

ANNA ZAGÓRSKA
(UNIWERSYTET WARSZAWSKI)
ORCID 0000-0003-3703-1364

DEFINING COLLOCATIONS FOR THE PURPOSES OF LSP LEXICOGRAPHY

ABSTRACT

A widely held view of linguistic researchers claims that collocations constitute a difficult and crucial aspect of vocabulary knowledge. The above-mentioned view is reflected in linguistic theory and practice. Collocations are justly devoted attention by linguists as well as all the individuals who aspire to fluently use a foreign language. Nevertheless, the term itself remains a subject of controversy, since linguistic literature provides various definitions of collocations. The differences between various definitions are related to their contexts as well as purposes for which they are formed. The present article accepts the context-dependence of various definitions of collocations and aims to propose a definition to be employed for the purposes of LSP lexicography.

KEYWORDS: collocation, dictionary, lexicography, LSP

STRESZCZENIE

Znajomość kolokacji danego języka jest niezbędna do posługiwania się nim na wysokim poziomie zaawansowania. Sam termin „kolokacja” pozostaje jednak przedmiotem kontrowersji, a w różnych pracach podaje się różne jego definicje. Rozbieżności między poszczególnymi definicjami wiążą się z kontekstami, w jakich są one formułowane oraz celami, dla których są tworzone. Pomimo problemów wynikających z opisywanego stanu rzeczy, w niniejszym artykule przyjmuje się, że wspomniane różnice stanowią konieczność. Celem artykułu jest zaproponowanie definicji kolokacji odpowiedniej w kontekście leksykografii języków specjalistycznych.

SŁOWA KLUCZOWE: kolokacja, słownik, leksykografia, język specjalistyczny

Various LSP dictionaries differ in their content and in the approaches to its description. Their authors adopt various theoretical principles at the stage of selecting, classifying and presenting the content of dictionary entries (Nuccorini 2003: 368; cf. Gledhill 2000: 7–18; Nesselhauf 2005: 12–13). The aim of the present article is to propose a definition of collocation to be employed for the purposes of a LSP lexicography. As Bowker and Pearson (2002: 25) succinctly explain:

LSP is the language that is used to discuss specialized fields of knowledge. It is actually more accurate to talk about LSP in the plural (i.e. languages for special purposes) since different LSPs are used to describe different areas of specialized knowledge.

As for LSP lexicography, in this work it is considered in the light of lexicographic practice as opposed to lexicographic theory. Accordingly, it can be defined as the complex of activities involved in the process of the design and compilation of LSP dictionaries (cf. Hartmann, James 1998: 129).

THE NOTION OF COLLOCATION

The notion of collocation is defined and understood in a number of ways. Some definitions of the term are quoted below:

The term **collocation** will be used to refer to sequences of lexical items which habitually co-occur, but which are nonetheless fully transparent in the sense that each lexical constituent is also a semantic constituent. These are of course easy to distinguish from idioms; nonetheless, they do have a kind of semantic cohesion – the constituent elements are, to varying degrees, mutually selective (Cruse 1986: 40).

Collocations (...) are fixed, recurrent combinations of words in which each word basically retains its meaning (Benson 1990: 85).

Collocation is a term used to describe a group of words which occur repeatedly in a language (Carter 1994: 47).

A collocation is a predictable combination of words: *get lost, make up for lost time, speak your mind*. Some combinations may be very highly predictable from one of the component words – *foot the bill, mineral water, spring to mind*. Some ‘strong’ collocations have the status of idioms – *shrug your shoulders* – they are not guessable and are non-generative. Some may be so common that they hardly seem worth remarking upon – *a big car, a nice flat, have lunch* (Hill 2000: 51).

Collocation is the way in which words co-occur in natural texts in statistically significant ways. It sounds an innocent definition, but one very important point needs to be made: collocation is about the way words naturally co-occur in what David Brazil brilliantly called ‘used language’ (Lewis 2000: 132).

Collocations exist in every language. The term itself denotes word combinations, consisting of two or more words, that repeatedly co-occur (Osuchowska 2001: 7).

What is a collocation? It can be said to be a set of two or more words that frequently occur in juxtaposition, and that seem to ‘fit together’ (Douglas-Kozłowska 2004: 9).

As employed in this chapter, the term **collocation** refers to combinations of two lexical items each of which makes a distinct semantic contribution, belongs to a different word class and shows a restricted range (Gramley, Pätzold 2004: 51).

Collocations are recurrent co-occurrences of words in texts. They certainly are statistically significant; but this is not enough. They also have to be semantically relevant. They have to have a meaning of their own, a meaning that isn't obvious from the meaning of the parts they are composed of (Teubert 2004: 188).

