



# From binary collocations to grammatically extended collocations: Some insights in the semantic field of emotions in French.

Agnès Tutin, Olivier Kraif

## ► To cite this version:

Agnès Tutin, Olivier Kraif. From binary collocations to grammatically extended collocations: Some insights in the semantic field of emotions in French.. Mémoires de la Société néophilologique de Helsinki, 2016, Collocations Cross-Linguistically. Corpora, Dictionaries and Language Teaching,, pp.245-266. hal-01337486

**HAL Id: hal-01337486**

**<https://hal.science/hal-01337486>**

Submitted on 15 Dec 2018

**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.

## **Preliminary version**

---

# **From binary collocations to grammatically extended collocations: some insights in the semantic field of emotions in French**

AGNÈS TUTIN

LIDILEM, Université Grenoble-Alpes

BP 25, 38040 Grenoble Cedex 09

agnes.tutin@univ-grenoble-alpes.fr

OLIVIER KRAIF

LIDILEM, Université Grenoble-Alpes

BP 25, 38040 Grenoble Cedex 09

olivier.kraif@univ-grenoble-alpes.fr

Collocations have long been recorded by lexicographers but the theorization of this kind of phraseological unit is quite recent. In line with the lexicological “continental” tradition of collocation, we propose a set of semantic and syntactic criteria to delimit clearly this phenomenon. We then show that, in the semantic field of emotions, collocations are more regular than expected because a) they are especially productive in specific constructions; b) few semantic relations, as shown by ECD Lexical Functions, are very productive; c) many collocates are combined with large paradigms of bases. We also highlight the importance of grammatical elements in collocations, which are essential but have failed to receive the attention they require. We demonstrate that corpus-based linguistic studies can offer interesting insights for this issue, through NLP tools such as the Lexicoscope.

## **1. Introduction**

Collocations have been a central topic for several years now, even though the definition of this linguistic phenomenon is not always clear. In this paper, we aim to show that, while collocations are often unpredictable, most of them display regularities, especially in the lexical field of emotions. We also wish to highlight the importance of grammatical elements in collocations, which are essential and yet have failed to receive the attention they require. We will also demonstrate that corpus-based linguistic studies can offer interesting insights for this issue, through NLP tools such as the Lexicoscope (Kraif & Diwersy, 2014).

After outlining some of the historical milestones related to this notion, we will provide a set of linguistic criteria that can be used to identify collocations, based on syntactic and semantic

properties. We will then focus mainly on the semantic field of emotion and show how most collocations exhibit regular properties. Finally, we will discuss our methodology, based on the Lexicoscope, and give detailed information about the syntactic properties that should be considered when providing a usage-based description of collocations.

## 2. A definition of collocation

### 2.1 Some historical milestones

Although the term “collocation” itself is new, this kind of expression has long been recorded by lexicographers. According to F.J. Hausmann (1979), the oldest French monolingual dictionary is a dictionary of “collocations”, *Les épithètes de M. de La Porte*, published in 1571, which included 80 000 noun-adjective associations, such as those shown in Figure 1. For example, for the noun *tristesse* [‘sadness’] attributive adjectives such as *languide* [‘languid’], *amère* [‘bitter’] and *mortelle* [‘mortal’] are mentioned.

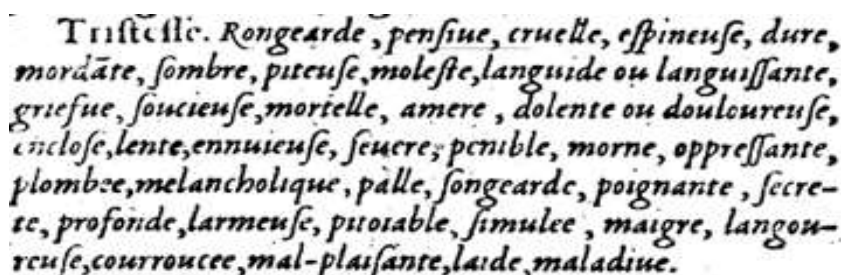


Figure 1 : An excerpt from *Les Epithètes de M. de la Porte*, 1571

In French-speaking linguistics, Charles Bally, one of Saussure’s disciples, proposed a fine-grained typology of multiword expressions in his 1909 *Traité de stylistique*, in which the notion of collocation was already present, but referred to as a *série phraséologique* [‘phraseological series’] or *groupement usuel* [‘usual word combination’], different from “indecomposable units”.

Entre les extrêmes (groupements passagers et unités indécomposables) se placent des groupes intermédiaires appelés **séries phraséologiques** (p.ex. les séries d’intensité ou les périphrases verbales). (1909: 66) [Between the extremes (free word combinations and indecomposable units) lie intermediary groups called **phraseological series** (e.g. series of intensity markers or verbal periphrases)]

Il y a série ou **groupement usuel** lorsque les éléments du groupe conservent leur autonomie, tout en laissant voir une affinité évidente qui les rapproche, de sorte que l’ensemble présente des contours arrêtés et donne l’impression du “ déjà vu ”. (1909: 70) [There is a series or **usual word combination** when elements of the group retain their autonomy, while also revealing a clear attraction drawing them together, in such a way that the whole presents a clear-cut outline and gives an impression of déjà vu]

The first mention of the term “collocation” is much more recent, in 1933, and is attributed to Palmer, a teacher of English as a Foreign Language, who defined it as “a succession of two or more words that must be learnt as an integral whole and not pieced together from its component parts” (Palmer, 1933). This notion remains essential in the field of foreign language teaching and learning (e.g. Meunier & Granger, 2008).

According to G. Williams (2003), in contemporary linguistics, there are two key trends regarding the notion of collocation. On the one hand, the British tradition, following Sinclair (1991) for example, mainly considers collocations as recurrent statistical elements. On the other hand, the “continental tradition” of lexicography and lexicology mainly defines a collocation as a privileged recurrent binary association between meaningful words, which have a syntactic and semantic relation. We mainly endorse the latter approach, whose most eminent representatives are Hausmann (1979) and Mel’čuk (1998), but in our model we do also include usage-based parameters as well some descriptive elements inspired by cognitive linguistics.

