

The notion of *motif* at the crossroads of disciplines – folkloristics, narrativity, bioinformatics, automatic text processing, linguistics

Dominique Legallois and Stefan Koch

Abstract:

The aim of this chapter is to present the main elements necessary to understand the use of the term "motif" in different disciplines, ranging from 19th century folkloristics, narratology, bioinformatics, automatic text processing and linguistics. Although each discipline has its own specificities, it is possible to highlight common concerns, such as the definition of the motif, its nature (is a motif a thematic or a linguistic unit?), its identification. This article argues for collaboration between disciplines that are scientifically very distant, but which all face the problem of motif.

Keywords: motif, narratology, bioinformatics, theme, automatic text processing, linguistic unit.

1. INTRODUCTION

The aim of this chapter is quite ambitious since it consists in proposing a presentation and discussion of the notion of *motif* in several disciplinary fields. These disciplines can be very distant from each other, which at the same time complicates and enriches the task.

Why be interested in this notion of *motif* in the first place? It ~~has-seemsappeared~~ that this notion is one of the few that transcends (in part) ~~the~~ different disciplines, whether they be disciplines of the humanities or natural sciences, which often have little in common. Moreover, in linguistics, specifically in the subdisciplines concerned with semantic or stylistic characterisation of texts, the notion *motif* has undoubtedly enjoyed certain success in recent years. This success can be measured by the various proposals and innovations brought forward by different linguists, either to further develop this concept or to develop tools that allow the automated extraction of motifs from texts. In the present chapter, as in the different chapters of this book, applications of *motifs* are presented.

The term *motif* is extremely complex. Even when tackled within the same discipline and field of analysis, as in this publication, it is subject to somewhat different conceptions. In other words, *motif* remains a notion whose definition is not stable and that is constantly undergoing new developments. These developments are on the one hand technically driven by advances in computer technology, but also by conceptual innovations.

It therefore seems necessary to us to provide the linguist and the literary text specialist with historical, epistemological and methodological information on the use of the term *motif* in scientific research. The different approaches and draw inspiration from them for their own research, or put their research in a particular perspective. In return, non-linguists will be able to find, elements in the linguistic approach that will allow them to think through and develop their own conception of *motif*. The term *motif* and its different concepts seem to us to be a solid basis for a dialogue between subjects and disciplines that usually ignore each other.

The disciplines we will review first are narratology and folkloristics. In these disciplines, the notion is very well represented since the work carried out in the 19th century on popular traditions. It was then developed by structuralists and narratologists/semioticians.

Secondly, we will present the role of the term *motif* in information retrieval in science, whether from Harris' et al. (1989) perspective or the perspective of **bioinformatics**. These techniques used to analyse patterns of different kinds are not as far away – as we shall see – from some of the older approaches in **narratology**.

The third “family” of approaches is the one of **natural language processing** (NLP), which has recently developed the notion of *motif* as a particular linguistic unit that allows computational processing of texts and corpora: characterization, clustering, classification, etc.. This work was initiated by Köhler (2006).

Finally, over the past ten years, a series of studies from a linguistic perspective have been carried out in France, including the domain of computational linguistics, but trying to cover several dimensions of the text by recurring to stylometry, semantic characterizations, genre characterizations etc. The Franco-German, ANR-DFG-financed research project *PhraseoRom*, from which many of the contributions in this book stem, is part of this body of research.

At the end of this overview, the reader will have at his disposal a set of designs from which he will be able to valuable conclusions to be able to evaluate the possible ramifications of the concept of *motif*.

Since *motif* is also often being associated with music and painting, i.e. in non-linguistic or textual domains, our article could have been even more ambitious and addressed other “arts” than the text. For this article we limit our presentation and reflection to textual motifs, though, try to elaborate a specific application to literary texts at the end of the article.

2. MOTIF IN FOLKLORISTICS AND NARRATOLOGY

With the term *motif*, cultural anthropology and folkloristics refer to a recurrent element in the popular traditions of a cultural community. The combination of several *motifs* forms patterns, and allows researchers to describe, analyze and interpret the differences between one culture and another. Thus, in a wide sense, *motif* can be applied to the field of literature – including oral literature, such as tales –, visual and textile arts and music. In the latter two, a motif is then either any repeated decorative element or the smallest distinctive recurring musical line from a melody or rhythm. This musical *motif* is different from the one proposed by Boroda (1982), which we mention later.

2.1 TALES AND POPULAR NARRATIVES

For our purpose, the study of *motifs* in tales or popular narratives is of particular interest. There, *motifs* are recurrent micro-stories, recognizable thanks to a stable but malleable and flexible physiognomy, according to the texts in which they are instantiated. More precisely, they are stereotypical narrative sequences whose identity is definable and irreducible (Vincensini 2000, 2). See the following examples, some of which are developed in the work of Vincensini:

- “Punishment, sitting in the water”¹, which appears in many novels of the Middle Ages, but also in the Lives of the Saints.
- “The perilous cemetery”: the hero is confronted, at night, with supernatural opponents (Cadot 1980).

¹ This motif is also what in the 12th and 13th centuries are called EXEMPLA (sg. EXEMPLUM), i.e. short moral narratives given as true, used for educational purposes in sermons. They were also used to “christianise” traditional stories. There were EXEMPLA repertories, for example the collection SCALA COEDI by Jean Gobi around 1300 (Vincensini 2000, 13).

- “The constrained gift”: a donor makes unlimited offers without even being asked by the recipient. This motif appears in the Gospels of Saint Matthew as well as in the Arthurian legend.
- “The accusatory bleeding”: a corpse starts bleeding in the presence of the murderer. This motif is used in many medieval texts, but also in *le Second livre des Amours* by Ronsard, in *Richard III* by Shakespeare, in *les Tragiques* by Agripa d’Aubigné.
- “The animal skin delimiting a territory”: this motif is about a trick used by the hero to claim a large territory.

The conception of *motif* in ancient and traditional texts necessarily has to do with **intertextuality** and **collective memory**. In a way, *motifs* are always *leitmotifs* in a culture.

Based on the first collections made by the **Grimm brothers**, in the *Verzeichnis der Märchentypen*, published in 1910 by the Finnish folklorist Antti Aarne (1867–1925; see Aarne 1961 for an English version), **tale-types** (Märchentypen) are classified into categories, each defined by a number and a title in German. The concept of *tale-type* allows variation of the same tale in different countries or repertoires. The American folklorist Stith Thompson (for example 1946) continued this work, which in the end lead to the international classification of tales known as the **Aarne-Thompson classification**. But he also proposed a categorization of *motifs* whose result is the famous **Index of Motifs** in the six-volume *Motif-index of folk-literature: a classification of narrative elements in folktales, ballads, myths, fables, mediaeval romances, exempla, fabliaux, jest-books, and local legends*, published between 1932–1936².

Several main categorizations are distinguished:

a) mythological motifs; b) animal; c) tabu; d) magic; e) the dead; f) marvels; g) ogres; h) tests; j) the wise and the foolish; k) deceptions; l) reversal of fortune; m) ordaining the future; n) chance and fate; p) society; q) rewards and punishment; r) captives and fugitives; s) unusual cruelty; t) sex; u) the nature of life; v) religion; w) traits of character; x) humor; z) miscellaneous groups of motifs.

For each category, Thompson made sub-divisions. For example, the motives relative to *The Wise and the Foolish* are included in the following sub-divisions:

- J0-J199 Acquisition and possession of wisdom/knowledge
- J200-J1099 Wise and unwise conduct
- J1100-J1699 Cleverness
- J1700-J2799 Fools (and other unwise persons)

The many distinguished classes are the subject of the following grouping, made by Thompson in his book *The Folktale* (1946):

A motif is the smallest element in a tale having a power to persist in tradition. In order to have this power it must have something unusual and striking about it. Most motifs fall into three classes. First are the actors in a tale – gods, or unusual animals, or marvelous creatures like witches, ogres, or fairies, or even conventionalized human characters like the favorite youngest child or the cruel step-mother. Second come certain items in the background of the action – magic objects, unusual customs, strange beliefs, and the like. In the third place there are single incidents – and these comprise the great majority of motifs (Thompson 1946, 415–416)

²A revised and enlarged edition was published between 1955 and 1958.