Collocation is the co-occurrence of two items in a text within a specified environment (Sinclair *et al.* 2005: 10).

Le collocazioni sono espressioni formate da due o più parole che per uso e consuetudine lessicale formano una unità fraseologica non fissa ma riconoscibile (Tiberii 2012: 3).

Some of the definitions are quite elaborate and detailed (e.g. Cruse 1986; Hill 2000), whereas others, especially those included in dictionaries, as opposed to theoretical linguistic literature, are relatively simple and general (e.g. Douglas-Kozłowska 2004; Osuchowska 2001; Tiberii 2012). Besides, some of them may be perceived as relatively narrow (Cruse 1986: 40; Teubert 2004) whereas others are quite broad (Carter 1994; Hill 2000: 50–51). Even those relatively complex definitions give the impression that the notion of collocation deserves a more detailed explanation (cf. Meer 1998; Fontenelle 1994). In fact, after formulating a definition of collocations, numerous researchers provide further information on the notion. The additional information is often absolutely crucial. It can even be assumed that in order to properly describe the notion of collocation, one needs to go beyond its definition, since the notion under discussion cannot be accurately described using just a few words or sentences. Probably for this reason numerous researchers who devote their attention to collocations resign from formulating any definition of the term and instead provide a detailed description of the concept.

Furthermore, while examining the definitions quoted above, it is not difficult to notice that there are several researchers who perceive collocations as word combinations which are recurrent or which occur repeatedly or frequently (Carter 1994; Osuchowska 2001; Douglas-Kozłowska 2004). Nevertheless, Nation (2001: 324) notices that collocations cannot be satisfactorily defined solely in this way, as corpus-based frequency research returns many combinations such as *of the*, *although he* or *but if*. The combinations in question do have a high frequency of occurrence, nonetheless intuition suggests that they are not collocations. According to Nation (*ibidem*), the expressions which can be termed as collocations are “closely structured” and demonstrate “some element of grammatical or lexical unpredictability or inflexibility”. In addition, a number of researchers assume that collocations are statistically significant (Lewis 2000; Sag *et al.* 2002; Teubert 2004) or mutually selective word combinations (Cruse 1986; cf. Herbst 1996: 383).

It should be highlighted that some authors may decide to employ relatively simple definitions of collocations for specific reasons. For the sake of example, the definitions by Osuchowska (2001), Douglas-Kozłowska (2004) and Tiberii (2012) appear in collocational dictionaries which are presumably intended most of all for

non-native users of English and Italian. It is probable that a significant part of dictionary users could find more complex definitions confusing. In fact, it can be assumed that all that they need is only a general and basic understanding of what collocations are. Besides, the seemingly very simple definitions of collocations are often followed by a more detailed description of the notion. For instance, Sinclair *et al.* (2005: 9) immediately after their definition propose a distinction between significant collocations, whose component parts “co-occur more often than their respective frequencies, and the length of text in which they appear, would predict”, and casual collocations, which are non-significant collocations.

Furthermore, various definitions of collocations share the tendency to refer to lexical items. In this work we assume the definition of lexical item proposed by Sinclair *et al.* (2005: 9), who postulate that “a lexical item is a unit of language representing a particular area of meaning which has a unique pattern of co-occurrence with other lexical items”. Some lexical items can be identified with orthographic words. Others correspond to various senses of a polysemous word. Others again take the form of a group of words associated paradigmatically (e.g. kick, kicked, kicking, kicks) or syntagmatically (e.g. *to beat about the bush*). Lexical items are sometimes juxtaposed with grammatical items, i.e. units of language whose presence in a given text is related to their grammatical function and not their specific meaning (*ibidem*). Besides, they are sometimes referred to as lexemes, though *lexeme* may also be understood as an uninflected word form (Carter 1994: 22; cf. Aprile 2005: 9–10).

Some authors assume that collocations may involve lexical items as well as grammatical items (Benson 1989; Carter 1994: 47–48). Accordingly, the researchers in question divide collocations into grammatical and lexical. Other authors, however, present collocations as combinations of lexical items rather than those grammatical (Cruse 1986; Hill 2000; Gramley, Pätzold 2004). Nevertheless, sometimes their examples of collocations do include some grammatical words, such as prepositions (e.g. *completely obsessed with, moved to tears*, Hill 2000). Finally, it should be mentioned that collocation is sometimes juxtaposed with colligation. In such cases the former pertains to “constrained lexical choice” (Bartsch 2004: 31), whereas the latter refers to what is elsewhere called grammatical collocation (e.g. *argument for, discrimination against*, Flowerdew 2009: 87).