## 2.2 Criteria for a definition

Numerous studies have been conducted on collocations, for example in lexicology (e.g. Heid, 1994; Fontenelle, 1997; Bartsch, 2004; Siepmann, 2005), computational linguistics (Evert, 2008; Seretan, 2011; Kraif & Diwersy, 2014) and on pedagogical issues (Lewis, 2000; Nesselhauf, 2005; Cavalla, 2008) but we will not go into a detailed overview here.

Following previous work (mainly Tutin & Grossmann, 2002; Grossmann & Tutin, 2003; Tutin, 2010; Tutin, 2013), we propose a set of criteria for defining collocations:

1. Collocations are **associations** which typically involve **two semantically full linguistic elements**. We thus exclude associations between a full element and a grammatical word, such as *give* and *to* in *he gave a book to Laura*. But we include collocations such as *out of love* in *she committed the crime out of love*, where *out of* has a causal explanation.
2. Most collocations are **binary collocations** and most ternary collocations such as *pay close attention* can be analyzed as merging binary collocations such as *pay attention* + *close attention* or as recursive collocations such as *mettre (en colère)* where *en colère* is both a collocation and the base associated with *mettre*.
3. Collocations are involved in specific **syntactic relations** such as subject-verb, verb-complement, noun-adjective, adverb-adjective, etc. which can be represented with the help of

dependency relations such as in Tesnière's model (1959). Collocations are **endocentric** multi-word expressions which generally follow usual syntactic rules. Some collocations may sometimes have specific syntactic constructions such as Adj Adj constructions (e.g. *ivre mort* lit. 'drunk dead') or N N (*amour passion* lit. 'love passion'). They also may involve **specific constructions** associated with specific meanings, such as similes (e.g. Adj *comme* N, *heureux comme un roi*, lit. 'happy like a king') which generally have an intensive meaning, or color adjectives with the pattern Adj N (e.g. *bleu roi*, *vert pomme*, *gris souris*, lit. 'king blue', 'apple green', 'mouse gray').

4. As regards semantics, collocations have the following properties:
  - a. They are fully or largely **compositional**: one element of the collocation, the **base**, retains its usual meaning. The other part, the **collocate**, can be a) **regular** (it applies to a large set of elements, e.g. *grand* in *grand travailleur*, lit. 'big worker'), b) **transparent**, if it is easy to decode, but unpredictable (e.g. *faim de loup*, lit. 'wolf's hunger' but *appétit d'ogre* lit. 'ogre's appetite'), or c) **opaque**, when it is not easily predictable (e.g. *peur bleue* lit. 'blue fear' for 'intense fear') (for a semantic typology of collocations, see Tutin & Grossmann, 2002; Grossmann & Tutin, 2003). Regular collocations seem to be by far the most numerous.
  - b. Collocations can generally be analyzed as **predicate-argument relations** (Tutin, 2013). The predictable element, the base, is the argument of the relation, whereas the unpredictable element, the collocate, is a predicate. In *peur bleue*, the collocate *bleue* intensifies *peur*; in *drive crazy*, *drive* is the causative predicate of the argument *crazy*.
  - c. As regards **usage**, the constituent elements of collocations are **frequent combinations** which have a **strong attraction**, which is often computed with the help of **association measures** (see Evert, 2008, for an overview on this topic). While the meaning of the collocation is generally transparent, the collocates are quite arbitrary to encode. Non-native speakers generally understand them but may be unable to produce them. This is why collocations are a major issue in language teaching and learning as reflected in the large number of dictionaries of collocations dedicated to second language acquisition (Cf. Nesselhauf, 2005). These multiword expressions must be learnt and stored by non-native speakers.

This definition allows us both to consider syntactic variability and to account for regularities, which are far more numerous than might be expected.

### 3 Collocations in the field of emotions: regularities rather than idiosyncrasies

Scholars working on collocations and combinatorics have often been interested in the semantic field of emotion, since collocations are abundant and varied in this field, especially in the theoretical framework of the Explanatory and Combinatorial Lexicology, for example Mel'cuk & Wanner (1996), Alonso Ramos (2004), Sanroman (2004). Studies in this field are extremely interesting because they show that, unlike true idioms, few collocations are completely unpredictable and really idiosyncratic. Few are in fact unique combinations. Examples such as *peur bleue* (lit. 'blue fear') where *bleu* means 'intense' are very rare and most collocations follow regular patterns.

In fact, three kinds of regularities can be highlighted: grammatical constructions involving collocations, semantic types of collocations and base-collocate associations (Cf. Tutin 2012).

#### 3.1 Grammatical constructions and meaning

Some collocations are involved in specific constructions, associated with specific semantic patterns. This is the case for similes which are known to be used in the semantic expression of intensity where several syntactic patterns are used:

- N de N: *appétit d'ogre*, *faim de loup* [lit. 'ogre's appetite', 'wolf's hunger'].
- Adj comme N: *malin comme un singe*, *muet comme une carpe* [lit. 'smart as a monkey', 'silent as a carp'].
- Adj à Vinf: *triste à mourir*, *bête à manger du foin* [lit. 'sad enough to die', 'stupid enough to eat hay']
- V comme N: *dormir comme une souche*, *mentir comme un arracheur de dents* [lit. 'to sleep like a log', 'to lie like a puller of teeth'].
- V à Vinf: *il pleut à boire debout* (français québécois), *rire à s'en décrocher la mâchoire* ['it's raining enough to drink standing up' (Quebecois French), 'to laugh enough to dislocate your jaw']

The comparing element can be considered as a collocate, according to our definition above. This is a modifier and it generally has an intensive meaning, even though intensity is derived from the comparison. We can assume that the comparative construction, quite independently of the lexical items used, gives an intensive meaning. Some lexical items are obviously preferred, e.g. *comme un singe* (lit. 'as a monkey' is associated with *malin* ('smart') in French or *clear* with *crystal* in

English, but a specific meaning seems to be associated with these syntactic constructions. We thus have both a “syntactic cliché” and a “lexical cliché”, to quote Schapira’s (1999) terminology. The construction considered here is Adj à Vinf (*triste à pleurer*, lit. ‘sad enough to cry’, *ennuyeux à mourir*, lit. ‘boring enough to die’, etc.), which has been less studied than the Adj *comme* N construction. This construction is not actually a true comparative structure and could more accurately be described as a “consecutive construction” as shown by the following paraphrases:

- 1) a. Ce film est triste à pleurer. lit. ‘this film is sad enough to cry’  
 b. Ce film est tellement triste qu’on en pleurerait. Lit. ‘this film is so sad that we could cry’

In these constructions, the comparative construction is not to be interpreted literally, because it has hyperbolic meaning and works metaphorically: the meaning is not fully compositional. A small corpus study was performed on a subpart of the Frantext corpus<sup>1</sup> including novels, in order to study the semantic structure of this construction. The study was limited to predicative contexts, i.e. patterns such as *être* Adj à Vinf (e.g. *est triste à pleurer*, lit. ‘is sad enough to cry’) in order to limit the number of examples. The results of this query show that this structure is not extremely frequent, but is governed by several trends: most adjectives are axiological and negative (*triste, laid, hideux*, ‘sad’, ‘ugly’, ‘hideous’). Moreover, four main semantic fields tend to be associated with the verbs: physical manifestation (*pleurer, hurler*, ‘to cry’, ‘to scream’), death (*mourir, crever, périr*, ‘to die’, ‘to snuff it’, ‘to perish’), feeling verbs (*faire peur, faire pitié*, lit. ‘to scare [s.o.]’, ‘to generate pity’) and content verbs (*déborder*, ‘to overflow’). Some expressions are fully lexicalized (e.g. *fou à lier*, lit. ‘crazy enough to tie up’) but, very interestingly, many of them are more creative (*noir à gerber, barbare à donner le frisson*, etc., lit. ‘black enough to throw up’, ‘barbaric enough to make you shiver’). This corpus study shows that this construction has a specific meaning and produces a specific meaning, even though there are some lexical constraints with certain expressions. These “syntactic clichés” are especially frequent for collocations, alongside more standard constructions such as Subj-V or V-Obj constructions.

### 3.2 Semantic types of collocations

The ‘Lexical Functions’ of the *Explanatory and Combinatorial Lexicology* (Mel’cuk 1998) probably provide one of the most complete approaches to the semantic modeling of collocations:

---

<sup>1</sup> Prose in the categorized database from 1900 to 2007, i.e. 70 219 581 words.

standard Lexical Functions are the most recurring associations, which are universal and result in a large number of values. Bally, in his *Traité de stylistique* (1909), had already highlighted the importance of specific associations such as the *séries d'intensité* (Magn in the LF language) or the pervasiveness of *périphrases verbales* (Oper in the LF language). Collocations seem to be especially frequent with specific universal semantic relations such as intensity (Magn), cause (Cause) and the notion of set (Mult) or portion (Sing), which involve syntagmatic LFs. However, to our knowledge, the productivity of LFs had not yet been checked against corpus evidence. We wanted to examine the extent to which these trends were confirmed in the semantic field of emotions.

A linguistic study involving annotation of lexical functions in a corpus of 670 000 words was therefore carried out in the semantic field of emotions (Augustyn & Tutin, 2009). The XML annotation was performed semi-automatically and included information about the LF, a semantic gloss and syntactic information. In Figure 2, we show an example of annotated text where two collocations *avoir peur* (lit. ‘have fear’) and *peur terrible* (lit. ‘terrible fear’) are annotated. The first one is a support verb construction (Lexical Function Oper1, which is glossed as “feel”). The second one has both ‘intense’ and ‘negative’ qualities (LF: Magn + Antibon). The collocation is annotated on the collocative word, since a base word frequently has several collocates (here, *peur* is involved in two collocations, *avoir peur* and *peur terrible*).

```
Pour le coup le petit Chose <COLLOC BASE="peur" CAT="vt"
FL="Oper1" TYPESEM="éprouver">eut</COLLOC> une <LEXIQUE
TYPE="affect" CAT="N" DOMAINE="peur" NV LANGUE="courant"
INTENSITE="moyen" POLARITE="négatif">peur</LEXIQUE> <COLLOC
BASE="peur" CAT="adj" FL="Magn+AntiBon" TYPESEM="intense et
mauvais">terrible</COLLOC>; il se voyait déjà dans la rue, sans
ressources... (Le Petit Chose, A. Daudet)
```

Figure 2. An example of annotated text with Lexical Function

Analysis of the results showed that an overwhelming majority of collocations could be encoded with the help of standard LFs. Moreover, among these LFs, three of them (Magn, Oper and Caus) totaled more than half of the annotated LFs. These results show the empirical validity of this model. Moreover, a large number of collocations were used for few semantic functions (intensity, action, cause), which can be considered according to Wierzbicka's (1996) kinds of semantic primitives.



### 3.3 Regularities in base-collocate associations

We also observed a third type of regularity with the lexicalization of collocates. The lexical choice of the collocate was rarely completely arbitrary and generally corresponded to predictable or at least explainable rules. It has long been noted by linguists (inter alia, Cowie 1981; Allerton 1990; Mel'čuk, & Wanner, 1996) that the lexical elements of collocations show regularities. Very few collocations are completely idiosyncratic associations such as *peur bleue* (lit. 'blue fear') or *heureux comme un pape* (lit. 'happy as a pope'), in which the collocate is unpredictable. For most collocations, collocates can be used in co-occurrence with several bases which share semantic features. For example, *perpétrer* ('perpetrate') is used with bases such as *délit* or *crime* ('offence' and 'crime'). Several studies in the field of emotions (e.g. Mel'čuk & Wanner, 1996) support this view. In collaboration with F. Grossmann (Grossmann & Tutin 2007), we conducted a study on collocations of intensity associated with nouns of emotion. This issue had already been addressed by Mel'čuk and Wanner (1996) in German, in a study that dealt with regularities in base-collocate associations with inheritance mechanisms. Our hypothesis was stronger: we wanted to show that most associations could be explained or were even partially predictable. Our corpus-based semantic study highlighted several semantic dimensions for intensity such as /measure/ (*grande joie*, *joie profonde*, 'great joy', 'deep joy'), /authenticity/ (*vrai Bonheur*, 'true happiness'), /completeness/ (*Bonheur total*, 'total happiness'), /polarity/ (*affreux désespoir*, 'awful despair'). Associations between nouns of emotion and intensive adjectives proved to be strongly motivated and could be explained with the help of several association principles. For example, the 'polarity principle' can explain how a negative adjective such as *affreux* takes on intensive meaning in co-occurrence with a polar negative noun such as *désespoir*: as a general principle, polar adjectives tend to become intensive in co-occurrence with a noun of the same polarity, through a process of semantic bleaching. For example, the adjectives *affreux* and *horrible* become intensive in co-occurrence with *déception* ['disappointment' or 'deception'] but also with *remords* and *détresse* ['remorse' and 'distress']. On the other hand, these adjectives keep their usual meanings with positive nouns such as *joie* ['joy']. The semantic regularities of collocations can be even better observed in specific syntactic configurations.