One of the current conceptions of *motif* in narratology seems to correspond to Thompson's comments. The striking character of the *motif*, in a way, its expressive character, lies in its concrete value. In a recent reference book on narratology, Abbott (2008, 95), based on Prince (2003, 55) states, that "[m]otifs are, in Gerald Prince's words, the "minimal thematic unit" (Dictionary, 55)". He then goes on to define *motif* and *theme* (See Abbott 2008, 237, 242):

Motif: A discrete thing, image, or phrase that is repeated in a narrative. Theme, by contrast, is a more generalized or abstract concept that is suggested by, among other things, motifs. A coin can be a motif, greed is a theme. (Abbott 2008, 237).

Whereas:

Theme: A subject (issue, question) that recurs in a narrative through implicit or explicit reference. With motif, theme is one of the two commonest forms of narrative repetition. Where motifs tend to be concrete, themes are abstract (Abbott 2008, 242).

Abbott (2008, 95) also exemplifies his definitions:

Windows, for example, are a motif in *Wuthering Heights* and, given the way Brontë deploys them, they support a highly complex interplay of three themes: escape, exclusion, and imprisonment. When, for another example, the character Barkis in *David Copperfield* continues to repeat his cryptic phrase, "Barkis is willin'," it becomes a motif, a signature phrase for the theme of shy, honest-hearted devotion in love that Barkis exhibits in his pursuit of Peggotty.

The *motif* in contemporary narratology is therefore a minimal unit that "concretizes" the *themes*³.

2.2 THE RUSSIAN FORMALISTS

Since the 1920s the *motif* has also been considered, among Russian formalists, a minimum unit, but the difficulty of defining this minimality has also been pointed out. Additionally, these researchers develop elements for an articulation between motifs and linguistic forms.

Thus, for Tomaševskij (1966), there is an equation between motif and clause (See also Ducrot and Todorov 1972). To our knowledge, this is the first reflection on the linguistic dimension of *motif*. This point will be developed further below. Tomaševskij (1966) also proposes the idea that some *motifs* can be dynamic: they change the narrative situation, they make it evolve. Those who do not change the narrative situation are so-called stative motives. We recognize this as a first reflection on the *gestalt* complementarity between foreground and background or figure and ground. Tomaševskij is also the author of the famous distinction between fabula and story, sometimes referred to as the story/narrative distinction.

For Veselóvskij (1913a; 1913b), each proposal has its own *motif*, that is a smallest particle of the thematic material. The notion of *motif* is therefore undergoing a radical redefinition, which moves it away from the folklorist conception. However, according to Propp

³ See for example the helmet motif in L.F. Céline's *Casse-Pipe*, studied by Richard (1979)

(1968 [1928]), by seeking indecomposability in the linguistic clause, one misses the set of possible substitutions that make up the richness and variety of tales:

If a motif is something logically whole, then each sentence of a tale gives a motif. (A father has three sons: a motif; a stepdaughter leaves home: a motif; Ivan fights with a dragon: a motif; and so on). This would not be so bad if motifs were really indivisible; an index of motifs would then be made possible. But let us take the motif ‘a dragon kidnaps the tsar’s daughter’ (this example is not Veselóvskij’s). This motif decomposes into four elements, each of which, in its own right, can vary. The dragon may be replaced by Koščej, a whirlwind, a devil, a falcon, or a sorcerer. Abduction can be replaced by vampirism or various other acts by which disappearance is effected in tales. The daughter may be replaced by a sister, a bride, a wife, or a mother. The tsar can be replaced by a tsar’s son, a peasant, or a priest. In this way, contrary to Veselóvskij, we must affirm that a motif is not monomial or indivisible (Propp 1968, 13).

Propp looked for another method to replace Veselóvskij’s (1913b: 9–12) *motifs* with the notion of *narrative function*, that is, action defined in terms of its meaning in the development of the plot independently of the characters and the way in which these functions are fulfilled. Propp (1968) distinguishes – following the French philologist Bédier – the constant patterns from the variable patterns, and calls the former *function*, the latter *attribute*. The names of the characters change, but their actions or functions do not.

The 31 narrative functions in the traditional Russian tale are certainly *motifs* – narrative atoms –, but Propp seems to discard a cumbersome terminology. The functions (for example: *acceptance of the mission by the hero*; *departure of the hero*; *testing of the hero by a donor*, etc.), find their place in sequences; they are ordered and arranged. For example: *lack – compensation for lack*, *prohibition – violation of the prohibition*, *combat – victory*, et cetera.

2.2. SOME DEVELOPMENTS OF THE FORMALIST APPROACH

The impact of Propp’s model was very high from the 1960s onwards. An excellent example is the American folklorist Alan Dundes (1934–2005), who, using the terminology of the linguist Pike (1967), sought to apply Propp’s method to a corpus of Amerindian stories. Dundes (1962) proposes to rename the Proppian function *motifemes*. He also developed the term *allomotif* to designate the various forms in which the *motif* is realized in the tale. Allomotifs are to motifemes as allophones are to phonemes.

The Czech researcher Dolezel (1972) has repatriated the notion of *motif* into the field of literary text analysis. His approach has three levels: *motifeme*, *structural motif* and *texture motif*. The first level is metalinguistic. It consists in assigning a series of expressions designating an action to another series of expressions that define the actant. Dolezel thus defines the *motifeme* as a clause similar to Propp’s *function*. The second level, also metalinguistic, focuses on the narrative segments, where an action is assigned to a character. Indeed, for Dolezel, the structural motif is limited to dynamic motifs only. Finally, at the third level, the texture pattern corresponds to a sentence in the text itself. In the author’s words:

Examples of *motifeme strings*; (a) one-actant strings: The hero returned; The hero passed the test. (b) Two-actant string: The hero defeated the villain.

A MOTIF (m) is a proposition predicating an action to a character. Examples of motif strings: Ivan killed the dragon; Yanko rescued from death a grandchild of Swaffer; The judge sentenced Meursault to death.

A **MOTIF TEXTURE** is a narrative sentence (i.e. a sentence of the narrative text). Examples: *And suddenly, one fine morning in spring, he rescued from an untimely death a grandchild of old Swaffer* (Conrad). *Le president m'a dit dans une forme bizarre que j'aurais la tête tranchée sur une place publique au nom du peuple français* (Camus) (The president told me in a strange way that my head would be cut off in a public square on behalf of the French people).

The motifeme/motif relationship will be called SPECIFICATION. We will say that a motifeme is specified by a motif (motif sequence) or that a motif (motif sequence) specifies a motifeme. The motif/texture relationship will be called VERBALIZATION. A motif will be said to be verbalized by a texture; a texture will be said to verbalize a motif (Doležel 1972, 59–61; translation Camus DL).

These perspectives show the need to conceive *motif* at several levels: an abstract level (for example the *motifeme*) of a stereotyped entity, identifiable in a literary culture or tradition, which is linguistically embodied in various forms (for example the *texture motif*).

To conclude this panorama – necessarily incomplete – in narratology, it is important to address a very formal aspect of *motif*, which appears in particular in Propp's criticism of Vólkov. Vólkov considers that a *motif* can be a quality of the heroes (e.g. *two wise sons-in-law and the third madman*), an act of the heroes (e.g. *the injunction of a father so that someone watches over his grave after his death, an injunction that is executed by the madman alone*), an object (e.g. *houndstooths*, *talismans*). In Vólkov's annotation system, each of the *motifs* receives a conventional sign: a letter and one or two numbers. Similar *motifs* are marked by the same letter with a different number. This formalisation allows Vólkov to translate the *motifs* into alpha-numeric signs and thus to compare combinations. This method seems naive to Propp:

The only “conclusion” that can be drawn from this transcription is that similar tales resemble each other – a conclusion which is completely noncommittal and leads nowhere. (Propp 1968, 15).

What interests us is not the conception of *motif* in Vólkov's work, but rather his formalization process, which strikes us because of the analogy that can be seen with recent research, which does not concern narrative or tale, but a completely different field: bioinformatics. What is the relationship between bioinformatics and narratology? The *motifs* of Russian or Amerindian tales certainly take root in stories that are very different from those told by genome sequences. But perhaps both areas have something to share concerning methodology.

