Armine Garibyan

Interpreting collocations: An analysis of collocation dictionaries of: English, German and Russian

Abstract: Unless online collocation tools have not become common yet, printed collocation dictionaries are still one of the few sources of checking collocative possibilities of a lexical unit. The aim of this paper is to analyse one printed collocation dictionary in each of the 3 languages: English, German and Russian and see which interpretation of collocation was focused on and what user groups were targeted. Also, it was attempted to see whether the morphological specificity of the languages was taken consideration of as compared to the structure of these dictionaries, and further determine whether inclusion or exclusion of certain language-specific information aims at a particular group of target users.

Keywords: collocation, dictionary, language-specificity, target user.

1 Limiting the scope of collocation dictionaries

1.1 Expressing the need for collocation dictionaries

Although most people would not know what a collocation dictionary is, nowadays there is hardly any linguist who would deny the necessity of collocation dictionaries. This necessity can be explained by many target groups of users who might need such a product. And whereas it is for some reason conventional that it is language learners

Armine Garibyan, Friedrich-Alexander-Universität Erlangen-Nürnberg, Bismarckstr. 1, 91054 Erlangen, e-Mail: armine.garibyan@fau.de

https://doi.org/10.1515/lex-2019-0007

who find themselves in search for the necessary collocate of a certain lexical unit, the user groups are certainly not limited to them. Here are several scenarios which in perfect conditions would inevitably lead to the use of a collocation dictionary:

1. Foreign language learners could experience negative transfer between their mother tongue and the second foreign language (L2), e. g. a Russian learner of English or German could mistakenly translate *жидкий суп* as *liquid soup** or *liquide Suppe** whereas the correct alternatives are *thin soup* and *dünne Suppe*.
2. A non-native speaker teacher of English might become unsure while grading some of their students' written assignments, e. g. while deciding whether *perform a mistake* is as 'correct' as *make a mistake*, or it should be marked as an error.
3. A native speaker of English or an advanced language learner may not be familiar with collocations belonging to a particular filed, e. g. *lodge a complaint, place a complaint* or *formulate a complaint* when writing a course paper in English.
4. A linguist may carry out research on the differences between British and American collocations.

There is a vast body of research reporting challenges in collocation use or competence which targeted different groups of participants: foreign language learners (Greenbaum 1970; Herbst 1996; Jiang & Nekrasova 2007; Yamashita & Jiang 2010; Wolter & Gyllstad 2013; Matsuno 2017; Choi 2017), native speakers (Herbst 1996; Wolter & Gyllstad 2013; Dabrowska 2014; Matsuno 2017), etc. Also, although non-native teachers are not usually the target participants of research on collocational competence, some evidence on challenges being experienced by them can be found, for example in the methodological section of a study by Chen (2011: 65) where while analyzing students' data by a non-native teacher, it was stated that

> when the teacher was not sure about the correctness of certain collocations provided by students, she consulted a native speaker. When both of them were not sure about a certain answer, the teacher used the *Oxford collocations dictionary for students of English* ...

However, a collocation dictionary could well be the first choice given the number of studies mentioned earlier according to which native speakers are also subject to unawareness or doubts regarding collocations.

1.2 Defining collocations

So, what does one want to find in a perfect collocation dictionary? The answer to this question will imminently lead to the various existing interpretations of this phenomenon. Following Herbst (1996), three interpretations are at least possible: 1. text-oriented approach in which collocations are seen as a purely text linguistics phenomenon (Halliday & Hasan 1976); 2. statistically-oriented approach (Palmer 1968; Sin-

clair 1991); and 3. significance-oriented approach (Cowie 1981; Hausmann 1984). Yet, whereas the approaches being outlined may give a naïve impression of interpretation possibilities which lie far apart from each other, in reality, fuzzy boundaries can be seen.

According to the text-oriented approach, one can find a marked cohesive effect in a pair of words occurring in close proximity with each other (Halliday & Hasan 1976). The following combinations could thus be classified as collocations in this respect: *work... computer, drink... cup, book... thick, plane... fly, tailor... suit*, etc. Given that such combinations can be called collocations, it seems that according to the text-oriented approach, constituents of a collocation can be interrupted by other units, which means they co-occur in a given text and in a certain – rather broad – span but are not necessarily adjacent. So, although there is no doubt that combinations of this kind stand in lexicosemantic relations with each other, the reasons for their co-occurrences, as Herbst claims (1996), could be better explained by extralinguistic factors rather than linguistic ones. Thus, in a cookery book, one is likely to find lexical units *knife, bake, onion, oven, tomato, sieve*, etc. in close proximity with each other only because they are referring to extralinguistic objects involved in the act of cooking.

