? *Corpus 11*. <http://corpus.revues.org/2202>
- Legallois, D., & Tutin, A. (2013). Présentation : vers une extension du domaine de la phraséologie. *Langages*, 189, 3-25.
- Legallois, D., & Koch, S. (2020). The notion of motif where disciplines intersect: folkloristics, narrativity, bioinformatics, automatic text processing and linguistics. In I. Novakova & D. Siepmann (Eds.), *Phraseology and Style in Subgenres of the Novel: A Synthesis of Corpus and Literary Perspectives* (pp. 17-46). Cham: Palgrave Macmillan.
- Lits, M. (2011). *Le roman policier dans tous ses états : d'Arsène Lupin à Navarro*. Limoges : Pulim.
- Longrée, D., & Mellet, S. (2013). Le motif : une unité phraséologique englobante ? Étendre le champ de la phraséologie de la langue aux discours. *Langages*, 189, 68-80.
- Mahlberg, M. (2007). Clusters, key clusters and local textual functions in Dickens. *Corpora*, 2(1), 1-31.
- Mahlberg, M. (2013). *Corpus Stylistics and Dickens's Fiction*. London / New York: Routledge.
- Martin, R. (1983). *Pour une logique du sens*. Paris : PUF.
- Mel'čuk, I., Clas, A., & Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve : Duculot.

- Mellet, S., & Longrée, D. (2012). Légitimité d'une unité textométrique : le motif. In A. Dister, G. Purnelle, & D. Longrée, *Actes des JADT 2012 : 11e journées internationales d'analyse statistique des données textuelles* (pp. 715-728). <http://hdl.handle.net/2268/122518>
- Monte, M., & Philippe, G. (dir.). (2014). *Genres et textes : Déterminations, évolutions, confrontations*. Lyon : Presses universitaires de Lyon.
- Novakova, I., & Siepmann, D. (2020a). Literary style, corpus stylistic, and lexico-grammatical narrative patterns: Toward the concept of literary motifs. In I. Novakova & D. Siepmann (Eds), *Phraseology and Style in Subgenres of the Novel: A Synthesis of Corpus and Literary Perspectives* (pp. 1-15). Cham: Palgrave Macmillan.
- Novakova, I., & Siepmann, D. (Eds.). (2020b). *Phraseology and Style in Subgenres of the Novel: A Synthesis of Corpus and Literary Perspectives*. Cham: Palgrave Macmillan.
- Quiniou, S., Cellier, P., Charnois, T., & Legallois, D. (2012). Fouille de données pour la stylistique : cas des motifs séquentiels émergents. In *Actes des 11es journées internationales d'analyse statistique des données textuelles*. Liège, 13-15 June 2012 (pp. 821-833). <http://lexicometrica.univ-paris3.fr/jadt/jadt2012/Communications/Quiniou%2C%20Solen%20et%20al.%20-%20Fouille%20de%20donnees%20pour%20la%20stylistique.pdf>
- Philippe, G. (2016). *French style : l'accent français de la prose anglaise*. Bruxelles : Les Impressions Nouvelles.
- Philippe, G., & Piat, J. (2009). *La langue littéraire : une histoire de la prose en France de Gustave Flaubert à Claude Simon*. Paris : Fayard.
- Rastier, F. (2011). *La mesure et le grain : sémantique de corpus*. Paris : Honoré Champion.
- Renouf, A., & Sinclair, J. (1991). Collocational frameworks in English. In K. Aijmer & B. Altenberg (Eds), *English Corpus Linguistics: Studies in Honour of Jan Svartvik* (pp. 128-144). London: Longman.
- Rewald, S. (2011). *Rooms with a View: The Open Window in the 19<sup>th</sup> Century*. New Haven / London: Yale University Press.
- Saint-Gelais, R. (1997). Rudiments de lecture policière. *Revue belge de philologie et d'histoire*, Tome 75, Fasc. 3, 1997. Langues et littératures modernes – Moderne taal- en letterkunde. (pp. 789-804) [https://www.persee.fr/doc/rbph\\_0035-0818\\_1997\\_num\\_75\\_3\\_4196](https://www.persee.fr/doc/rbph_0035-0818_1997_num_75_3_4196).
- Sawyer, A. (2015). Science fiction: the sense of wonder. In C. Berberich (Ed.), *The Bloomsbury Introduction to Popular Fiction* (pp. 87-107). London: Bloomsbury.
- Siepmann, D. (2005). Collocation, colligation and encoding dictionaries. Part I: Lexicological aspects. *International Journal of Lexicography*, 18(4), 409-444.
- Siepmann, D. (2015). A corpus-based investigation into key words and key patterns in post-war fiction. *Functions of Language*, 22(3), 362-399.

- Siepmann, D. (2016). Lexicologie et phraséologie du roman contemporain : quelques pistes pour le français et l'anglais. *Cahiers de lexicologie*, 108(1), 21-41.
- Sinclair, J. (2004). *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
- Stefanowitsch, A., & Gries, S. (2003). Collocations: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209-243.
- Steyer, K., & Brunner, A. (2014). Contexts, patterns, interrelations: New ways of presenting multi-word expressions. In *Proceedings of the 10<sup>th</sup> Workshop on Multiword Expressions (MWE 2014), Gothenburg, Sweden, 26-27 April 2014* (pp. 82-88). <https://aclweb.org/anthology/papers/W/W14/W14-0814/>.
- Stockwell, P. (2000). *The Poetics of Science Fiction*. Harlow et al.: Longman.
- Stubbs, M. (2005). Conrad in the computer: Examples of quantitative stylistic methods. *Language and Literature*, 14(1), 5-24. doi/10.1177/0963947005048873
- Stubbs, M., & Barth, I. (2003). Using recurrent phrases as text-type discriminators: A quantitative method and some findings. *Functions of Language*, 10(1), 61-104.
- Todorov, T. (1978). *Poétique de la prose*. Paris : Éditions du Seuil.
- Tutin, A. (2010). *Sens et combinatoire lexicale : de la langue au discours*. Habilitation thesis, University of Grenoble. [https://www.academia.edu/7332851/HDR\\_Tutin](https://www.academia.edu/7332851/HDR_Tutin)
- Vaudrey-Luigi, S. (2013). *La langue romanesque de Marguerite Duras : une liberté souvenante*. Paris : Garnier.
- Vidotto, I., & Goossens, V. (2020). D'une fenêtre à l'autre. Étude d'un motif spécifique à la littérature blanche et au roman sentimental. In L. Fesenmeier & I. Novakova (dir.), *Phraséologie et stylistique de la langue littéraire / phraseology and stylistics of literary language : approches interdisciplinaires / interdisciplinary approaches* (pp. 49-70). Berlin : Peter Lang.
- Wolf, M.J.P. (2012). *Building Imaginary Worlds: The Theory and History of Subcreation*. New York/London: Routledge.
- Worthington, H. (2011). *Key Concepts in Crime Fiction*. Houndmills, Basingstoke: Palgrave Macmillan.
- Zymner, R. (2003). *Gattungstheorie: Probleme und Positionen der Literaturwissenschaft*. Paderborn: Mentis.