3. FROM NARRATOLOGY TO BIOINFORMATICS

With the notion of sub-language, Harris et al. (1989) show, how it is possible, through a set of formal transformations, to “translate” scientific reports into a sequence of formulas which represent the information contained in this reports. The work is mainly concerned with forming classes of words which have the same grammatical relation to particular other words. For example the words of class A which appear as subject in *offound in the lymph nodes after injection of an antigen*. These words A include *antibodies*, *agglutinin*, etc. Starting with this, we may then

form a class V of operators *is found in, is contained in, is produced by*, whose subjects are A (e.g. *antibodies, agglutinin*). In this way we find that certain word classes recur in a particular grammatical relation to certain others word classes, e.g. A as subject of V with object or “complement” T (*lymph nodes, lymph. serum*). This creates a sentence type (structure), AVT, which is obtained at the same time as we set up (extensionally) the word classes A, V, T, since these are defined as occurring in respect to each other in the operator-argument relation which constitutes a sentence structure. The grammatical transformations applied previously will have strengthened this result, since for example a sentence *Lymph nodes produce antibodies* will have been recognized as a transform of *Antibodies are produced by lymph nodes*, hence as a case of AVT (Harris et al. 1989, 8).

This procedure could be carried out by computer programs applied to scientific papers to identify recurrent informatively relevant patterns. With Darányi et. al. (2012) we consider a comparison with the morphology of tales of particular interest: despite their digitalization, tales have not, to our knowledge, been the object of precise computational analysis. Combinations of *motifs* could thus be analyzed from a very large number of occurrences. The method that would be used would be very similar to that used in Harris’ et al. sublanguages or to those used in *motif detection* in infobiology, for example to detect relations between genes and rare diseases in biomedical texts.

Since Frederick Sanger invented DNA sequencing in the second half of the 1970s, technological advances in this field have been groundbreaking so that the volume of DNA sequences available has increased exponentially. Infobiologists align sequences (linguists would say *concordances*) to recognize recurring sequences, called *motifs*. This makes it possible to find similarities between two sequences and determine their possible homologies. For example, in the genome sequences in figure 2.1 a certain pattern emerge (see D’haeseleer 2006: 423)

(2.1; from D’haeseleer 2006: 423)

```
HEM13  CCCATTGTTCTC
HEM13  TTTCTGGTTCTC
HEM13  TCAATTGTTTAG
ANB1   CTCATTGTTGTC
ANB1   TCCATTGTTCTC
ANB1   CCTATTGTTCTC
ANB1   TCCATTGTTCGT
ROX1   CCAATTGTTTGG
```

This domain-specific practice from life sciences can be recognized in formal descriptions of narrative content, too. Darányi et. al. (2012) therefore propose to deal with narrative *motifs* and *motifs* from biomedical texts in the same way. From a set of tales indexed and indexed^[SK1] in Aarne-Thompson-Uther Tale Type Catalog, the patterns receive an alpha-numeric annotation. Each tale is thus reduced to a set of annotations. -For example, the tale *The Dragon-Slayer* is represented in the following way, along with other tales which show a closed motif structure:

ATU tale type 300: The Dragon-Slayer.

[B421 B312.2] B11.2.3.1 [B11.10 S262] T68.1 D1975 D1978.2 [B11.11 B524.1.1 R111.1.3] H105.1 K1933 K1932 H151.2 [H83 H105.1]

Sequence variants

B421	B11.2.3.1	B11.10	T68.1	D1975	D1978.2	B11.11	H105.1	K1933	K1932	H151.2	H83
B421	B11.2.3.1	S262	T68.1	D1975	D1978.2	B11.11	H105.1	K1933	K1932	H151.2	H83
B421	B11.2.3.1	B11.10	T68.1	D1975	D1978.2	B524.1.1	H105.1	K1933	K1932	H151.2	H83
B421	B11.2.3.1	S262	T68.1	D1975	D1978.2	B524.1.1	H105.1	K1933	K1932	H151.2	H83
B421	B11.2.3.1	B11.10	T68.1	D1975	D1978.2	R111.1.3	H105.1	K1933	K1932	H151.2	H83
B421	B11.2.3.1	S262	T68.1	D1975	D1978.2	R111.1.3	H105.1	K1933	K1932	H151.2	H83
B421	B11.2.3.1	B11.10	T68.1	D1975	D1978.2	B11.11	H105.1	K1933	K1932	H151.2	H105.1
B421	B11.2.3.1	S262	T68.1	D1975	D1978.2	B11.11	H105.1	K1933	K1932	H151.2	H105.1
B421	B11.2.3.1	B11.10	T68.1	D1975	D1978.2	B524.1.1	H105.1	K1933	K1932	H151.2	H105.1
B421	B11.2.3.1	S262	T68.1	D1975	D1978.2	B524.1.1	H105.1	K1933	K1932	H151.2	H105.1
B421	B11.2.3.1	B11.10	T68.1	D1975	D1978.2	R111.1.3	H105.1	K1933	K1932	H151.2	H105.1
B421	B11.2.3.1	S262	T68.1	D1975	D1978.2	R111.1.3	H105.1	K1933	K1932	H151.2	H105.1
B312.2	B11.2.3.1	B11.10	T68.1	D1975	D1978.2	B11.11	H105.1	K1933	K1932	H151.2	H83
B312.2	B11.2.3.1	S262	T68.1	D1975	D1978.2	B11.11	H105.1	K1933	K1932	H151.2	H83
B312.2	B11.2.3.1	B11.10	T68.1	D1975	D1978.2	B524.1.1	H105.1	K1933	K1932	H151.2	H83
B312.2	B11.2.3.1	S262	T68.1	D1975	D1978.2	B524.1.1	H105.1	K1933	K1932	H151.2	H83
B312.2	B11.2.3.1	B11.10	T68.1	D1975	D1978.2	R111.1.3	H105.1	K1933	K1932	H151.2	H83
B312.2	B11.2.3.1	S262	T68.1	D1975	D1978.2	R111.1.3	H105.1	K1933	K1932	H151.2	H83
B312.2	B11.2.3.1	B11.10	T68.1	D1975	D1978.2	B11.11	H105.1	K1933	K1932	H151.2	H105.1
B312.2	B11.2.3.1	S262	T68.1	D1975	D1978.2	B11.11	H105.1	K1933	K1932	H151.2	H105.1
B312.2	B11.2.3.1	B11.10	T68.1	D1975	D1978.2	B524.1.1	H105.1	K1933	K1932	H151.2	H105.1
B312.2	B11.2.3.1	S262	T68.1	D1975	D1978.2	B524.1.1	H105.1	K1933	K1932	H151.2	H105.1
B312.2	B11.2.3.1	B11.10	T68.1	D1975	D1978.2	R111.1.3	H105.1	K1933	K1932	H151.2	H105.1
B312.2	B11.2.3.1	S262	T68.1	D1975	D1978.2	R111.1.3	H105.1	K1933	K1932	H151.2	H105.1

Figure 2.2: The Dragon Slayer as a motif chain and its equally valid story variants (Ofek et. al. 2013, 3).

DNA motif sequences and narrative motif sequences are identical in several respects, including some transformations:

adding grammar and moving over to permutations, sequences start to play a role. Canonical nucleotide sequences generate secondary and tertiary – in fact spatial – structures such as the famed double helix; canonical motif sequences may contribute to the evolution of tale types, themselves representatives of tale variants. Whereas most mutations are neutral and have little or no impact on the functionality of the product, their adding up can dramatically affect the survival rate of the outcome, leading to new genotypes and phenotypes in the course of evolution. In the same vein, deletion and translocation could be standard tools in the narrative building toolkit; inversion is suggested to play a central role in the Bible (Christensen, 2003), and duplication is evident e.g. in the case of the Proppian narrative scheme where complete tale moves may be repeated several times or combined with one another by different embeddings (Propp, 1968). This indicates the need for a theory of text evolution as a series of narrative element recombinations, forming from simple to more complex structures by “mutation mechanisms”. (Darányi et. al. 2012, 4).

These similarities in motif analysis in two very different scientific fields show the great potential of collaborations between natural scientists, computer scientists, narratologists and linguists.

4. MOTIFS IN AUTOMATIC TEXT PROCESSING

Very recently, the term *motif* has been used with a particular definition in the context of automatic text processing, at the initiative of Köhler (initially 2006; see 2015): a *motif* is defined as the longest continuous sequence of equal or increasing values representing a quantitative property of a linguistic unit. It is a matter of considering the motif as a unit that can reflect the sequential organization of a text, without resorting to grammatical or linguistic analysis. Without going into the details of the diversity of reasons, we give here some examples.