To conclude this section, many regularities can be observed in collocations, which pose complex challenges when describing the latter. To our knowledge, at the moment, dictionaries of collocations do not adequately address these regularities.

## 4 Beyond binary associations: taking grammatical words and syntax into account with corpus-based methods

As previously mentioned, collocations are often described as binary combinations because they mainly entail a relation between two meaningful words. In fact, as they are syntactic constructions, collocations also show specific syntactic properties, mainly concerning alternations (i.e. passive alternations) or syntactic distributions, but also specific properties with grammatical words (colligations of collocations, in a way) (see Tutin 2004). We would like to show that these properties are an important part of the collocation.

In this section, we will first present the tool we used in order to observe these phenomena, the Lexicoscope, and before then discussing in more detail the syntactic properties which should be taken into account.

### 4.1 The Lexicoscope, a corpus-based collocation (and complex collocation) extractor

In order to extract collocations and their lexical properties, we used a tool designed by Olivier Kraif and Sascha Diwersy (Kraif & Diwersy, 2014), which uses treebanks in five European languages: French, English, German, Spanish and Russian. In a way, this tool is comparable to other kinds of collocation extractors like SketchEngine (Kilgarriff 2006), but it is free and it extracts complex collocations, that is to say collocations beyond two elements. Furthermore, there is also a concordancer for treebanks which allows the syntactic properties of collocations to be computed easily.

Simple binary collocations can be extracted with the help of a syntactic relation and an association measure (several measures are available). For example, Table 1 shows associations extracted from a large corpus of newspapers (112 M words) and novels (16 M) involving verbs related to the noun *peur* ('fear') as a direct object (with the following thresholds: frequency > 9, and log-likelihood measure  $\geq 10.83$ ). The corpus has been syntactically annotated with the help of Connexor (Tapanainen & Jarvinen, 1997). The results show that 3 V-obj-N collocations are very frequent with *peur*: *avoir peur*, *faire peur* and *prendre peur* (lit. 'have fear', 'make fear', 'take fear').

l1	l2	f.deprels	f	f1	f2	N	f.disp	am.log.likelihood	r.log.likelihood
peur_*	avoir_V	~obj	5487	8429	423602	6789513	8	19982,4109	1
peur_*	faire_V	~obj	1763	8429	340426	6789513	8	2599,0133	2
peur_*	prendre_V	~obj	245	8429	90682	6789513	8	118,4989	3
peur_*	vaincre_V	~obj	25	8429	1003	6789513	8	103,1044	4
peur_*	surmonter_V	~obj	20	8429	1500	6789513	8	58,9427	5
peur_*	apprivoiser_V	~obj	10	8429	271	6789513	4	48,8716	6
peur_*	conjur_V	~obj	12	8429	512	6789513	7	48,0540	7
peur_*	braver_V	~obj	11	8429	453	6789513	5	44,7965	8
peur_*	semer_V	~obj	17	8429	1713	6789513	5	41,0840	9
peur_*	ressentir_V	~obj	19	8429	2883	6789513	5	32,7031	10
peur_*	éprouver_V	~obj	15	8429	2766	6789513	5	21,1627	11
peur_*	alimenter_V	~obj	16	8429	3276	6789513	4	20,0240	12
peur_*	inspirer_V	~obj	14	8429	3127	6789513	6	15,7243	13
peur_*	susciter_V	~obj	24	8429	7881	6789513	7	14,6882	14
peur_*	avouer_V	~obj	15	8429	4309	6789513	6	11,6631	15

Table 1: Collocation extraction with the help of the Lexicoscope. (V-obj-*peur* collocations)

The system is also very convenient when it comes to making queries on specific elements, with syntactic relations, thanks to the dependency syntactic analysis computed with the help of the Connexor system, as shown in Figure 1 for the sentence *Il avait une peur malade de mourir* ('he had a pathological fear of dying').

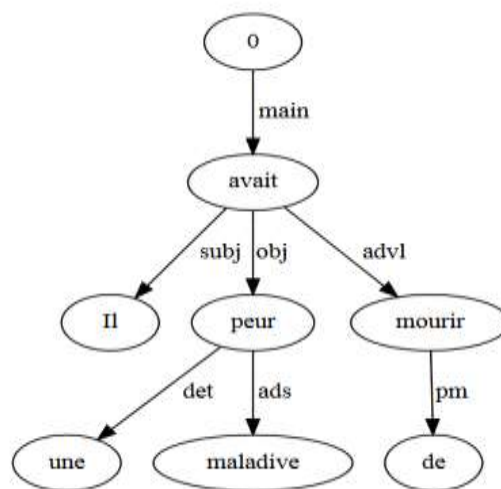


Figure 3: The syntactic analysis of *Il avait une peur malade de mourir* using Connexor

With the following query (see Figure 4), contexts could be extracted where *peur* was related to an attributive adjective in the collocation *avoir peur*. In our corpus, 71 occurrences of this kind were found.

**Nombre total d'occurrences 71**  
 Dispersion 8

### Requêtes

- Requête : <|=**peur**,#1> && <|=**avoir**,#2> && <c=A,#3> :: (obj,2,1)(ads|attr,1,3)

Show 10 entries Search:

Identifiant <sup>A</sup>	Contexte gauche	Pivot	Contexte droit
litmod_fr-orig-cnx-prs-tel-utf8.xml-s104789	Scoletti avait une	<b>peur</b>	panique de Grohm.
litmod_fr-orig-cnx-prs-tel-utf8.xml-s1049799	Il avait une	<b>peur</b>	maladie de mourir.