Then, collocations in the sense of the frequency-based approach are most closely associated with Sinclair's (Sinclair 1991: 170) contribution to this notion. He defines a collocation as the occurrence of two or more words within a short space of each other in a text (Sinclair (1991): 170). Examples of this type of collocations are *as a result, last year, good morning*, etc. Then, a more phraseological view of collocations was suggested by Hausmann (1984) who characterises collocations as 'typical, specific and characteristic relations between two words' and calls them 'semi-prefabricated products of the language' (translation mine) (Hausmann 1984: 398–399). Typical examples are *commit a crime, false teeth, painfully shy, cry bitterly*, etc. So, these combinations can be rare, but they are mutually expectant. For example, *painfully* occurs in the BNC (*British National Corpus*) 665 times, whereas *shy* occurs 1053. However, the word combination *painfully shy* returns 8 hits, which proves that there is something more than pure frequency, i. e. specific lexical relations between two lexical units, which explains why combinations of this type are called collocations.

Nonetheless, while it is less challenging to define an appropriate approach for collocational interpretation and then come up with matching examples, it is more so working the opposite way. For instance, it is not so straightforward to determine whether a combination like *make a mistake* is a statistically significant or a phraseological one. On the one hand, it is relatively frequent because it occurs 1555[1] times in the BNC, but on the other, certain lexical relations hold between the two content words since it is not possible to say *do a mistake*, the latter being a good predictor for

1 The following query was used for the search: make/V ** mistake/N to allow for an optional determinative and/or premodifier.

the phraseological type of collocation. An example of this type should draw one's attention to the gradient phenomenon of collocation where at the one end of the scale it is strictly frequent and on the other – phraseological and therefore unpredictable. Figure 1 below roughly illustrates this kind of gradience including various interpretations of collocation.

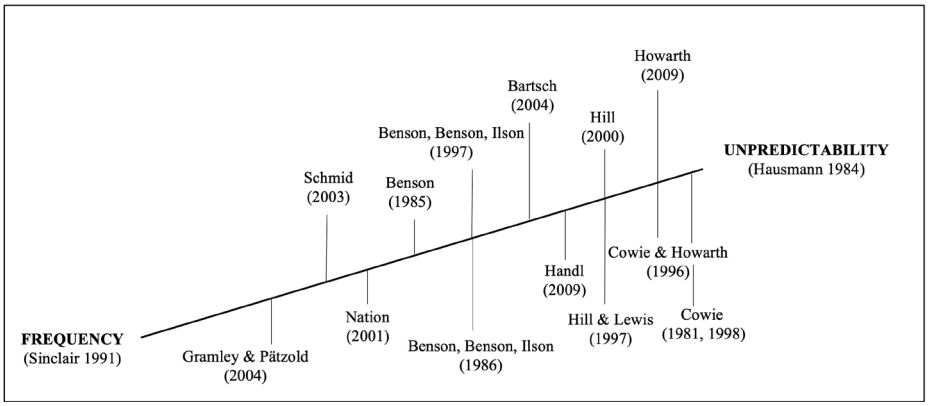

Figure 1: The continuum between frequent and unpredictable collocations as researched by various linguists.

But what kind of collocations does one want to see in collocation dictionaries? It could be assumed that there would be little linguistic use, if any, listing collocations related by a topic (=text-oriented approach). Then, it can be also argued that dictionary users would benefit both from frequent collocations and phraseological ones since the former occur frequently in the language and are thus more appealing and relevant to foreign language learners and the latter carry more problem-solving value, meaning that one might not know what the suitable collocate for a certain lexical unit is, or has already made a mistake and would then want to consult a dictionary. However, it might be reasonable that given the limited space that (paper-based) dictionaries have in their disposal, one would probably opt for semi-transparent phraseological collocations rather than frequent but transparent combinations which one has better chances of coming across beyond the dictionary scope rather than phraseological collocations.

Criticism of such a view can certainly be expected. One piece of it comes from Tarp (2008) who condemns Hausmann's (1985) distinction between banal and typical combinations and the conclusion that only typical combinations should be included in dictionaries. Tarp elaborates on the examples *einen Vogel fangen* and *einen Vogel in der Hand halten* and poses the question why the first combination should be considered a collocation whereas the second should not. However, here it is essential not to miss the purpose for such a distinction. It is certainly clear that any attempts at rigid

distinctions lead to even more cases in between. However, it is hard to believe that the distinction was made by Hausmann just for the sake of making it. It can rather be surmised that it was made with the aim of limiting the scope of a collocation dictionary to the most useful entries. For these reasons, in the present study, it is the third view, a more idiomatic perspective of collocations, that has been adopted to investigate their nature and presence in collocation dictionaries.