The following example is taken from Köhler (2015:92): the sentence “Word length studies are almost exclusively devoted to the problem of distributions.” is analysed according to the above mentioned definition of *motif* (longest continuous sequence of equal or increasing values...).

If we count syllables in each word (Word=1, length=1, studies=2, etc.), one obtains a sequence of 5 motifs, so called **word length motifs**: (1-1-2) (1-2-4) (3) (1-1-2) (1-4)

A word with a lower number of syllables than the previous word is the first element of a new *motif*. Applied to an entire text, the method makes it possible to observe, according to Köhler, that motifs display a rank-frequency distribution of the **Zipf-Mandelbrot** type: they behave in this respect in a way similar to traditional units of linguistic analysis. From this point, the method can be used to make comparisons between texts, between authors or to perform textual classification. But the example of “syllabic” patterns is not the only one. Motifs can be defined for any linguistic unit (phoneme, type of sentence, type of clause, etc.). A more semantic example is given by the Köhler (2015:98) as well: the segmentation of newspaper commentaries in argumentative relations: [“elaboration”], [“elaboration”, “concession”], [“elaboration”, “evidence”, “list”, “preparation”, “evaluation”, “concession”], [“evidence”, “elaboration”, “evaluation”]. Again, on the basis of these motifs, textual typologies can be established.

Finally Čech et al. (2017) take verbal valence as elements of the *motif*. To each verb of a text, they assign a number corresponding to the number of its arguments; we obtain something like in figure 2.3:

[4, 3, 3, 4, 2, 3, 2, 2, 1, 4, 3, 3, 3, 2, 0, 2, 1, 3, 3, 2, 1, 1, 1, 2, 2, 2, 2, 0, 2, 2, 2, 2, 0, 3, 3, 3, 1, 3, 1, 0, 1, 0, 2, 2, 1, 2, 1, 2, 1, 3, 3, 2, 2, 2, 1, 1, 1, 2, 2, 2, 1, 4, 2, 3, 2, 2, 2, 1, 2, 2, 4, 0, 2, 2, 2, 2, 1, 2, 2, 2, 5, 2, 0, 3, 4, 1, 3, 1, 3, 3, 3, 4, 1, 3, 2, 3, 2, 3, 2, 2, 2, 2, 2, 1, 2, 2, 2, 3, 3, 2, 0, 2, 1, 2, 2, 2, 4, 2, 3, 5, 4, 4, 3, 2, 2, 2, 1, 3, 2, 2, 2, 4, 3, 2, 1, 2, 2, 2, 1, 4, 4, 2, 1, 2, 2, 2, 1, 2, 4, 3, 2, 2, 3, 3, 4, 0, 3, 2, 2, 0, 2, 2, 2, 3, 2, 1, 1, 2, 1, 2, 3, 2, 2, 1, 4, 1, 3, 2, 3, 2, 1, 2, 3, 3, 2, 3, 3, 1, 4, 1, 1, 3, 2, 4, 4, 2, 1, 1, 2, 2, 3, 1, 1, 2, 2, 3, 2, 2, 1, 2, 2, 2, 1, 2, 2, 1, 1, 2, 2, 3, 2, 2, 2, 4, 1, 3, 3, 2, 2, 4, 1, 3, 3, 1, 3, 1, 3, 2, 2, 1, 2, 1, 1, 2, 2, 5, 3, 1, 3, 3, 1, 3, 1, 3, 1, 3, 2, 2, 2, 1, 4, 2, 2, 6, 2, 3, 2, 2, 1, 3, 1, 3, 3, 2, 3, 1, 2, 2, 4, 1, 1, 1, 2, 2, 2, 1, 3, 2]

Figure 2.3 (Cech et al. 2017, 16)

The motifs and their frequency are then identified; for the first five:

- 4
- 3-3-4
- 2-3
- 2-2-2
- 1-4
- Etc.

Figure 2.4 (Cech et al. 2017, 16)

Again, the study shows that this type of motifs can be considered as a unit sharing the same distributional patterns as well-established traditional units, including the rank frequency and spectrum abide by the Zipf-Mandelbrot distribution. The authors state that **valency motifs** can be considered regular language entities.

Interestingly Köhler bases his approach and the design of “his” *motif* on the so called **F-motif**, forged by the musicologist Boroda in the 1970s and 80s (Boroda 1973; 1982). Boroda was looking for a unit in music that could be seen as correspondent to *word* in language and finally came up with the *F-motif*, which is, just as Köhler’s linguistic *motifs*, a sequence of equal or increasing values, only that in the case of Boroda it is tones and not syllables. Boroda also

already tested the ditributional patterns of these *motifs* against the Zipf-Mandelbrodt-Law, concluding that they match nearly perfectly into the projected distribution by the Zipf-Mandelbrodt-Law. Köhler reimported this concept to linguistics. Boroda’s motif does not coincide with the classic notion of *motif* in musicology, that of a recurring melody etc. (see above).

This conception of *motif* as a unit of analysis of texts is still limited to the formal level, but it is undoubtedly possible, as the study on argumentative relations shows, to extend it to more semantic considerations. Let us conclude with Köhler (2015, 107) that

motifs provide a means to analyse text with respect to their sequential structure with respect to all kinds of linguistic units and properties; even categorical properties can be studied in this way. The granularity of an investigation can be adjusted by iterative application of motif-formation, and proven statistical methods can be used for the evaluation. The full potential of this approach has not yet been explored.

This method of automatic text processing, proposed by Köhler and his colleagues, is profoundly different from the “bag-of-words” model, which is generally applied in linguistics and stylistics, and does not take into account the sequentiality of the units composing a text, nor their topology.

5. THE CONTRIBUTION OF COMPUTER LINGUISTICS IN THE FRENCH DOMAIN

In recent years several approaches in France, halfway between computer processing and text linguistics, have developed, first independently and then partly in collaboration. The main objective was to create methods and tools to identify, what in these approaches is also called *motifs*. In the following we propose an overview over these perspectives, who all have in common, that they consider a *motif* a statistically significant unit or a unit composed of statistical significant linguistic representations of it.

5.1.GANASCIA (2001)

One of the first works on textual *motifs* in France is the approach of Ganascia (2001). In this case it is opportunde to speak of textual *motifs* because it is indeed a question of characterizing speeches. But, these characterizations are of a syntactic nature, in other words, texts are characterized by the syntactic specificities of the sentences that compose them.

Ganascia automatically extracts recurrent sequences from syntactic trees. Applied to texts, the implemented method makes it possible to identify the stylistic characteristics of these texts. More recently, Ganascia and his team at Sorbonne University have been working on a simpler model, which consists of identifying motifs of POS in texts by factor analysis. The text is first segmented into a set of sentences and then each sentence is mapped into a sequence of syntactic (PoS tag) items. For example the sentence⁴:

“J’aime ma maison où j’ai grandi.”

(I love my house where I grew up)

is first mapped onto a sequence of PoS tags,

<PRO:PER VER:pres DET:POS NOM PRO:REL PRO:PER VER:pres VER:pper SENT>

⁴ From Frontini and al. (2018 :122).

In Frontini and al. (2018), the authors characterize the style of the dialogues of the characters in Molière’s plays. Thus, *Scapin* and *Harpagon*, the two comical characters, are both characterised by patterns of lower syntactic complexity. This is especially the case with Harpagon, whose patterns convey the image of a self-centered person, who wants to have things his way, and who is subject to violent disappointments (especially when money is involved):

- Example 4 - [PRO:PER] [PRO:PER] [VER:pres] [VER:pper]
- on m' a privé ... [they have deprived me ...]
 - on m' a dérobé ... [they have robbed me ...]
 - on m' a volée ... [they have stolen me ...]
 - on m' a pris ... [they have taken me ...]

The technique is partly implemented in the Eremos online platform (<<http://eremos.lip6.fr/index.php>>). Eremos is a web application capable of extracting and manipulating syntactic patterns through a simple, fast and ergonomic user interface.