Figure 4: Concordances for expressions avoir –obj->peur-mod->Adj

The Lexicoscope is also designed to take complex collocations into account – it can extract co-occurring elements with a specific collocation. In figure 5, we can see an extraction of extended collocations from Verb-Object collocations involving the noun *admiration*. In the first step, for example, the collocation *forcer-admiration* (lit. ‘force-admiration’) was extracted. In the second step, the collocation was extended with the definite determiner, which is very specific for this collocation. Another interesting example was the collocation *caler-admiration* (lit. ‘hide-admiration’) which was extended with two specific elements: a negative adverb and a possessive determiner, that is *ne pas caler son admiration* (‘not hiding his/her/one’s admiration’).

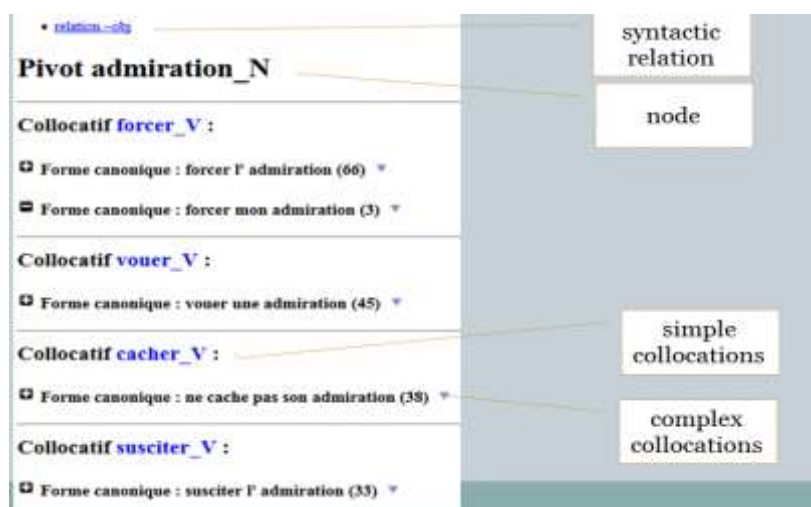


Figure 5: Extraction of complex collocations with Lexicoscope

Moreover, from extended collocations, the Lexicoscope is able to extract a lexical paradigm with the same lexico-syntactic configuration. For example, from the complex collocation *ne pas cacher son admiration*, it extracts other nouns which can be combined in the same way in the same construction *ne NegAdv cacher PossDet N*, as can be seen in Figure 6. Some grammatical words such as negative adverbs or possessive determiners are automatically extended. We can observe that the lexical paradigm extracted is mainly a subclass of emotion nouns. As we will see below, grammatically extended collocates are far more specific of semantic classes than simple collocations. In this case, the presence of negation and of the possessive determiner seem to heavily constrain the semantic class.

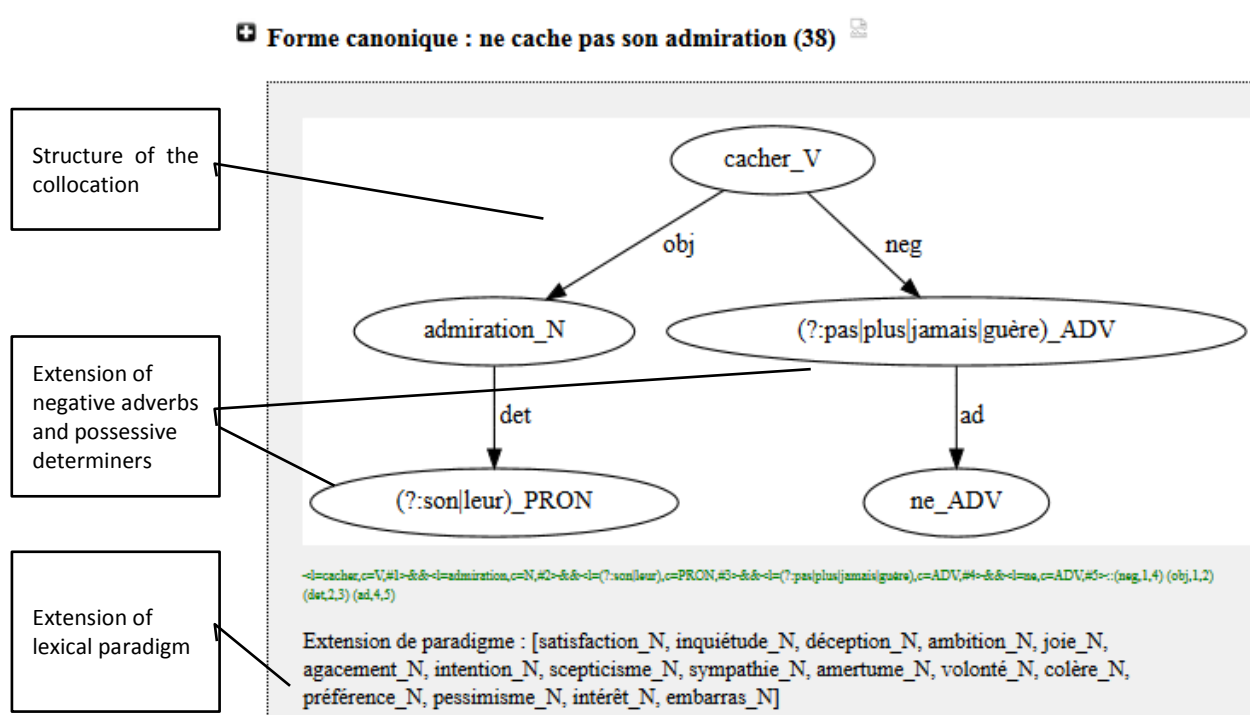


Fig. 6: Extension of lexical paradigm from the extended collocation *ne pas cacher* PossDet N

## 4.2 Alternations and distributions

As shown previously, we define collocations as combinations of two meaningful words related by a deep syntactic relation. It means that for example, the verb and the noun in a verb-object-noun collocation are for us related by a “deep” object relation (Cf Hagège & Roux, 2003). For example, the collocation *formuler-hypothèse* (‘formulate-hypothesis’) can occur in a simple direct object construction, as in (2):

- 2) La littérature spécialisée ayant montré le caractère assez peu généralisable (à travers les tâches) de la stratégie synthétique nous ne **formulerons** pas d'**hypothèses** spécifiques à son sujet. (article-psycho) ['Given that specialist literature has shown that the synthetic strategy can rarely be generalized (through tasks), we **will not formulate** any specific **hypotheses** in this regard.']

