For an extended definition of lexical collocations

Agnès Tutin

For an extended definition of lexical collocations

Agnès Tutin
Laboratoire de Linguistique et Didactique des Langues Etrangères et Maternelles
(LIDILEM), Université Grenoble 3
BP 25, 38040 Grenoble Cédex 09
e-mail : agnes.tutin@u-grenoble3.fr
Tél : 04 76 82 43 00   Fax : 04 76 82 43 95

Abstract

In this paper, we address two main problems raised by the classical definitions of restricted lexical collocations: 1) the binary status of collocations; 2) the problem of function words which are often implicitly excluded from the definition of collocations We show that collocations are fundamentally argument-predicate structures and that most ternary collocations can be considered as combinations of binary collocations. We think that content grammatical words should be considered as parts of collocations, while pure functional words are not central elements of the collocations.

1. Introduction

In European tradition, two main conceptions of collocations co-exist: in the British contextualist framework (Firth; Halliday & Hasan 1978; Sinclair 1993; Williams 2003), collocations can be broadly defined as recurrent lexical elements which contribute to the text cohesion. In the "continental" tradition (Williams 2003), collocations are also called "restricted lexical collocations" and considered as lexicalised phrases where two recurrent lexical elements have a syntactic relationship. In this paper, we will address several problems related to the continental definition of collocations.

Restricted lexical collocations have now been studied and encoded in dictionaries for over than twenty years, and stable definitions have been provided for this notion by numerous scholars working on collocations (e.g. Hausmann 1989, Mel'čuk 1998, Heid 1994, Tutin & Grossmann 2002). They are roughly defined as recurrent combinations of two linguistic elements which have a syntactic relationship. One of the elements of the collocation, called "base" keeps its usual meaning (autosemantic words (Hausmann 2004)) while the other, the "collocate" is dependent on the other (synsemantic words) and usually has a less transparent meaning.

Nevertheless, even though such a definition is operational for a large number of lexical associations, it raises several problems. The first one is the binary status of the collocation and the unequal status of the two parts of the collocation, which has been questioned by several linguists (*inter alia* Siepmann 2006, Bartsch 2004) who suggest expanding the definition to associations of three or more elements. A second problem concerns the grammatical status of the collocations. Should functional words – and to what extent – be included in the definition of collocation? For example, in expressions such as *for fear of*, the whole combination can be analysed as a preposition, and not as a

phrase contrary to prototypical collocations such as *pay attention* (verb phrase), *major problem* (noun phrase), *seriously injured* (adjective phrase). However, *fear* in *for fear of* can be considered as relatively transparent, and according to us should be considered as a collocation.

In this paper, we address these two issues in detail and call for an extended typology of restricted collocations. We examine the lexicographical consequences of such an extended definition.

2. Are collocations fundamentally binary associations?

Collocations are considered as binary associations in classical definitions of lexical restricted collocations in two ways:

1. Collocations are associations of <u>two</u> lexical units, or broadly speaking, two linguistic elements :

2. The two elements of the association have a <u>different status</u>: the base is the prominent element while the collocate depends on the base.

These two characteristics are present in these classical definitions of Hausmann, Mel'cuk and Heid.

> On appellera collocation la <u>combinaison de deux mots</u>[1] … (Hausmann 1989 :10010)
> [We will call collocation a two-word combination].

> […] collocations are combinations of exactly <u>two lexemes</u> (of category noun, verb, adjective or adverb), realizing two concepts where the choice of one of them depends on (or : is restricted by) the other. (Heid 1994 : 228)

> A COLLOCATION **AB** of language **L** is a semantic phraseme of **L** such that its signified 'X' is constructed out of the signified of one of its <u>two constituent lexemes</u> – say, of **A** – and a signified 'C' ['X' = 'A⊕C'] such that the lexeme **B** expresses only 'C' contingent on **A**. (Mel'čuk 1998 : 30)

These two characteristics seem to be operational for a large number of multiword units since in a large number of dictionaries (e.g. BBI, OCDSE, LTP, DC) provide collocational information within the entries of the base, and most collocations (all of them in LTP and DC[2]) are binary lexical collocations. For example, *pay attention* is recorded in the BBI, OCDSE or LTP dictionaries within the entry dictionary of *attention*. We observe the same facts for the French equivalent *prêter attention* in DC.

It must be emphasized that the two elements are not necessarily simple words. The base and the collocate can be fixed idioms, for example: a) *travailler d'arrache-pied* ('to

---

[1] The underlining is ours.