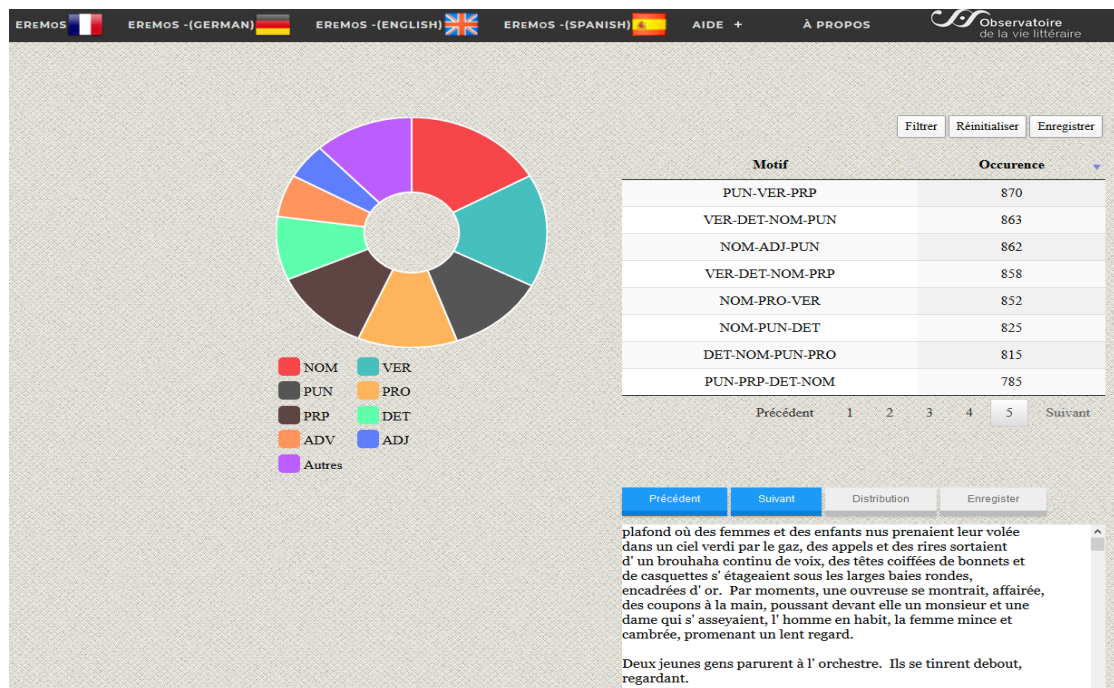


Fig.2.5: Eremos

Another platform dedicated to the identification of syntactic patterns is also worth mentioning: SDMC (<https://tal.lipn.univ-paris13.fr/sdmc/>; see

Saharoui et al. 2017). From a text or a corpus, SDMC extracts regularities, that is, sequences of words, lemmas and POS tags that frequently appear together. A set of constraints can be applied to filter the extracted patterns: minimum frequency, or text-specific constraints like “contains an adjective, or a verb”. SDMC facilitates the analysis of these **sequential motifs** with a visualization:

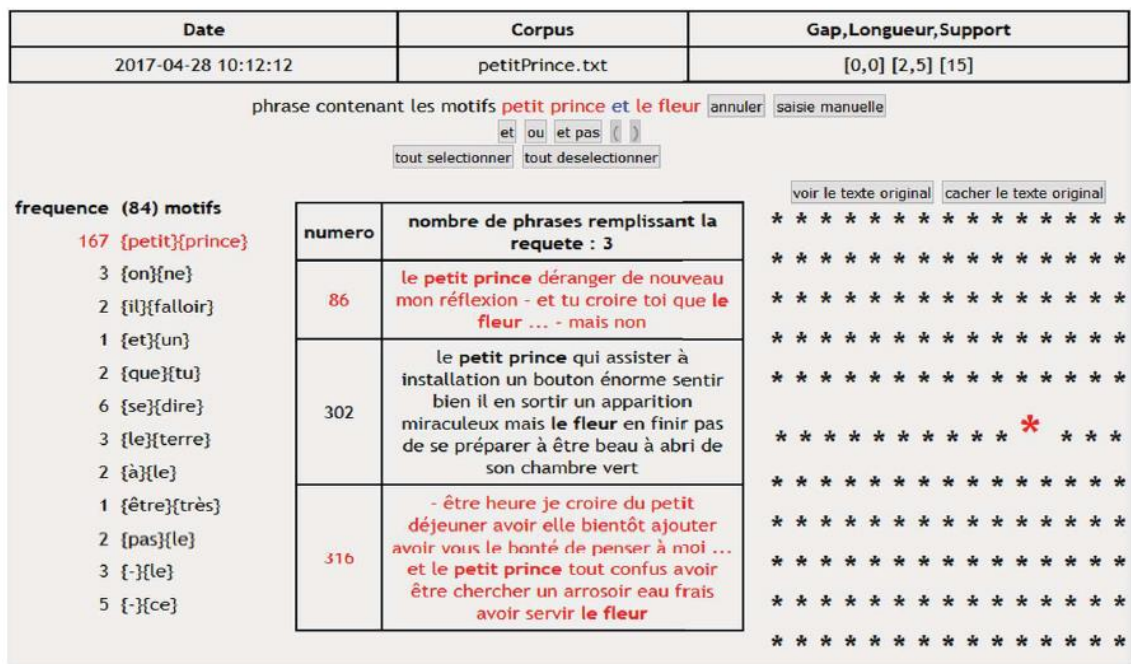


Fig. 2.6 : SDMC

It is possible to search not only for one *motif* but also cooccurrences of *motifs*. On the left side of the visualization, there is the index of *motifs*; in the middle column, we find the sentences containing the *motif*. And in the right part, the tool proposes a compact version of the text, wherein the sentences are replaced by stars. SDMC therefore makes it possible to analyze the topological dimension of a pattern.

5.2 THE APPROACH OF LONGRÉE AND MELLET

In a different context, that of textometry of Latin texts, Longrée and Mellet (2018, 156) propose to consider *motif* as a textual unit:

What is a motif? In a formal way the motif is defined as an ordered subset of the textual ensemble, formed by the recurring combination of n elements provided with its linear structure. Thus, if the text is formed by a certain number of occurrences of elements A, B, C, D and E, a motif can be the recurring micro-structure ACD or AAA, etc., without here prejudging the nature (lexical, grammatical, metrical, ...) of the elements A, B, C, D and E in question: the ‘motif’ is only the framework – or the collocational pattern – accommodating a range of parameters to be defined and capable of characterizing the diverse texts of a corpus, or even the different parts of a text. The motif is involved in the temporal dynamics of the narration, in the relations between sentences and between the different textual sequences such as descriptions, narrations, argumentations and so on. As general pattern, the motif is able to characterize a genre; but its different realizations or tokens may be specific to different authors in a given genre.

The objective is clearly the topological characterization of texts, A text is understood as an ordered and multidimensional space. The aim is to highlight the units that contribute to the

organisation of this space. The authors' definition emphasizes the non-linear dimension of the pattern and its variability, as the following example⁵ shows:

(5)
here I would like/wish to specify however
here, however, I would like to clarify
Ø I wish to specify, however, that
here I would like to specify Ø
here I would like to specify however

The authors' perspective currently tends towards a cognitive validation of *motifs*; it involves linking the associations between the lexical elements of a *motif* as measured by textual statistics and the priming effects of psycholinguistic analysis.

5.3. THE LST APPROACH

The notion of *motif* is also used in the work undertaken in the PhraseoRom project, co-funded by DFG and ANR, which seeks to identify lexical pivot patterns specific to six novel subgenres (crime, love, historical, science fiction, fantasy, general literature). To do this, the involved authors use a method based on corpora of **syntactic dependency trees**: the automatic extraction of recurrent lexico-syntactic trees (LST) (see Kraif, 2016). LST are lexical units linked by syntactic dependency relationships. They are constructed from statistically significant series of co-occurrences based on a statistical association measure, in this case the calculation of the **log-likelihood ratio** (LLR), using the **Keywords Method** (Bertels and Speilmann, 2013). Their extraction is done automatically thanks to the programme Lexicoscope (Kraif, 2016). LST extraction takes as its starting point a lexical pivot (verbal or nominal), selected by frequency and dispersion criteria (among the authors composing the corpus), and produces a set of recurrent subtrees.