But it can also occur in several syntactic alternations such as passive construction (3), reduced passive (without any auxiliary) (4) and passive pronominal constructions (5).

- 3) Aucune **hypothèse** n'était **formulée** quant à l'effet de la centration en situation de soumission forcée standard [...]. (article psycho) [No **hypothesis** was **formulated** regarding the effet of focus in a standard situation of forced submission [...]]
- 4) Qu'en est -il alors des deux premières **hypothèses formulées** par Hennin? (article-histoire) ['What, then, of the first two **hypotheses formulated** by Hennin?']
- 5) Cette **hypothèse se formule** comme suit : le système réalise les meilleures approximations rationnelles des fréquences au sens des fractions continue... (article physique). ['This **hypothesis** is formulated [or lit. '**formulates itself**'] as follows: the system produces the best rational approximations of frequencies in the sense of continuous fractions']

These alternations and constructions, which have long been observed by several linguists, are not systematic (Weller & Heid, 2010). For example, the very frequent collocation *avoir-peur* (lit. 'have fear') only enables the standard direct object construction and no alternation whatsoever is allowed.

Furthermore, in some collocations the "standard" construction is far less frequent than certain alternations. For example, in some subject-verb collocations, passive constructions are more frequent than active constructions. This is often the case with collocations in the field of emotions, for example *envahir-SUBJ->tristesse* where the passive construction (*être envahi de tristesse*, 'to be overcome by sadness', e.g. (7) is far more frequent than the active construction, e.g. (6)).

- 6) Souvent, quand cette **tristesse l'envahissait**, Élyse lâchait à son entourage : « Je me sens mal, comme si j'avais perdu une jumelle ». (Ouest-France) [Often, when this **sadness would overcome** her, Elyse would come out and say to those around her: 'I don't feel well, as if I'd lost a twin'].
- 7) En regardant autour de moi, je suis **envahie de tristesse** car je vois la mort de mes rêves. (Literary corpus) ['Looking around me, I'm **overcome** with **sadness** as I can see the death of my dreams']

In lexicographic and lexicological descriptions, it is important to be able to provide the "instructions for use" of the expression since some constructions are really far more frequent than others.

Another interesting case to deal with is the distribution of adjectives. In French, adjectives can have two distributional positions when they are attributive: they can occur in front of the noun (pre-posed adjectives) or after the noun (post-posed adjectives). In some collocations, for example highly idiosyncratic ones such as *peur bleue* (lit. ‘blue fear’ or *célibataire endurci* ‘hardened singleton’, the position is generally fully fixed, either preposition or postposition. But in some cases, adjectival collocations, especially regular ones, can occur in several positions. For example, for the collocation *joie*-ADJ->*immense*, in our corpus, we observe 12 pre-posed adjectives (e.g. (8)), 14 post-posed adjectives (e.g. (9)) and even 4 predicative adjectives (e.g. (10)).

- 8) Le "profond soulagement" du gouvernement de Tony Blair n'avait d'égal que l'**immense joie** des familles, après l'annonce, [...], de la libération des 15 marins britanniques capturés par l'Iran le 23 mars. (Le Monde, 2008) [‘The “profound relief” of Tony Blair’s government was only matched by the **immense joy** of the families, after the announcement [...] of the liberation of the 15 British navy personnel captured by Iran on March 23’]
- 9) Ma famille et moi-même avons éprouvé une **joie immense** à cette annonce. (Ouest-France 2007) [‘My family and I felt immense joy [lit. **joy immense**] after this announcement’]
- 10) Forcément, si on gagne avec un but décisif de ma part, la **joie** sera **immense**. (Ouest-France, 2007) [‘Of course, if we win with a deciding goal on my part, the **joy** will be **immense**’]

#### 4.3 Colligations of collocations

Not only alternation and distribution should be taken into account, but also modifiers and subcategorized elements. For example, some V-OBJ->N collocations show very different properties with determiners and modifiers. For our demonstration, we will take the three most frequent verbal collocations of our corpus in which *peur* is a direct object (see Table 1): *avoir peur* (5487 occ.), *faire peur* (1763 occ.) and *prendre peur* (245 occ.).

If we first look at the determiners (cf. Table 2), we can observe that *prendre peur* has no determiner, while it is quite rare with *faire peur* and a bit more frequent with *avoir peur*. We observe the same tendencies with the adjectival modifier, which is actually often associated with the determiner (using an attributive adjective entails the presence of an adjective, generally the indefinite article, as in the following example).

- 11) Il **avait une peur malade** de mourir. [‘He **had a pathological fear** of dying’]  
Bernard Werber, *L’ultime secret*, 2001.

We also notice that examples with plurals, which are very sparse, only occur with *avoir*. If we now look at the complement of the collocation, we can observe that with the collocation *avoir peur*, the noun *peur* often has a verbal complement (see example 11: *peur de mourir*, ‘fear of dying’) or a nominal complement as in (12: *peur des policiers*, ‘fear of the police’):

- 12) Les gens ont peur. Mais, alors qu'avant ils **avaient peur des criminels**, ils **ont** aujourd'hui **peur des policiers**. *Libération*. 14/02/2007. [‘People are scared [lit. have fear]. But whereas before, they were scared [lit. **had fear**] **of criminals**, today they are scared [lit. they **have fear**] **of the police**.’]



Collocation	Number of occurrences	% of nouns in the plural	With a determiner	With an adjectival modifier	With a nominal complement. Eg: <i>peur du noir</i> ('fear of the dark')	With an infinitive complement Eg: <i>peur de sauter</i> ('fear of jumping')
<i>faire peur</i> lit. 'make fear'	1763	0 (0%)	9 (0.5%)	8 (0.5%)	3 (0.02%)	8 (0.5%)
<i>avoir peur</i> lit. 'have fear'	5487	6 (0.1 %)	107 (1.9%)	60 (1.1 %)	998 (18.1%)	1070 (19.5%)
<i>prendre peur</i> lit. 'take fear'	245	0	0 (0%)	0 (0%)	8 (3.2 %)	0 (0%)

Table 2: Syntactic properties of verbal collocations with *peur*

Furthermore, some specific determiners occur in collocations. For example, the verbal expression of emotions often takes possessive determiners (*crier, dire son admiration*, lit. 'to shout, to say his/her/one's admiration'). For some verbal collocations, constraints can also be noted with the negative adverb. For example, the collocation *cache-Obj->admiration* is especially frequent with two grammatical words: the possessive determiner and a negative adverb, as we can see in Table 3.