Moreover, various works devoted to collocations differ in the approach adopted with respect to idioms and free combinations. Let us firstly focus on the former. In this work, idioms will be defined as multi-word items with holistic meanings which are not retrievable from the meanings of their component elements (e.g. *to kick the bucket*, or Italian *darsi delle arie*, Moon 1997: 46). Hill (2000: 50–51) and Geller and Dąbrówka (2007: XIV) state that idioms constitute a kind of collocation (cf. Palmer 1976: 98; Moon 1997: 43), whereas Cruse (1986: 40), Benson (1989), Jędrzejko (1998: 74), Crowther *et al.* (2002: vii) and Rokicka (2007: 81) claim that they are a different type of word combination. Besides, Moon (1997) takes an intermediate approach assuming that multi-word items, such as idioms, phrasal

verbs, compounds, fixed phrases and prefabs, “in many respects can be seen as extreme cases of fixed collocations” (*ibidem*: 43). Finally, Palmer (1976: 96) criticizes clear distinctions between collocations which are predictable from the meanings of their components and those which are unpredictable. The researcher postulates that the notion of collocation should not be restricted in any precise way, though understandably every researcher is allowed to limit their research to any items that they find interesting.

Let us now turn to the so-called free combinations, which are also termed as occasional or casual collocations (Roos 1976; Sinclair 1966: 418; Sinclair *et al.* 2005: 10). They are expressions of the smallest degree of restrictedness whose components combine with vast numbers of items, sometimes coming from different grammatical categories. For instance, the verb *to see* may be accompanied by a variety of nouns (for instance *to see a car, a cat, a teacher, a flower etc*), prepositional phrases (*to see somebody on the street, in the park*) and adverbs (*to suddenly see, to see today*) (Roos 1976). Some linguists like Crowther *et al.* (2002: vii), Proost (2007: 165), Hausmann and Blumenthal (2006: 4) and McKeown and Radev (2000) distinguish between collocations and free combinations. Contrastingly, Howarth (1996: 33–34) assumes that free combinations should be perceived as a kind of collocation and Hill (2001: 51) provides examples of collocations which in fact constitute free combinations (*a nice flat, a big car*).

Furthermore, some researchers who notice that the notion of collocation is problematic in terms of its definition approach the term in a non-conventional way. To give an example, Meer (1998: 315) assumes that it is not possible to formulate a clear-cut definition of collocation, as the notion does not feature clear-cut distinctions in natural language. Thus, the researcher proposes to formulate a definition of the prototypical collocation (cf. Herbst 1996: 385). He tentatively defines the prototypical collocation as a combination:

1. of two or more lexical units, with meanings also occurring independently in other combinations;
2. of lexical units which are used non-metaphorically;
3. which appears normally, repeatedly and conventionally in a language;
4. which is available as a whole to the language user and serves to express conventional, established concepts;
5. whose constituents are in a modifier-modified relation;
6. whose constituent words naturally select each other since the sense definition of the modifier includes the modified (and sometimes vice versa) in a non-banal way (semantic motivation);
7. which typically functions as part of a larger group and not as a complete utterance

Meer (1998: 316) points out that the above-mentioned criteria eliminate a number of fixed expressions which are not prototypical collocations, such as

proverbs, idioms or catchphrases. Besides, the researcher distinguishes collocations from free combinations, assuming that the former convey conventional concepts.

TYPOLOGIES OF COLLOCATIONS

The variety of word combinations which may be termed collocations has generated the need of divisions within the notion. A number of researchers have attempted to classify collocations according to various criteria or have identified different types of collocations. This section aims at presenting the divisions that seem the most useful. Their review will provide further insight into the nature of collocations.

First and foremost, collocations are often divided into lexical and grammatical (Benson 1989; Carter 1994). The former type of collocation may consist, for instance, of a noun and an adjective (for example the English *high probability*, Italian *atto illecito*) or a verb and a noun (for instance English *to mount resistance*, Polish *tryskać radością*) (Benson 1989: 88). As far as grammatical collocations are concerned, they consist of a dominant word and a grammatical word, which is usually a preposition. The dominant word may be a noun (for instance *ambassador to*, *counselor to*, or Polish *informacja o*), a verb (for example *to abstain from*, *to hint at*, Italian *partire per*, *riuscire a*) or an adjective (e.g. *typical of*, *adverse to*). Some linguists assume that grammatical collocations also include certain syntactic structures which are typical of a given item. The structures in question may involve *that*-clauses (which, for instance, typically follow the adjective *adamant*) or *-ing* forms (e.g. required after the verb *to avoid*) (Fontenelle 1994).