The criteria used in the project for the extraction of LST that are specific for the novel subgenres in question, are the following:

- A frequency threshold for nominal and verbal pivots relative to the subcorpus size (greater than 5 or 10);
- An LST size of at least 3 elements, with no upper limit;
- An LLR greater than 10.83;
- A variable dispersion according to the compared corpora, which can be an up to 50% of the authors of the corpora.

The figure below gives an example of an extracted LST corresponding to the expression *il pousse la porte* (he pushes the door), specific to French crime novels compared to contemporary French general literature (Gonon et al. 2018).

⁵ Our example.

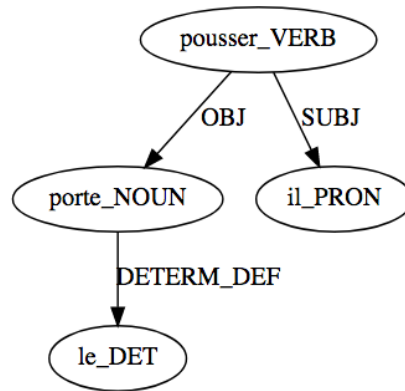


Fig. 2.7: An example of an LST

Compared to purely sequential approaches, such as n-grams, LST extraction has the advantage of associating only elements that are united by a syntactic relationship, which considerably reduces noise. On the other hand, LST make it possible to free oneself from linearity by exploiting syntactic relationships and thus limit redundancy. Thus, examples (6) and (7) correspond to a single LST in which *cœur* (heart) is the subject of the verb *battre* (to beat).

(6) **Le cœur** de Kurt **battait** à coups immenses (Kurt's heart was beating with immense blows). (Berthelot Francis, *Khanaor 1 – Solstice de fer*)

(7) **Le cœur** encore **battant**, il se rendit compte que c'était lui qui avait crié (his heart still beating, he realized that it was he who had screamed). (Achard Marion, *Tout seuls*)

These two examples also illustrate another specificity of LST: they allow gaps between the elements forming a pattern, which is not possible with a sequential approach. Another step is still possible. Indeed, since this method is based on the lexicon, as many LST as paradigmatic and syntagmatic variations of the motif can be extracted. For example, just as the LST “open the door” presented above, the LSTs “close the door”, “push the door”, “close the door”, “open the door” and “the door opens” are specific to crime fiction (compared to general literature fiction). A grouping of these LST is therefore necessary and is done manually for the moment.

We see how this conception of *motif* could be linked to more narratological perspectives, such as the one proposed by Muryn et al. (2016). These authors are working with the notion of *semantic motif* in the crime fiction:

The semantic motif is an abstract representation of all structures realizing the same predicate-argument model [...] (it) can thus take the form of a complex sentence, a simple sentence, an SN, etc., or be inferred in whole or in part (Muryn et al. 2016, 4).

It seems to us that we are close here to the complex conception of Dolezel's (1972) *motif* presented above. Thus, one of the *motifs* inherent in the crime scene, in a crime fiction novel, is the discovery of a body. This narrative moment (*motifeme*) is expressed by various possibilities (*texture motifs*), but can be abstracted into a set of headings:

N<HUM>	<DESC>	<POSITION>	<MANNER>	<LOC>
The corpse of a little girl	disarticulated	huddled up	∅	In the middle of a forest
A man	naked	lying	Unconscious	In a house in the 13th arrondissement

Fig. 2.8.

The aim of Muryn et al. (2016) is thus to create a semantic grammar of a particular literary genre.

5.4. MOTIFS AND CONSTRUCTIONS

Without using the term *motif*, some approaches, based on n-grams or **lexical bundles** (for example Mahlberg 2013) aim to characterize, for example, the way in which characters in novels are constructed. This is no longer a matter of narratological *motifs* in the sense of minimal concrete themes, but this type of research is fundamental to understanding writing processes and to analyzing the specificity of textual genres. The following example is from Siepmann (2015, 381), who is interested in units that he considers as lexically specific narrative patterns and that he calls **key word patterns** or **syngram**:

(8) PossDet + *thoughts were on* / of NP / V+ing

It is a pattern used to introduce a passage which depicts a character's train of thought:

Like the categories of discourse markers known as 'announcers' or 'topic initiators' (Siepmann 2005:93), which prepare the reader for the next step in the author's exposition or reasoning (*I will now briefly describe*) or initiate a new topic in the current text, this kind of unit frequently, though not systematically, serves to indicate a change in focalization, most typically a shift from external to internal focalization. It is notably common for this marker to be followed by a proper noun⁶ (Siepmann, 2015, 381).

The work carried out by Legallois (see Legallois et al. 2018)) attempts to identify these types of patterns, but also formally more abstract patterns. We can thus speak of **grammatical ally motifs**, particularly within the method first presented by Quiniou et al. (2012), by which the authors sought to identify recurring lexico-grammatical patterns in a corpus of 19th century poetry. Since then, the method has been developed and has been applied to different corpora, in particular to distinguish textual genres: the distinction between travel stories and fiction by the same authors (see Diwersy and Legallois (forthcoming), the characteristics of the novels

⁶ And, parting from them abruptly, he ran down the stairs and out into the air. He bolted into a hansom, and drove to the Goat's Club. His thoughts were on Holly and what he must do before her brother showed her this thing in tomorrow's paper (J. Galsworthy, *In Chancery*).

Harlequin (see Legallois et al. 2016) or the stylistic *motifs* in 19th century French novelists (see Legallois et al. 2018).

The method is based on a hybrid annotation of corpora: the set of labels corresponds to general morpho-syntactic categories for nouns, adjectives and for most verbs and adverbs. The most common verbs (auxiliaries, modals et cetera) are reduced to their lemma. The form of invariable words (prepositions, conjunctions, certain adverbs et cetera) is preserved. For example:

- (9) La lecture a ceci de magique qu'elle permet de remonter le temps (Thilliez, L'anneau de moebius)
 (Reading is magical in that it allows us to go back in time)

is annotated this way :

- (10) le N avoir ceci de ADJ que SUB il permettre de INF le N.

Then, by extracting repeated segments whose length can be varied, we obtain a set of patterns. These patterns are therefore sequential. Finally, the specificity of each pattern for a given text is statistically calculated in relation to the other texts in the corpus – the idea being to characterize these texts by units of a dimension other than lexical ones. In order to give an idea of the type of unity that can be captured in this way, we recur to a comparison between two different genres: the *crime*- and the *love*-genre. The relatively small corpus includes:

- 12 “detective” novels, written by Thilliez, Benacquista, Vargas et cetera (2 texts per author), 479129 words
- 12 novels from the *Harlequin* collection, 336902 words.

We are only interested here in a few *motifs* related to the *thriller*. These *motifs* have for us a construction value, in the sense of **Construction Grammar** (for example Croft and Cruse 2004; Goldberg 2006). A construction is a linguistic unit (a sign), which can be a morpheme, a word, the **argumental schema** of a verb, or a **phraseological unit**. Constructions can be generic (such as the transitive construction “subj V Obj”), semi specific (“subj *break* obj”) or specific (*give the Devil his due*).

A construction is associated with a meaning that can be semantic, pragmatic or functional. Hence this representation of a construction:

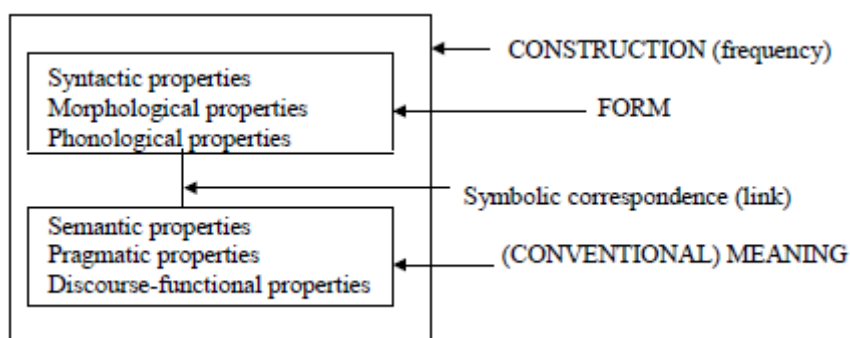


Fig 2.9: The symbolic structure of a construction (Croft 2001:18)

The theoretical framework of Construction Grammar is an important contribution to understanding the role of *motif* in texts. Here, the *motif* is a construction – a linguistic sign – considered not only from the grammatical point of view, but from the point of view of its discursive dimension in a particular text or a particular genre. Nevertheless, these *motifs*, like the constructions studied in Construction Grammar, contribute to the linguistic competence of “speakers” (here writers and readers), a competence understood as knowledge not of forms as such, but of their use.