Collocation	Number of occurrences
<i>Cacher admiration</i> 'Hide admiration'	56 (100%)
<i>Cacher mon/ton/son admiration</i> 'Hide my/your/his-her)one's admiration'	49 (87.5%)
<i>Ne jamais/plus/pas cacher mon/ton/son admiration</i> 'Never/no longer/not hide my/your/his-her-one's admiration'	36 (64%)

Table 3: Colligations of the collocation *cache-admiration*

These grammatical elements belong to the collocation and should be taken into account in dictionaries. While, to our knowledge, “classical” dictionaries of collocations for French do not include this information, one lexical resource for French does provide quite complete information about collocations. It is the dictionary of word combinations in the writing aid *Robert Correcteur* (Fig. 7), which offers a graphical presentation of the most relevant collocations with their colligations, including the grammatical colligations of collocations. For example, for the collocation *hypothèse vraisemblable* (‘likely hypothesis’), the most likely determiner, *une* (‘a’), is presented while the adverb *peu* (lit. ‘little’) appears as a variant of the collocation. However, the alternations presented above (see 4.2) are not yet mentioned in this very convenient lexical resource.



Figure 7: The *Robert correcteur* dictionary of word combinations

#### 4.4 Extended collocates and semantic classes

We previously observed that some collocations usually work within a specific grammatical context: reciprocally, we have also noticed that for a given lexical collocate, some colligational features may induce a specific class of bases. For instance, we have extracted from a literary corpus various occurrences of the nouns which appear as object of the collocate verb *éprouver* (‘to feel’). Then we made a classification of these nouns in two groups: *Emo+*, the nouns which are related to the semantic class of affect, and *Emo-*, the nouns which are not (to have more details about the criteria that were used for this classification, see Kraif, 2014). It appears that the nature of the determiner and the presence of an adjectival modifier is highly correlated with the former semantic classification:

Extended collocates	<i>Emo+</i>	<i>Emo-</i>
<i>éprouver</i> + definite article + noun	24 (55,8 %)	19 (44,2 %)

<i>éprouver</i> + indefinite article + noun	157 (89,7 %)	18 (10,3 %)
<i>éprouver</i> + indefinite article + adj + noun	94 (91,26 %)	9 (8,74 %)

Table 4: Correlation between extended collocates and the *Emo+* semantic feature

We note that in the context of the definite article *le/la*, the object of *éprouver* is not strongly related with the semantic field of affect (44,2% of occurrence are not), whereas in the context of the indefinite article *un/une*, which is far more frequent, we find a very pronounced correlation. If we consider the context of *un/une* associated with an adjective, the correlation appears to be even stronger.

To determine the semantic consistency of that kind of extended collocate, we have selected several extended collocations, and tried to search all the nouns that could commute with the original base. For instance, taking the extended collocation *en concevoir une amertume*, ('to feel bitterness') we made the query corresponding to the right-handed tree in figure 8 :

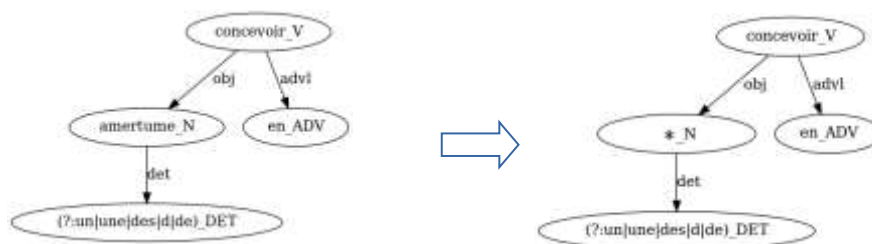


Figure 8 : from a collocation to an extended collocated

From this query, using the Lexicoscope, we found the following nouns, with relatively homogenous senses, expressing grief and sadness :

- 13) *en concevoir un/une + amertume, chagrin, déception* [bitterness, grief, disappointment]

The latest example is extracted from the literary part of EmoBase, but within the newspapers corpus we can find similar phenomena, for instance with the extended collocation *pour calmer le/la + noun* (we underlined the few nouns semantically inconsistent with the remainder of the class):

- 14) *pour calmer le/la + colère, jeu, esprit, grogne, tension, surchauffe, ardeur, inquiétude, mécontentement, fronde, ire, impatience, douleur, crainte, prix, monde, crise* [to calm the anger, game, spirit, discontent, tension, overheating, revolt, wrath, impatience, pain, fear, price, world, crisis]

It appears that most of the nouns which commute in this context are related to the senses of anger and public discontent.

## Conclusion

Collocations have long been recorded by lexicographers and language teachers, but have only been theorized recently, probably because of their hybrid nature, halfway between syntax and lexicon. As part of prefabricated language, they fully belong to the field of phraseology. Compared to other kinds of phrasemes, they present two remarkable properties: they are highly regular (especially regarding semantic properties) and they also present strong syntactic variability. These two properties still need to be encoded in dictionaries of collocations in order to provide “instructions for use” of these expressions, especially for language learners.