So much for the distinction between grammatical and lexical collocations. Let us now turn to two independent divisions of collocations which are proposed by McCarthy (1994). Firstly, the researcher divides collocations into strong and weak (McCarthy 1994: 12). The division in question hinges upon the collocational range of lexical items which form a given collocation. A strong collocation is, for instance, the expression *blond hair*, since the adjective *blond* does not collocate with any noun except for *hair*. Contrastingly, the expression *brown hair* is an example of weak collocation, since both *brown* and *hair* collocate with many other lexical items (cf. Italian *naso camuso* and *naso dritto*). Secondly, McCarthy (1994: 12–13) divides collocations into marked and unmarked. The division is based on the typicality or atypicality of a collocation. For the sake of example, a very typical collocation, such as *a major problem*, is an unmarked collocation. In contrast, quite an original expression, such as *major amount*, is an example of marked collocation. Very marked collocations, such as *a major man*, are in most cases considered to be language errors.

Furthermore, according to Carter (1994), collocations can be divided according to their degrees of restrictedness. The researcher postulates their division into

unrestricted, semi-restricted, familiar and restricted. Unrestricted collocations are typical of the simplest and most rudimentary words, such as the verbs *to have*, *to make*, *to take* or *to run*, which collocate with a considerable number of lexical items. As for semi-restricted collocations, they differ from those unrestricted in that in their case “the number of items which can be substituted in different syntactic roles is more determined”. As an example, Carter (1994: 63) points to the collocations of the verb *to harbour*, such as *to harbour a doubt*, *to harbour uncertainty* or *to harbour suspicion*. Contrastingly, familiar collocations consist of words which typically accompany each other, for instance *an innocent bystander*, *a lukewarm reception*, *an amicable divorce* or Italian *pioggia torrenziale*. Last but not least, the category of restricted collocations encompasses fixed expressions, such as *pitch black* or *pretty sure*. Besides, it includes idioms (e.g. English *to beat about the bush*, Polish *spalić za sobą mosty*) and so-called irreversible binomials, such as *swings and roundabouts*, *ups and downs* or Italian *sano e salvo* (cf. Benson 1989).

Carter’s classification of collocations seems similar to the one proposed by Hill (2000: 63–64), who nevertheless does not refer to restrictedness, but to collocational strength. Firstly, Hill (2000) distinguishes unique collocations, such as *to foot the bill* or *to shrug one’s shoulders*. The researcher points out that the verb *to shrug* does not collocate with other body parts, whereas the verb *to foot* does not collocate, for instance, with *invoice* or *coffee* (cf. Italian *digrignare i denti*). The same type of collocation is termed by Cruse (1986: 41) “bound collocation”. Secondly, Hill (2000: 63) refers to strong collocations. While describing them, he notices that “predictably, we may talk of *trenchant criticism* or *rancid butter*, although this does not mean that other things cannot be *trenchant* or *rancid*”. Though the collocations in question are not unique, the knowledge of the adjectives *trenchant* or *rancid* certainly would not be complete without the knowledge of their most common collocates (cf. the Polish collocation *dosiadać konia*). As for the third type of collocations mentioned by Hill (2000), they are weak collocations, which are created, for instance, by the adjectives *long*, *cheap*, *good*, *bad* etc. Finally, the researcher identifies medium-strength collocations, which are in between those strong and weak. Some example medium-strength collocations include the English expressions *to hold a conversation* or *to recover from a major operation*, Polish *wykonywać pracę* or Italian *raccogliere fiori*.

Let us now turn to a more complex division of collocations, which is proposed by Nation (2001). The researcher presents a set of ten scales for classifying items as collocations and for distinguishing their categories. The scales under discussion, which were created based on previous research (e.g. Kennedy 1998; Kjellmer 1982; Renouf, Sinclair 1991; Sinclair 1991, all quoted in Nation 2001: 329–331), are presented in the table below (see *Table 1*). The column on the left shows the features of word combinations to be assessed on scales between two contrasting characteristics (e.g. *semantically opaque* and *semantically transparent*, *grammatically regular* and *grammatically unique* etc). As for the column on the right, it contains explanations and sample collocations. According to Nation (2001: 332),

it is possible to use only some of the scales under discussion or prioritise them depending on the purpose for which they are used.