Given the small size of the corpus, the number of occurrences of these *motifs* is itself naturally modest. Nevertheless, their distinctive dimension is obvious. This is how we remember:

(11) Motif [Inf, inf, inf...] (8 occurrences)

(a)- Mettre le feu dans le passé, tout perturber, créer des failles, des incohérences, et, peut-être, permettre l'arrestation de l'assassin, avant qu'il s'en prenne à Sylvie (Set fire to the past, disrupt everything, create loopholes, inconsistencies, and perhaps allow the murderer to be arrested, before he attacks Sylvie). (Pol_thilliez_anneau; translation DL)

(b)- Foncer à Méry-sur-Oise, faire le tour des établissements scolaires, comprendre quel rôle elle jouait dans cet incroyable micmac (Go to Méry-sur-Oise, visit all the schools, understand what role she played in this incredible Mi'kmaq). (Pol_thilliez_anneau; translation DL)

(c)- Run, turn around, find the network and alert the reinforcements. (Pol_thilliez_chambre).

The *motif* present, essentially in Thilliez's work, is an accumulation of infinitives referring to important trials in the narrative. The content of these averbial sentences is anchored in the character's inner monologue, which urgently produces a reflection to determine a course of action.

(12) Motif [N, N, N, ...] (7 occ.)

(a)- Catacombes, laboratoires de biologie, nécropoles, hôtels bizarroïdes, et j'en fais ensuite une espèce de grosse soupe, dans mon cerveau (Catacombs, biology laboratories, necropolises, weird hotels, and then I make a kind of big soup out of it, in my brain) (Pol_thilliez_chambre ; translation DL).

(b)- Depuis l'automne, mélancolie, mutisme, brusquerie, insomnies et sautes d'humeur (Since autumn, melancholy, silence, abruptness, insomnia and mood swings) (Pol_vargas_seine ; translation DL).

This time it is an accumulation of nouns (concrete in the first example, abstract in the second), that summarizes past dysphoric experiences. The *motif* also gives us access to the consciousness of the character who is in full mental confusion.

(13)- Motif [pas de N] (9 occ.)

(a)- En pleine nuit, pas de lumières, il n'avait aucune chance d'être vu (In the middle of the night, no lights, he had no chance of being seen) (Pol_thilliez_chambre ; translation DL)

(b)- A priori, pas de traquenard (A priori, no stumbling blocks). (Pol_benacquista_malavita ; translation DL)

(c)- Il n'y a personne, ses volets sont baissés, pas de voiture (There's no one there, his shutters are down, no car). (Pol_thilliez_chambre; translation DL)

The syntactic reduction constructs an obvious effect: the expression is anchored in the inner discourse of the character observing a situation and evaluating it to determine if there is danger.

(14) Motif [Q alors que P ?] (6 occurrences)

(a) Que faisait-il au domaine accompagné de Victor, à boire des canettes, alors que toutes les polices de France le recherchaient? (What was he doing at the property with Victor, drinking cans, when all the police in France were looking for him?) (Pol_benacquista_ego; translation DL)

(b)- Pouvait-on imaginer qu'il épargnerait des témoins, alors qu'il avait étranglé une fillette innocente? (Could we have imagined that he would spare witnesses when he had strangled an innocent girl?) (Pol_thilliez_chambre; translation DL))

This pattern expresses several values: the illogicality of a situation (drinking cans – Q – when one is wanted – P –, where P, normally, prohibits Q), or, on the contrary, its predictability (how could Q not happen in the future when P has happened in the past and yet P is less likely than Q?; P “allows” therefore Q). In a different form, because it is more absurd (killing animals is less likely here than killing people), Vargas exploits this *motif* whose form is no longer written *Q whereas P, but P then why Q?*.

(15) Je tue pas les bêtes, alors pourquoi je tuerais les gens? (I don't kill animals, so why would I kill people?). (Pol_vargas_Neptune; translation DL).

These few examples show that the method makes it possible to identify patterns specific to a textual genre whose discursive function or functions can be described. We believe that these patterns are constructions insofar as they correspond to a pairing between (variable but identifiable) forms and meanings. Like the grammatical constructions, the *motifs* can be schematic, semi specific or even specific in the case of phraseologisms.

CONCLUSION

This article proposes a kind of map, very incomplete because many territories are still missing or “uncharted”, to the notion of *motif*. Because of the polysemy of the word, because of the different conceptions it embodies, it is obviously impossible – and indeed undesirable – to consider that there exists an object designated by the word *motif*, which is common to all the

disciplines reviewed here. *Motif* is a “crossroads” word, based on which disciplines can and should interact to share concepts, methods and tools. Beyond its function of presenting a heterogeneous notion of a term (*motif*), this article pleads for a collaboration between perspectives, which would make it possible to refine certain definitions, to point out phenomena ignored in certain approaches, in short, to give an account, each in its own specificities, of elements in common that motivate the recurrent use of the word *motif*. Thus, we started from narratology and folklore, which conceive the *motif* as a minimal unit, first independent of a particular form (the *motif* is a kind of concrete theme), the *motif* has acquired a “morphology” in the Russian structuralist tradition: it is linked to generic proposals, which are embodied in various statements, but which are associated with defining functions of the story. The “import” of the *motif* into modern literary criticism is certainly fructuous, but not without problems: narrative functions are in principle less stereotypical. However, at the level of the analysis of a literary genre, as marked as the crime or sentimental novel, we can detect characteristic patterns, whether these patterns are recurrent lexico-grammatic patterns (similar, in a way, to formal scientific patterns) associated with a particular semantics, or abstractions giving rise to concrete achievements. Subsequently, it will be appropriate for the linguistic analysis of literary texts, and for literary criticism itself, to produce a higher level analysis, namely the study of the articulation of these minimum units into more global units, to assess the thematic developments of the texts. A substantial work in perspective!

Aarne, Antti. 1961 [1910]. *The Types of the Folktale: A Classification and Bibliography*. Helsinki: The Finnish Academy of Science and Letters.

Boroda, Mojsej. 1973. “K voprosu o metroritmitsheski elementarnoj edinice v muzyke”. *Soboshtshenija AN GSSR* 71, No. 3: 745–202.

Boroda, Mojsej. 1982 [1975]. “Die melodische Elementareinheit.” In *Sprache, text, Kunst. Quantitative Analysen*, edited by Jurij Orlov, Mojsej Boraoda, and Isabella Nadarejshvili, 205–221. Bochum: Brockmeyer.

Christensen D.L. 2003. *The Unity of the Bible: Exploring the Beauty and Structure of the Bible*. Mahwah, New Jersey: Paulist Press.

Courtès, Joseph. 1982. “Motif et type dans la tradition folklorique: Problèmes de typologie”. *Littérature* 45: 114–127.

Croft, William. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press

Croft, William, and Alan Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Ganascia, Jean-Gabriel. 2001. “Extraction automatique de motifs syntaxiques” In *Actes de la 8ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN’2001)*. Tours (France), edited by Jean Véronis, Laurence Danlos, Pierre Zweigenbaum, Nathalie Gasiglia, and Pascal Amsili (Bureau et Conseil d’Administration de la ATAL), TALN’2001, Jean-Gabriel Gnascia, [1–10]. Paris: ATAL – Association pour le Traitement Automatique des

Langues. Accessed January 28, 2019. <http://talnarchives.atala.org/TALN/TALN-2001/taln-2001-long-017.pdf>.

Goldberg, Adele. 2006. *Constructions at Work*. Oxford: Oxford University Press.

Ducrot, Olivier, and Tzvetan Todorov. 1972. *Dictionnaire encyclopédique des sciences du langage*. Paris: Éditions du Seuil.

Daranyi, Sándor, Peter Wittek, and Lázló Forró. 2012. "Toward Sequencing "Narrative DNA": Tale Types, Motif Strings and Memetic Pathways." In *Proceedings of Computational Models of Narrative 2012*, edited by Mark Alan Finlayson, 2–10. Cambridge, Mass.: Massachusetts Institute of Technology. Accessed January 28, 2019. <http://narrative.csail.mit.edu/cmn12/proceedings.pdf>.