[2] In the LTP, collocations are provided within the base entries. Syntactic models are associations of two content words: V-N, N-V, A-N, N-N, V-ADV. We find the same kinds of associations in Beauchesne's DC.

work hard') where *travailler* is the base and *d'arrache-pied* is a collocate; b) *point de vue classique* ('a classical viewpoint') where *point de vue* is the base and *classical* the collocate. More interestingly, some expressions like similes, used as collocates, (*to sleep* (base) *like a child* (comparison phrase as a collocation) cannot really be considered as lexicalized. These cases show that the notion of lexical element in the definition should be extended to idioms and even to expressions like similes for collocates.

However, several linguists have questioned this binary status (Hausmann 2007; Siepmann 2006; Bartsch 2004), because 1) some apparent collocations are larger than two constituents, and because 2) no dominant constituent seems to emerge in some examples.

2.1 Some collocations are larger than two constituents

2.1.1 Some collocations like *to pay close attention* apparently include three or more elements. Nevertheless, most of them in this case can be analysed as a merging of two or three binary collocations ("collocational chains" for Hausmann (2004; 2007), "collocational cluster" for Spohr (2005)). For example, the sequence *to pay close attention* can be decomposed into two collocations: *to pay attention* (a support verb construction) and *close attention* (an intensive, "Magn-like" collocation in ECD words, according to Mel'čuk's terminology). Though *close attention* very often collocates with *pay,* this collocation can be encountered in other contexts: a rapid Google search yields for example *close attention is required, to recommend close attention, this issue needs close attention ...*

Even tricky cases can often be analysed as merged collocations. For example, D. Siepmann (2006) gives the example of *avoir un geste déplacé* as an example of a ternary collocation (Lit. "to have an inappropriate gesture") and shows that the adjective cannot be removed: **avoir un geste* (Lit. "to have a gesture"). But, for us, this problem is largely due syntactic constraints, viz. to the presence of the indefinite determiner. In this context, other determiners *il a eu ce geste, le geste qu'il a eu ...* seem perfectly appropriate with this collocation without any modifier. On the other hand, the verb can easily be deleted and the collocation *un geste déplacé* (lit "inappropriate gesture") can be encountered in other contexts: *se permettre un geste déplacé, l'auteur d'un geste déplacé, à la suite d'un geste déplacé ...*

In other words, there is a syntactic constraint on the collocation *avoir un geste* and a modifier is required, but there is no lexical constraint on this modifier (*geste d'horreur, geste rapide ...*). The sequence *avoir un geste déplacé* can thus be analysed as the merging of *avoir un geste* (+ Modifier) and *geste déplacé.* This case shows that close attention should be paid to syntactic constraints on collocations, and that in this case, specific constraints should associated to the determiner. As suggested by Heid & Gouws (2006), we think that morphosyntactic properties of collocations should be accounted for in detail in lexicographic subentries ("second level treatment units").

Syntactic constraints of collocations could also explain why two collocations cannot merge into a collocational cluster in some cases. For example, while *avoir peur* (Lit. "to have fear") and *une peur bleue* (Lit. "a strong (blue) fear") can be merged into *avoir une*

peur bleue), this is not the case for *prendre peur* (Lit. "to gain fear") and *peur bleue* : **prendre une peur bleue*. The difference lies in the syntactic structures of the collocations *avoir peur* and *prendre peur*. While the first one also has the structure *avoir une peur +* Modifier, this is not the case with *prendre peur,* where the noun cannot be modified, **prendre une peur +* Modif, probably because it has an inchoative value which precludes modification.

2.1.2 In some examples, analysing a complex collocation as a cluster is not possible. For example, in sequences like *in other words,* the sequence *in* Adj *words* only appears with *other*, and if *other words* is possible, it does not have exactly the same meaning as for the collocation. This collocation can be considered as a kind of real ternary collocation with *words* as a base, while *in* and *other* would be the collocates. If one wants to keep the binary status of the collocation (since *words* is intuitively the "base" of this complex collocation), one may suggest that *in other* is a kind of complex collocate.

2.2.2 The two previous cases must be distinguished from "recursive collocations", when collocations are inserted into collocations. In this case, collocations can themselves be used as a base or as a collocate. In the sequence *to fall in love*, *in love* can be considered as a collocation for *love* and exists independently: *a woman in love* (with). *fall,* the collocate, would be the specific inchoative verb associated to this adjectival collocation. In the sequence *freshly baked bread* (example of Bartsch (2004: 67)), *freshly baked* can be considered as a collocational collocate of *bread*, *freshly* being the collocate of the collocation *freshly baked* that can be encountered in other contexts: *freshly baked cakes, freshly baked buns*, *freshly baked cookies* … The existence of recursive collocations also provides a strong argument to develop collocational subentries in dictionaries as advocated by Heid & Gouws (2006).