## References

- Allerton, D.J. (1990). Linguistically Strange Combinations. *On strangeness*, ed. M. Bridges, 25-38. Tübingen: Gunter Narr Verlag.
- Alonso Ramos, Margarita 2004. Elaboración del Diccionario de colocaciones del español y sus aplicaciones. *Actes del I Symposium Internacional de Lexicografia*, 16-18 maig 2002. 149-162.
- Augustyn, Magdalena & Agnès Tutin 2009. Constitution d'un corpus annoté autour du lexique des émotions : collocations et fonctions lexicales, in D. Beck, K. Gerdes, J. Milićević, A. Polguère (eds), *Actes de la quatrième conférence internationale sur la théorie Sens-Texte*, 25-34. Montréal: OLSST.
- Bally, Charles. 1909. *Traité de stylistique française*. Paris : Klincksieck.
- Bartsch, Sabine 2004. *Structural and Functional Properties of Collocations in English. A Corpus Study of Lexical and Pragmatic Constraints on Lexical Co-Occurrence*. Tübingen: Narr.
- Cavalla, Cristelle 2008. Les collocations dans les écrits universitaires: un lexique spécifique pour les apprenants étrangers. *Apprendre une langue de spécialité: enjeux culturels et linguistiques*, Olivier Bertrand et Isabelle Schaffner (eds), 93-104. Paris : Editions Ecole Polytechnique.
- Cowie, Anthony P. 1981. The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics* 2: 223–235.
- Evert, Stefan 2008. Corpora and collocations. *Corpus Linguistics. An International Handbook*, 2, A. Lüdeling and M. Kytö (eds), 223-233. Berlin: Mouton de Gruyter
- Fontenelle, Thierry. 1997. *Turning a bilingual dictionary into a lexical-semantic database* (No. 79). Max Niemeyer Verlag.
- Grossmann, Francis & Agnès Tutin (eds) 2003. *Les acollocations : analyse et traitement*. Amsterdam: de Werelt.
- Hagège, Caroline & Claude Roux 2003. Entre syntaxe et sémantique : Normalisation de la sortie de l'analyse syntaxique en vue de l'amélioration de l'extraction d'information à partir de textes. *Actes de TALN, TALN 2003*. Batz-sur-Mer, 11–14 juin 2003.
- Hausmann, Franz J. 1979. Un dictionnaire des collocations est-il possible?. *Travaux de Linguistique et de Littérature Strasbourg* 17(1) : 187-195.

- Heid, Ulrich 1994. On Ways Words Work Together - Topics in Lexical Combinatorics. *Proceedings of EURALEX'94 Proceedings*, 226-257. Vrije Universiteit Amsterdam.
- Kraif, Olivier & Sascha Diwersy 2014. Exploring combinatorial profiles using lexicograms on a parsed corpus: a case study in the lexical field of emotions. *Nouvelles perspectives en sémantique lexicale et en organisation du discours*. Blumenthal, Peter, Novakova, Iva, Siepmann, Dirk (eds), 381-394. Frankfurt: Peter Lang.
- Kraif, Olivier (2014). *Corpus parallèles, corpus comparables : quels contrastes ?*, Synthèse d'habilitation à diriger des recherches, Université de Poitiers.
- Le Robert Correcteur*. Dictionnaires Le Robert. 2015.
- Lewis, Michael. 2000. *Teaching Collocations*. Boston: LTP.
- Mel'čuk, Igor & Leo Wanner (1996). Lexical functions and lexical inheritance for emotion lexemes in German. *Lexical functions in lexicography and natural language processing*, Leo Wanner (ed), 31, 209. John Benjamins Publishing.
- Mel'čuk, Igor 1998. Collocations and Lexical Functions. *Phraseology. Theory, Analysis and Applications*, Anthony. P. Cowie (ed), 23-53. Oxford : Clarendon Press.
- Mel'čuk, Igor, Clas, André & Alain Polguère 1995. *Introduction à la lexicologie explicative et combinatoire*. Bruxelles : De Boeck Supérieur.
- Meunier, Fanny & Sylvian Granger (eds.) 2008. *Phraseology in foreign language learning and teaching*. John Benjamins Publishing.
- Nesselhauf, Nadja 2005. *Collocations in a learner corpus* (Vol. 14). John Benjamins Publishing.
- Palmer, Harold E. 1933. *Second Interim Report on English Collocations*. Tokyo: Kaitakusha.
- Sanromán Vilas, Begoña 2005. Individual -level and stage-level predicates: the Spanish Emotion Nouns. *East West Encounter: Second International Conference on Meaning Text Theory*, 417-431 Moscow: Slavic Culture Language Publishing House..
- Seretan, Violeta 2011. *Syntax-based collocation extraction* (Vol. 44). Springer.
- Siepmann, Dirk 2005. Collocation, Colligation and Encoding Dictionaries. Part I: Lexicological Aspects. *International Journal of Lexicography*, 18(4): 409-443
- Sinclair, John 1991. *Corpus, concordance, collocation*. Oxford :Oxford University Press.
- Tapanainen, Pasi & Tmo Järvinen (1997). A non-projective dependency parser. *Proceedings of the 5th Conference on Applied Natural Language Processing*, Association for Computational Linguistics, Stroudsburg, 64-71.
- Tesnière, Lucien 1959. *Eléments de syntaxe structurale*. Librairie C. Klincksieck.
- Tutin Agnès, 2012. Les collocations dans le champ sémantique des émotions: la régularité plutôt que l'idiosyncrasie, *Meanings, texts and other exciting things : a festschrift to commemorate the 80th anniversary of Professor Igor Alexandrovic Mel'čuk*, Youru Apresjan et al. (éds.), 602-612. Moscou : Jazyki slavanskoj kul'tury.
- Tutin, Agnès & Francis Grossmann 2002. Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif. *Revue française de Linguistique appliquée*, vol. VII-1. : 7-25.
- Tutin, Agnès 2010. Le traitement des collocations dans les dictionnaires monolingues de collocations du français et de l'anglais. *2ème Congrès Mondial de Linguistique Française*. EDP Sciences, 1175-1090.
- Tutin, Agnès 2010. *Sens et combinatoire lexicale, de la langue au discours*. Mémoire d'habilitation à diriger des recherches. Grenoble : Université Stendhal.
- Tutin, Agnès 2013. Les collocations lexicales: une relation essentiellement binaire définie par la relation prédicat-argument. *Langages*. 189/1: 47-63.
- Weller, Marion & Ulrich Heid 2010. Extraction of German multiword expressions from parsed corpora using context features. *Proceedings of LREC 2010*, 3195-3201.
- Wierzbicka, Anna 1996, *Semantics: Primes and Universals*, Oxford: Oxford University Press.

Williams, Geoffrey 2003. Les collocations et l'école contextualiste britannique. *Travaux et recherches en linguistique appliquée. Série E, Lexicologie et lexicographie*, Francis Grossmann & Agnès Tutin (eds), 33-44. Amsterdam: de Werelt.