Table 1. A set of ten scales for classifying items as collocations and for distinguishing their categories as proposed by Nation (2001)

The feature	Description
The frequency of the occurrence of a pair or group of items	The scale goes from frequently co-occurring items to those co-occurring infrequently. It should be mentioned, however, that according to Nation (2001), the criterion of co-occurrence entails numerous questions. It is not clear whether the co-occurring items need to be adjacent, whether they can be inflected or whether they have to show strong grammatical relationships and whether frequent word pairs such as <i>and the</i> should be taken into account. Additionally, collocation frequency needs to be considered alongside collocational range, as some items appear in a limited number of text types.
The adjacency of a pair or group of items	The scale goes from items which co-occur immediately next to each other to those which are separated by one or more words.
Grammatical connections	The scale goes from items which are grammatically connected to those unconnected. The collocating items usually show grammatical connections. Nevertheless, sometimes collocational status is granted to items which are grammatically unconnected but which stand in a “lexical cohesion relationship” (Nation 2001: 330).
Grammatical structure	Some collocations can be classified as ‘well-structured’ as opposed to ‘loosely related’. The latter, such as <i>of the</i> or <i>although he</i> , show some grammatical connections but do not reflect divisions that would take place while analysing the clauses.
Grammatical uniqueness	Some collocations show grammatical uniqueness (e.g. <i>hell for leather</i>). Others constitute an exception to a rule (e.g. <i>to go to bed</i> , which does not require the article). Still others are grammatically regular.
Grammatical fossilisation	If a given collocation is grammatically fossilised, it allows no grammatical change nor a change in word order (e.g. <i>law and order</i> , <i>by and large</i>). Nevertheless, there are also collocations which allow a variety of changes, such as passivisation, inflectional change or change in parts of speech. In between the extremes Nation (2001) locates collocations which allow for minor changes.

Table 1. cont.

The feature	Description
Collocational specialisation	Some items occur only in one particular collocation (e.g. <i>kith</i> in <i>kith and kin</i> or <i>bubonic</i> in <i>bubonic plague</i>), whereas others have a significant number of possible collocates. The scale goes from items which always mutually co-occur to those which all appear in numerous collocations. In between the extremes there are collocations with one bound item.
Lexical fossilisation	The scale ranges from collocations in which neither of the collocates can be replaced by their synonyms or words of related meaning (e.g. <i>a bird's eye view</i> , <i>no fear!</i>) to collocations which allow the replacements under discussion for all of the collocates.
Semantic opaqueness	The scale goes from collocations which are semantically opaque, i.e. whose meaning cannot be deduced from the meaning of their parts (e.g. <i>to burn your bridges</i>), to collocations which are semantically transparent.
Uniqueness of meaning	Some collocations have only one meaning, whereas others have multiple meanings.

The ranges in each of the scales presented by Nation (2001: 329–332) are ordered from the most to least lexicalised. According to the researcher, an example of a highly lexicalised collocation is *hocus pocus*. The collocation is frequent, grammatically unique and has only one meaning which cannot be deduced from its parts. The structure of the collocation under discussion allows neither grammatical nor lexical changes. The two collocates usually co-occur immediately next to each other, show grammatical connection and rarely form other collocations.

Furthermore, taking a completely different perspective, Lewis (2000: 133–134; cf. Białek 2009: 22–26) provides a list of various types of collocations based on the parts of speech and structures of their collocates:

1. *A difficult decision* (adjective + noun)
2. *To submit a report* (verb + noun)
3. *Radio station* (noun + noun)
4. *To examine thoroughly* (verb + adverb)
5. *Extremely inconvenient* (adverb + adjective)
6. *To revise the original plan* (verb + adjective + noun)
7. *The fog closed in* (noun + verb)
8. *To put it another way* (discourse marker)
9. *A few years ago* (multi-word prepositional phrase)
10. *To turn in* (phrasal verb)
11. *Aware of* (adjective + preposition)

12. *Fire escape* (compound noun)
13. *Backwards and forwards* (binomial)
14. *Hook, line and sinker* (trinomial)
15. *On the other hand* (fixed phrase)
16. *A sort of...* (incomplete fixed phrase)
17. *Not half!* (fixed expression)
18. *See you later / tomorrow / on Monday* (semi-fixed expression)
19. *Too many cooks...* (part of a proverb)
20. *To be or not to be...* (part of a quotation).

Still another view on collocations is reflected in Spohr's (2012: 71–73) taxonomy. The researcher divides collocations into simple and complex. The former may consist of, for instance, morphologically simple units, derivations, compounds etc. As for complex collocations, they contain at least one other collocation (cf. Białek 2009: 24). Their further subdivision is presented schematically in Figure 1:

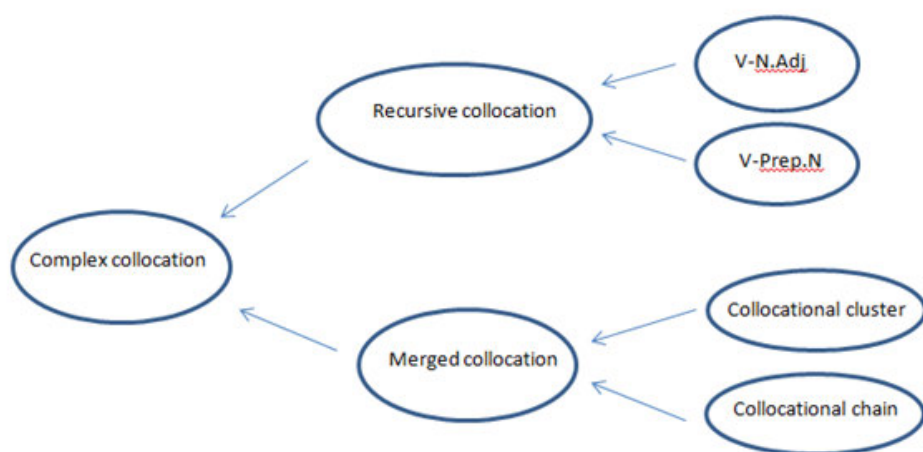


Figure 1. Subdivision of complex collocations proposed by Spohr (2012: 72)

Complex collocations are divided into merged collocations and recursive collocations. There are two types of merged collocations, which are referred to as collocational clusters and collocational chains. Collocational clusters share the same node. To give an example, *pay close attention* merges *pay attention* with *close attention*. In both of the collocations the noun *attention* constitutes the node as opposed to collocate (cf. the Italian collocation *prestare un'attenzione particolare*). Contrastingly, in a collocational chain, the node of one of the collocations constitutes the collocate of the other one. For instance, the collocation *strongly recommended book* merges *strongly recommended*, where *recommended* constitutes the node and *recommended book*, where *recommended* is the collocate. Finally, in

recursive collocations, one of the collocations which are merged constitutes a collocate or a node itself. For instance, in the collocation *to fall in love*, *fall* is a collocate, whereas the collocation *in love* constitutes its node.

The review of various types of collocation presented in this work is far from being exhaustive. As examples of some other typologies, Sinclair (1991: 115–116) distinguishes between upward, downward and neutral collocations. As for downward collocations, their nodes are used in a language more frequently than their collocates. Conversely, upward collocations consist of less frequently used nodes and more frequent collocates. Finally, neutral collocations constitute a “buffer area” between the two above-mentioned types of collocations (cf. Handl 2008: 60–61). Furthermore, Fontenelle (1994) refers to delexical collocations, which are a type of restricted collocations consisting of a grammaticalized verb (such as *to have*, *to make*, *to do* or *to get*) and its direct object (cf. Jędrzejko 1998: 17–21; Nesselhauf 2005: 20; Vetulani 2005: 153–155). The latter (2005: 155–160) devotes attention to further subdivisions of the collocations under discussion. Finally, Sinclair *et al.* (2005) distinguish between significant collocations, whose component parts appear together more often than their respective frequencies would predict, and casual collocations, which are non-significant collocations.

Beyond any doubt, differences between various typologies of collocations reflect different assumptions on what constitutes a collocation. They also demonstrate that collocations tend to be considered from various perspectives. Although some typologies of collocations seem to be similar in their content (e.g. McCarthy 1994; Hill 2000; Carter 1994), they differ in that some of them are more detailed (Hill 2000), whereas others are more general (McCarthy 1994). Additionally, given the collocations which are provided as examples of particular categories, it can be observed that the authors of the classifications seem to set different boundaries between various collocational types. In general, however, none of the presented typologies, except the one proposed by Lewis (2000), provides clear-cut boundaries between different types of collocations. Accordingly, the typologies in question correspond to the recommendation by Bartsch (2004: 33), who postulates that the dispute over the status of various multi-word expressions is best resolved by their classification “along a cline within a continuum without clearly definable borderlines” (cf. Carter 1994: 63).

DEFINING COLLOCATIONS FOR THE PURPOSES OF LSP LEXICOGRAPHY

Now it is time to consider the notion of collocation focusing on the context, i.e. the LSP lexicography. At the outset let us focus on the approach to idioms that should be recommended in the LSP dictionary making process. Most of all, it should be pointed out that, in general, idioms are not typical of LSPs. Besides, there is no

reason to assume that they should be treated as collocations since due to their semantic opaqueness they usually require a lexicographic approach different from the approach to semantically predictable word combinations. Nevertheless, the inclusion of idioms in an LSP dictionary might be certainly very useful if they are commonly employed in a particular LSP. Still, when it comes to the general definition of collocation formed for the purposes of LSP lexicography, the inclusion of idioms does not seem necessary.