The table 1 summarizes these different kinds of ternary combinations.

Phenomenon	Definition	Example (collocates are underlined)
Merged collocations into clusters (if syntactic patterns unify)	Two collocations which have the same base and can syntactically combine.	*pay <u>attention</u> + <u>close</u> attention* → *pay <u>close</u> <u>attention</u>*
Recursive collocations	Case 1 : the base is a collocation.	*<u>fall</u> (<u>in</u>(love))*
Recursive collocations	Case 2 : the collocate is a collocation	*(<u>freshly baked</u>) bread*
True ternary (or more) collocations	Two or more collocates can be associated to the base. The collocation cannot be decomposed	*<u>In other</u> words*

Table 1: Different kinds of ternary combinations

In dictionaries, merged collocations and recursive collocations should be decomposed whenever they can be, but very frequent clusters could be mentioned if necessary within the base entries. True ternary collocations could be mentioned within the base entry.

2.2. Are all collocations dissymmetric?

2.2.1 Most dictionaries of collocations (e.g. BBI, OCDSE, LTP) provide collocational information within the base entries: s.v. *attention* for *pay attention* or *close attention*, s.v. *bachelor* for *confirmed bachelor*, s.v. *injured* for *seriously injured* or *severely injured* and the dissymmetry of collocations is *de facto* institutionalized in this lexicographic practice, even though it may be assumed that the collocation dictionary user probably uses more syntactic clues than the notions of base and collocate to access lexical information. She knows that she will find verb-noun collocations within the noun entries and verb-adverb collocations within verb entries, and probably uses more this syntactic hint than the complicated notion of base and collocate.

According to us, the question of the dissymmetry between the two elements of the collocation should not only be analysed in terms of psycholinguistic salience of one element. We think that this notion is better analysed in terms of semantic properties than with lists of syntactic patterns. Collocations can in general be analysed as predicate-argument structures, where the base is the argument and the collocate the predicate, for example the modifying adjective of a noun (*heavy* (pred) *smoker* (arg)) or the subject or the object of a verb (*la tristesse* (arg) *envahit* (pred) ; *mourir* (pred) *de tristesse* (arg)). The semantic content of the argument is generally more precise (it is often a noun) than the semantic content of the predicate (generally an adjective or a verb) whose meaning is specified when it applies to a given argument.

2.2.2 According to several linguists (Siepmann 2006; Hausmann 2007; Gonzales Rey 2002; Bartsch 2004), in some lexical associations though, it is difficult to decide what element should be considered the base, since no dominant constituent seems to emerge. Interestingly, one can notice that the tricky cases often examined in the literature have the same syntactic properties, as pointed out by Hausmann (1996):

(a) **N prep N** pattern, e.g. in *a pack of dogs*[3] (is the base *pack* or *dogs*?) or *an outburst of anger*. If the quantitative interpretration is chosen ("Mult" in the language of Lexical Functions for *pack* or "Sing" for *outburst*), *dogs* or *anger* could be considered as the bases, while if the non quantitative interpretation is chosen, they would be considered as collocates of content words.

(b) **V Prep N** pattern like in *blush with shame,* where Prep N can be interpreted as an adverbial (the manifestation) or a causative complement.

In both structures, two predicate-argument structures can be pointed out, hence the ambiguity concerning the identification of the base. As advocated by Hausmann (2007), we think that these two-base collocations should be stored within both dictionary entries.

---

[3] (Bartsch 2004 : 36)

2.2.3 Apart from these cases, a difficult issue is the case of conjoined collocations, such as *rich and famous, sain et sauf, safe and sound, slowly but surely …* These associations share several characteristics with collocations: the sequence includes two recurring content words and it is quite transparent from a semantic viewpoint. The two elements are irreversible: **sauf et sain, *surely but slowly…* From a syntactic viewpoint, the coordination suggests that no dominant element emerges, and from a semantic viewpoint both elements seem to have an equal importance. If such combinations could be considered as (atypical) collocations in a way – due to their binary character and their transparency – they do not have a prominent element like a base. Several collocation dictionaries store these two-base collocations within both entries. For example, *safe and sound* in the OCDSE is recorded s.v *safe* and s.v. *sound*, which seems the best solution.

3. Should collocations include function words, and to what extent?

3.1 Determiners

Function words are the poor relation of phraseology. In studies about collocations, they are often regarded as secondary. Definitions of collocations generally assume that the syntactic relation holds between two content words or constituents, what would implicitly exclude function words from being part of a collocation. Nevertheless, as pointed out by Bartsch (2004) or Siepmann (2006), function words are essential in the syntactic description of collocations. In examples such as *to commit a crime*[4] or *faire une promenade* ("to take a walk"), the determiner is essential in the collocations (the determiner cannot be omitted: **to commit crime, *faire promenade* (Lit. 'to make walk')).