D'Haeselaar, Patrick. 2006. "What are DNA sequence motifs?" *Nature Biotechnology* 24, fasc. 4: 423–425. Accessed January 28, 2019. <http://146.6.100.192/users/CH391L/Handouts/nbt0406-423-primer-whataremotifs.pdf>.

Doležel, Lubomír. 1972. "From Motifemes to Motifs." *Poetics* 4: 55–90.

Dundes, Alan. 1962. "From Etic to Emic Units in the Structural Study of Folktales." *Journal of American Folklore* 75, No. 296: 95–105.

Frontini, Francesca, Mohamed Amine Boukhaled, and Jean-Gabriel Ganascia. 2018. "Approaching French Theatrical Characters by Syntactical Analysis: a Study with Motifs and Correspondence Analysis." In *Grammar of Genres and Styles. From Discrete to Non-Discrete Units*, edited by Dominique Legallois, Thierry Charnois, and Meri Larjavaara, 118–139. Berlin/Boston: de Gruyter.

Sahraoui, Hedi-Théo, Pierre Holat, Peggy Cellier, Thierry Charnois, and Sébastien Ferré. 2017. "Exploration of textual Sequential Patterns". In *14th International Conference on Formal Concept Analysis. Rennes, France, June 13–16, 2017. Supplementary Proceedings*, edited by Karell Bertet, Daniel Borchmann, Peggy Cellier, and Sébastien Ferré, 99–102. Rennes: IRISA. Accessed January 28, 2019. https://icfca2017.irisa.fr/files/2017/06/icfca_suppl_final_avec_isbn.pdf.

Muryn, Teresa, Malgorzata Niziolek, Alicja Hajok, Wojciech Prazuch, and Katarzyna Gabrysiak. 2016. "Scène de crime dans le roman policier: essai d'analyse lexico-syntaxique." In *5^e Congrès Mondial de Linguistique Française (SHS Web of Conferences 27)*, edited by F. Neveu, G. Bergounioux, M.-H. Côté, J.-M. Fournier, L. Hriba, and S. Prévost, 06007: 1–14. Les Ulis Cedex A/London: EDP Sciences. Accessed January 28, 2019. https://www.shs-conferences.org/articles/shsconf/pdf/2016/05/shsconf_cmlf2016_06007.pdf.

Köhler, Reinhard. 2006. "The frequency distribution of the lengths of length sequences". In *Favete linguis. Studies in honour of Victor Krupa*, edited by J. Genzor, and M. Bucková, 145–152. Bratislava: Slovak Academic Press.

Köhler, Reinhard. 2015. "Linguistic Motifs." In *Sequences in Language and Text*, edited by George K. Mikros and Ján Mačutek, 89–108. Berlin/Boston: de Gruyter.

Legallois, Dominique, Thierry Charnois, and Thierry Poibeau. 2016. "Repérer les clichés dans les romans sentimentaux grâce à la méthode des "motifs"." *Lidil. Revue de linguistique et de didactique des langues* 53: 95–117.

Legallois Dominique, Thierry Charnois, and Meri Larjavaara. 2018. "The balance between quantitative and qualitative literary stylistics: how the method of "motifs" can help". In *The Grammar of Genres and Styles: from Discrete to Non-Discrete Units*, edited by Dominique Legallois, Thierry Charnois, and Meri Larjavaara, 164–193. Berlin/Boston: de Gruyter.

Kraif, Olivier. 2016. "Le lexicoscope: un outil d'extraction des séquences phraséologiques basé sur des corpus arborés." *Cahiers de Lexicologie* 108: 91–106.

Mandelbrot, Benoît. 1953. "An Information Theory of the Statistical Structure of Language." In *Communication theory*, edited by W Jackson, 503–512. New York: Academic Press.

Mahlberg, Michaela. 2012. *Corpus stylistics and Dickens's Fiction*. London: Routledge

H-T Sahraoui, P. Holat, P. Cellier, Th. Charnois and S. Ferré (2017).

Harris, Zellig, M Gottfried, Th. Ryckman, A. Daladier, and P. Mattick. 1989. *The form of Information in Science: Analysis of an Immunology Sublanguage*. Dordrecht : Kluwer Academic Publishers

Longrée, Dominique, and Sylvie Mellet. 2018. "Towards a topological grammar of genres and styles: a way to combine paradigmatic quantitative analysis with a syntagmatic approach". In *The Grammar of Genres and Styles: From Discrete to Non-Discrete Units*, edited by Dominique Legallois, Thierry Charnois, and Meri Larjavaara, 140–163. Berlin/Boston: de Gruyter.

Ofek, Nir, Sándor Darányi, and Lior Rokach. 2013. "Linking Motif Sequences with Tale Types by Machine Learning." In *Workshop on Computational Models of Narrative 2013*, edited by Mark A. Finlayson, Bernhard Fisseni, Benedikt Löwe, and Jan Christoph Meister, 166–182. Wadern: Dagstuhl Publishing. Accessed January 28, 2019, <http://drops.dagstuhl.de/opus/volltexte/2013/4150/pdf/p166-ofek.pdf>.

Richard, Jean-Pierre. 1979. *Microlectures*. Paris: Seuil.

Siepmann, Dirk. 2016. "Lexicologie et phraséologie du roman contemporain : quelques pistes pour le français et l'anglais." *Cahiers de Lexicologie* 108 : 21–41.

Siepmann. 2015. "A Corpus-Based Investigation into Key Words and Key Patterns in Post-War Fiction." *Functions of Language* 22, fasc. 3: 362–399.

Siepmann, Dirk. 2005. *Discourse Markers across Languages*. London: Routledge.

Tomashevskij, Boris. 1966. *Theory of Literature/Theorie de la Litterature?*

Thompson, Stith. 1946. *The Folktale*. New York: The Dryden Press.

Thompson, Stith. 1932–1936. *Motif-index of folk-literature: a classification of narrative elements in folktales, ballads, myths, fables, mediaeval romances, exempla, fabliaux, jest-books, and local legends*. Six vol. Helsinki: Suomalainen Tiedeakatemia.

Čech, Radek, Veronika Vincze, and Gabriel Altmann. 2017. “On motifs and verb valency.” In *Motifs in Language and Text*, edited by Haitao Liu, and Junying Liang, 13–36. Berlin/Boston: de Gruyter.

Pike, Kenneth. 1967. *Language in Relation to a Unified Theory of the Structure of Human Behaviour*. Second, Revised Edition. The Hague/Paris: Mouton.

Prince, Gerald. 2003. *A Dictionary of Narratology*. Revised Edition. Lincoln: University of Nebraska Press.

Propp, Vladimir. 1968 [1928]. *Morphology of the Folktale*. Second Edition. Austin: University of Texas Press.

Abbott, H. Porter. 2008. *The Cambridge Introduction to Narrative*. Second Edition. Cambridge: Cambridge University Press.

Quiniou, Solen, Peggy Cellier, Thierry Charnois, and Dominique Legallois. 2012. “What about Sequential Data Mining Techniques to Identify Linguistic Patterns for Stylistics?” In *Computational Linguistics and Intelligent Text Processing*. Proceedings, Part I, edited by Alexander Gelbukh, 166–177. Berlin/Heidelberg: Springer.

Veselóvskij, Aleksandr. 1913a. “Poetika suzhetov i yeya zadatshi.” In *Poetika*. Tom Vtaroj (Vol. 2). *Poetika suzhetov (1897–1906)*, edited by The Imperial Academy of Sciences, 1–8. St. Petersburg: Tipografija Imperatorskoj Akademii Nauk.

Veselóvskij, Aleksandr. 1913b. “Motiv i suzhet.” In *Poetika*. Tom Vtaroj (Vol. 2). *Poetika suzhetov (1897–1906)*, edited by The Imperial Academy of Sciences, 9–12. St. Petersburg: Tipografija Imperatorskoj Akademii Nauk.

Vincensini, Jean-Jacques. 2000. *Motifs et thèmes du récit médiéval*. Paris: Nathan.

Zipf, George Kingsley. 1935. *The Psycho-Biology of Language*. Boston. Houghton Mifflin Company.