As far as free combinations are concerned, their inclusion also seems superfluous, since most users of LSP dictionaries are advanced enough to deal with them with the aid of a traditional bilingual dictionary. As for the unadvanced users, it is not probable that the inclusion of free combinations is likely to resolve their linguistic problems. Besides, it could also lead to a significant increase in dictionary size, which in some cases constitutes a problem. Finally, it should be mentioned that in general, free combinations are believed not to pose serious problems neither to language learners nor to translators (Hatch, Brown 1995). All in all, taking a different perspective, it can be postulated that an LSP dictionary should contain unique collocations, strong collocations and medium-strength collocations as understood by Hill (2000: 63–64).

All things considered, numerous LSP dictionaries contain neither idioms nor free combinations and the approach is understandable. Contrastingly, what seems important for the definition of collocation in an LSP dictionary is statistical significance, habituality of occurrence in natural language and semantic cohesion. Besides, an LSP dictionary should contain both lexical and grammatical collocations and, preferably, it should not be limited to collocations of only two words or collocations which are simple as opposed to complex ones (cf. Spohr 2012). Finally, collocations included in an LSP dictionary should not be required to be semantically unpredictable. The lack of semantic unpredictability does not seem to be a good reason to exclude any statistically significant and conventionally used collocation from the dictionary of the type.

In conclusion, for the purposes of LSP lexicography, the term *collocation* may be briefly defined as a combination of two or more words that:

- is statistically significant,
- is semantically coherent and semantically transparent,
- occurs habitually in natural language and
- may involve lexical as well as grammatical items.

It should be mentioned, however, that both the inclusion of particular types of collocations and the complexity of collocations included are dependent upon various factors, such as the available space and time span or the target user group of a particular dictionary etc. Besides, it should be pointed out that the notion of collocation is vague and is defined in different ways depending on the context of a given definition. A definition which is found in an LSP dictionary may differ from

a definition in an L2 learner's dictionary or a definition concluding a linguistic article on what should be understood as a collocation. Given the variety of mutually exclusive assumptions on what constitutes a collocation, it seems impossible to formulate its definition which would suit all purposes. Even the definition of prototypical collocation proposed by Meer (1998) cannot be perceived as universally applicable. To give an example, it is not sufficient for the process of compilation of a collocational dictionary, as the above-mentioned process requires establishing what should be included in the dictionary and what should be omitted. Finally, it can be postulated that in order to properly describe the notion of collocation, one needs to go beyond its definition, since the notion under discussion cannot be accurately described in just few words or sentences.

REFERENCES

- APRILE M. (2005): *Dalle parole ai dizionari*, Il Mulino, Bologna.
- BARTSCH S. (2004): *Structural and functional properties of collocations in English*, Gunter Narr, Tübingen.
- BENSON M. (1990): *The Collocational Dictionary and the Advanced Learner*, in: TICKOO M.L. (ed.), *Learners' Dictionaries: State of the Art. Anthology Series 23*, Regional Language Centre SAMEO, Singapore: 84–93.
- BIAŁEK E. (2009): *Kolokacja w przekładzie. Studium rosyjsko-polskie*, Wydawnictwo UMCS, Lublin.
- BOWKER L., PEARSON J. (2002): *Working with specialized language: a practical guide to using corpora*, Routledge, London and New York.
- CARTER R. (1994): *Vocabulary: Applied Linguistics Perspective*, Routledge, London and New York.
- CROWTHER J. et al. (2002) (eds.): *Oxford Collocations Dictionary for Students of English*, Oxford University Press, Oxford.
- CRUSE D.A. (1986): *Lexical Semantics*, Cambridge University Press, Cambridge.
- DOUGLAS-KOZŁOWSKA CH. (2004): *English Adverbial Collocations*, PWN, Warszawa.
- FIRTH J.R. (1957): *Papers in linguistics 1934–1957*, Oxford University Press, London.
- FLOWERDEW J. (2009): *Use of signaling nouns in a learner corpus*, in: FLOWERDEW J., MAHLBERG M. (eds.), *Lexical cohesion and corpus linguistics*, John Benjamins B.V., Amsterdam, Philadelphia: 85–102.
- FONTENELLE T. (1994): *What on Earth are Collocations?*, "English Today", 40/10 (4): 42–48.
- GELLER E., DĄBRÓWKA A. (2007): *Słownik stylistyczny języka polskiego*, Świat Książki, Warszawa.
- GLEDHILL C. (2000): *Collocations in Science Writing*, Gunter Narr, Tübingen.
- GRAMLEY S., PÄTZOLD K.M. (2004): *A Survey of Modern English*, Routledge, New York.
- HANDL S. (2008): *Essential collocations for learners of English. The role of collocational direction and weight*, in: MEUNIER F., GRANGER S. (eds.), *Phraseology in Foreign Language Learning and Teaching*, John Benjamins B.V., Amsterdam, Philadelphia: 44–66.
- HARTMANN R., JAMES, G. (1998): *Dictionary of Lexicography*, Routledge, London.
- HATCH E., BROWN CH. (1995): *Vocabulary, Semantics and Language Education*, Cambridge University Press, Cambridge and New York.
- HAUSMANN F.J., BLUMENTHAL P. (2006): *Présentation: collocations, corpus, dictionnaires*, "Langue française", 2 (150): 3–13.