However, one can notice that the information about the determiner is highly likely to vary according to contexts (*to commit crimes, he committed this horrific crime …*). In other words, the word *a* is not a stable element in the collocation *to commit a crime*. Collocational elements are here inserted in a syntactic construction where a determiner is compulsory and this determiner can vary (*to commit DET crime*). As pointed out above, the syntactic construction of the collocation (constraints about determiners, syntactic alternations, subcategorization of the collocation[5]) should be detailed, but according to us, variable determiners and subcategorized elements should not be considered as essential parts of the collocations, except in specific cases.

3.2 Prepositions

All prepositions do not have the same status. This is very clear in some classical syntactic models like Lexical Functional Grammar (Bresnan & Kaplan, 1982): prepositions which are used to introduce arguments (i.e. *I send that to him*) cannot be considered as predicates (they generally have a weak meaning (or no meaning at all) and their choice is determined by the verbal predicate, here *to send*), while prepositions used to introduce modifiers are generally considered as predicates. In *they did that for fear*, *for* has a causative meaning. The same prepositions can generally be used to

---

[4] Example of Bartsch (2004:36).
[5] For example, (Someone) gives a talk (about something) (to an audience).

introduce either arguments or predicates, even if they do not have the same meaning in these two contexts. For example, *de* in *je me souviens de ça* (Lit. "I remember of that") *de* is a pure function word and can be considered as an empty word while in *de dépit* ("out of pique") *de* introduces a causative meaning.

Functional empty words should not be considered as main elements of the collocations (even if the subcategorized prepositions of the collocation should be adequately described). Meaningful grammatical words should, however, be included in collocations such as *for fear, out of love, as a result* where they are predicates and can be considered as collocates of the base argument[6]. This is already the case for ECD-like collocation dictionaries like the DAFLES and the LAF, but the inclusion of meaningful grammatical words is not yet generalized in most collocation dictionaries. In OCDSE, subcategorized (empty) prepositions like *of* in *fear of failure* and meaningful prepositions like *out of (fear)* are stored within the same PREP paragraph, although they fulfil a very different function from a semantic viewpoint.

3.3 Complex prepositions and conjunctions including content words

Some sequences such as *for fear of, with the aim of, de peur que* ("for fear of"), in which the noun keeps its usual meaning, are problematic. A large set of these collocations are stored in some dictionaries of collocations like the DAFLES, the BBI or the OCDSE, as can be seen in table 2.

Collocation	DAFLES	DC	OCDSE	BBI	LTP
de peur de, de crainte de *for fear of*	+	-	+	+	-
dans le but de *with the aim/intention of*	-	-	+	-	-
dans l'espoir de/que *in the hope of*	+	-	+	+	-

Table 2 : Complex prepositions and conjunctions including content words in dictionaries of collocations

These collocations involve two function words which are discontinuous. The first ones, *for* (in *for fear* of) or *with* (in *with the aim of),* are meaningful prepositions, while the second ones (a preposition or a conjunction) are in a way subcategorized by the noun, and can be encountered in other contexts, as can be observed in (1) and (2) (Cf. Gross, 1986):

---

[6] Several linguists do not include this kind of syntactic pattern (Prep + N) within the set of collocational patterns (Heid and Gouws 2006; Spohr 2005; Hausmann 1989).

(1) Because they had a terrible <u>fear of failure</u>, they did not take the exam.
(1') Parce qu'ils avaient très <u>peur d'un échec</u>, ils n'ont pas passé l'examen.
(2) <u>For fear of failure</u>, they did not take the exam.
(2') <u>De peur d'échouer</u>, ils n'ont pas passé l'examen.

Moreover, most of these expressions can be encountered without subcategorized elements:

(3) They did that for fear.
(3') De peur, elle a fait ça.
(4) They did that with this aim.

These properties seem to show that sequences like *for fear of*, *with the aim of* or *de peur que* should not really be considered as complex conjunctions or prepositions, but like prepositional collocations in which the noun can subcategorize conjunctive sentences (*peur que* P (Lit 'fear that S')) or prepositional phrases (*fear of* NP).
Due to their syntactic properties, the sequences should be considered as prepositional collocations (equivalent to a PP) and their syntactic properties (in particular, subcategorization) should be adequately accounted for.

Conclusion

Collocations can be considered as predicate-argument structures, and as such, are prototypically binary associations, where the predicate is the collocate and the argument is the base. Most ternary (and over) collocations are merged collocations (collocational clusters) or recursive collocations. For language users, however, it can be useful to have access to these productive clusters or recursive collocations in dictionaries. Some atypical collocations do not show a dissymmetric base-collocate structure because they do not have a clear predicate argument structure.

Content grammatical words should be considered as parts of collocations (collocates), while other grammatical words like determiners and subcategorized prepositions or conjunctions should be adequately described but are not central elements of the collocations.

Dictionaries of collocations like the DAFLES, the BBI and the OCDSE are less timid than linguists in their description of collocations. They already include atypical collocations like conjoined collocations and collocations including grammatical words. Defining clear linguistic properties could probably enhance the microstructure organization of such dictionaries and enable adequate computational treatment of collocations.

Dictionaries

BBI : *The BBI dictionary of English word combinations* (1997), Benson M., Benson E. and Ilson R. Amsterdam ; Philadelphia, J. Benjamins.

DAFLES, Dictionnaire d'Apprentissage du Français Langue Étrangère ou Seconde:

http://www.kuleuven.ac.be/ilt/dafles.htm

DC : Beauchesne, J. (2001). *Dictionnaire des cooccurrences*. Montréal, Guérin.

DEC : Mel'čuk I. *et al.* (1984, 1988, 1992, 2000). *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques 1*, *II, III, IV*. Montréal, Les Presses de l'Université de Montréal.

LAF : Mel'čuk I., Polguère A. (2007). *Lexique Actif du Français*. Bruxelles, De Boeck.

LTP : *Dictionary of selected collocations* (1997), Hill J., Lewis M. Hove, Language teaching publications LTP.

OCDSE : *Oxford Collocations Dictionary for students of English* (2002), Crowther, J. et al. Oxford, Oxford University Press.

References

Bartsch, S. (2004), *Structural and functional properties of collocations in English*, A corpus study of lexical and pragmatic constraints on lexical co-occurrence. Tübingen, Narr.

Bresnan et al.82] Bresnan (J.) et Kaplan (R.), *The mental representation of grammatical relations*, Cambridge, MA, The Mit Press, 1982.

Firth J.R. (1934-51), *Papers in linguistics*, Oxford University Press.

Gonzalez-Rey I. (2002), *La phraséologie du français.* Toulouse, Presses universitaires du Mirail.

Gross G. (1986), *Les expressions figées en français : noms composés et autres locutions*. Paris, Ophrys.

Halliday, M.A.K. , Hasan, R. (1976). Cohesion in English. London, Longman.

Hausmann F. J. (1989). "Le dictionnaire de collocations". In Hausmann F.J., Reichmann O., Wiegand H.E., Zgusta L. (eds), Wörterbücher : ein internationales Handbuch zur Lexicographie. Dictionaries. Dictionnaires. Berlin/New-York, De Gruyter, 1010-1019.

Hausmann F.J. (2007), "Lexicographie française et phraséologie", in *Collocations, phraséologie, lexicographie*. Aachen, Shaker Verlag, 121-153.

Hausmann, F. J. (2004), **"Was sind eigentlich Kollokationen?"** in Steyer, K. (Ed..) *Wortverbindungen - mehr oder weniger fest*, Institut für Deutsche Sprache Jahrbuch 2003, 2004, 309-334.

Heid U., Gouws R. H. (2006), "A model for a multifunctional electronic dictionary of collocations", *Proceedings of the XIIth Euralex International Congress*, Torino, 979 – 988.

Heid U. (1994), "On Ways Words Work Together - Topics in Lexical Combinatorics". In *EURALEX'94 Proceedings*. Vrije Universiteit Amsterdam. 226-257.

Mel'čuk I. (1998). "Collocations and Lexical Functions". In A. P. Cowie (ed.), *Phraseology. Theory, Analysis and Applications*. Oxford, Clarendon Press, 23-53.

Siepmann D. (2006). "Collocations et dictionnaires d'apprentissage onomasiologiques bilingues : questions aux théoriciens et pistes pour l'avenir", *Langue Française*, N° 107, 99-118.

Spohr D. (2005). *A Description Logic Approach to Modelling Collocations*. Diploma thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Germany.

Tutin A, Grossmann F. (2002). "Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif", *Revue française de Linguistique appliquée*, « Lexique : recherches actuelles », vol. VII-1. 7-25.

Williams G, (2003), "Les collocations et l'école contextualiste", in Grossmann F, Tutin A, *Collocations : analyse et traitement*, Travaux et recherches en linguistique appliquée, série E no1,éd de Werelt, 33-44.