- HERBST T. (1996): *What are collocations: sandy beaches or false teeth?*, "English Studies", 77 (4): 379–393.
- HILL J. (2000): *Revising priorities: from grammatical failure to collocational success*, in: LEWIS M. (ed.), *Teaching Collocation. Further Developments in the Lexical Approach*, LTP Teacher Training, Hove: 47–69.
- HOWARTH P.A. (1996): *Phraseology in English academic writing: some implications for language learning and dictionary making*, Niemeyer, Tübingen.
- JĘDRZEJKO E. (1998) (ed.): *Słownik polskich zwrotów werbo-nominalnych*, Energeia, Warszawa.
- MCCARTHY M. (1994): *Vocabulary*, Oxford University Press, Oxford.
- MCKEOW R., RADEV D.R. (2000): *Collocations*, in: DALE R., MOISL H., SOMERS H. (eds.), *Handbook of natural language processing*, Marcel Dekker, Basel, New York: 507–523.
- MEER VAN DER G. (1998): *Collocations as one particular type of conventional word combinations. Their definition and character*, in: FONTENELLE T., HILIGSMANN P., MICHIELS A., MOULIN A., THEISSEN S. (eds.), *Actes Euralex '98*, Université de Liège, Liège: 313–322.
- MOON R. (1997): *Vocabulary connections: multi-word items in English*, in: SCHMITT N., MCCARTHY M. (eds.), *Vocabulary: Description, Acquisition and Pedagogy*, Cambridge University Press, Cambridge: 40–63.
- NATION I.S.P. (2001): *Learning Vocabulary in Another Language*, Cambridge University Press, Cambridge.
- NESSELHAUF N. (2005): *Collocations in a learner corpus*, John Benjamins B. V., Amsterdam, Philadelphia.
- NUCCORINI S. (2003): *Towards an 'ideal Dictionary of English Collocations*, in: VAN STERKENBURG P., (ed.), *A practical Guide to Lexicography*, John Benjamins B.V., Amsterdam, Philadelphia: 366–388.
- OSUCHOWSKA D. (2001): *English at Work. An English-Polish Dictionary of Selected Collocations*, Wydawnictwo Naukowe PWN, Warszawa.
- PALMER F.R. (1976): *Semantics. A new outline*, Cambridge University Press, Cambridge.
- PROOST K. (2007): *Conceptual structure in lexical items*, John Benjamins B.V., Amsterdam, Philadelphia.
- ROKICKA T. (2007): *Status kolokacji w Wielkim słowniku języka polskiego*, in: ŻMIGRODZKI P., PRZYBYLSKA R. (eds.), *Nowe studia leksykograficzne*, LEXIS, Kraków: 79–87.
- ROOS E. (1976): *Contrastive collocational analysis*, "Papers and Studies in Contrastive Linguistics", 5: 65–75.
- SAG I., BALDWIN T., BOND F., COPESTAKE A., FLICKINGER D. (2002): *Multiword expressions: a pain in the neck for NLP*, in: IVAN A., (ed.), *Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing*, Springer, London: 1–15.
- SINCLAIR J.McH. (1966): *Beginning the study of lexis*, in: BAZELL C.E., CATFORD J.C., HALLIDAY M. A.K., ROBINS R.H. (eds.), *In memory of J.R. Firth*, Longman, London: 410–430.
- Id. (1991): *Corpus, concordance, collocation*, Oxford University Press, Oxford.
- SINCLAIR J.McH., JONES S., DALEY R. (2005): *English Collocation Studies: The OSTI Report*, Continuum, London.
- SPOHR D. (2012): *Towards a multifunctional lexical resource: design and implementation of a graph-based lexicon model*, Walter de Gruyter, New York.
- TEUBERT W. (2001): *Corpus Linguistics and Lexicography*, "International Journal of Corpus Linguistics", 6: 125–153.
- TIBERII S. (2012): *Dizionario delle collocazioni*, Zanichelli, Bologna.
- VETULANI G. (2005): *Répercussions des activités humaines dans les collocations verbo-nominales*, "Neophilologica", 17: 153–161.