2 Analysing collocation dictionaries

Given the amount of interpretations of the notion of collocation, this is very likely to have had its implications on lexicography. At first glance, it might seem that the enormous attention paid to these linguistic units would result in a rather systematic account of collocations in dictionaries. In practice, numerous studies report challenges existing in the attempts to provide a comprehensive and exhaustive list of collocations (Herbst & Klotz 2003; Bartsch 2004; Lea 2007; Svensén 2009; Orlandi & Giacomini 2016). However, as mentioned in the previous section, one is more likely to consult a dictionary for the semi-transparent phraseological type of collocation than any other type. Also, although information about collocations can be found in a variety of dictionaries, e. g. a general monolingual dictionary (Herbst & Mittmann 2008), it is probably more efficient to consult a collocation dictionary to spare the effort.

Another question that has driven lexicographic analysis in the present study is connected with the comparison of collocations and their presentation in dictionaries across languages. In other words, given that a collocation dictionary is aimed at a very specific type of information, i. e. describing the collocative capacity of the base word, it is curious to see whether this leads to omitting language-specific information such as aspects connected to declination, conjugation, articles, etc. in dictionaries of different languages. The following languages have been chosen as occupying different positions on the scale from a limited inflection system to a richer one: English, German and Russian.

The process of selection of collocation dictionaries for the analysis varied in accordance with the degree of their availability in the respective language. Whereas English has a wide choice of collocation dictionaries, e. g. the *Oxford Collocations Dictionary*, the *Macmillan Collocations dictionary*, the *Longman Collocations Dictionary*, it has been decided to opt for the *Dictionary of Selected Collocations* which is significantly smaller than any of the dictionaries listed earlier. However, it was hoped that the word 'selected' pointed towards a more phraseological approach towards item selection. For German, only two dictionaries have been available: das *Wörterbuch der Kollokationen im Deutschen* (transl.: Dictionary of German Collocations) and *Feste Wortverbindungen des Deutschen* (transl.: German fixed word combinations). Yet, word combinations in the latter dictionary were subdivided into topics, e. g. family, food

and drink, communication, etc., and the dictionary was compiled by Swiss authors. To eliminate variations in dialect and a bias towards a certain range of topics, the Dictionary of German Collocations has been chosen for the analysis. Finally, for the Russian language, only one dictionary, the *Dictionary of Russian Word Combinations* (in the original: *Словарь сочетаемости слов русского языка*) has been suitable for the analysis since other materials were either aimed at foreign language learners only or served as teaching materials in class and were too scarce.

So, one collocation dictionary from each language was chosen in order to establish 1. whether these dictionaries have a more phraseological or frequency-based approach towards the choice of headwords and their collocates, and 2. to what extent language-specificity influenced their compilation.

In this respect, three printed dictionaries were examined:

- English: LTP Dictionary of Selected Collocations (LTP)
- German: Wörterbuch der Kollokationen im Deutschen (WKD)
- Russian: Dictionary of Russian Word Combinations (DRWC)

To begin with, a collocation dictionary is a production-oriented dictionary which provides more information about specific word combinations than a general monolingual dictionary would (Herbst & Klotz 2003). And although several groups were earlier identified as possible users of such dictionaries, there is one group which these materials would be particularly useful for and which are most often mentioned in the foreword of collocation dictionaries. Lea (2007: 261), one of the managing editors of the *Oxford Collocations Dictionary for students of English*, claims that 'native speakers, of whatever language, are masters of collocation almost by definition'. Although it is certainly true that being a native speaker does not guarantee a 100 % knowledge of all collocations in the language (Dabrowska 2014), according to the usage-based approach, it is the frequency of experience and input that facilitates language learning, be it L1 or L2 (Ellis 2002; Tomasello 2003; Goldberg 2006; Boyd & Goldberg 2009; Behrens 2009; Ellis & Collins 2009). Thus, given that it is native speakers who are normally exposed to authentic language input the most, the primary target user of a collocation dictionary is a second/foreign language learner. But how advanced should this learner be? A beginner or an elementary level learner would not be able to make use of such a dictionary since it would normally save space for collocates and lack definitions (which could be necessary on a lower level). Thus, a collocation dictionary would be preferably aimed at upper intermediate L2 learners.

According to Herbst & Klotz (2003), the information in a collocation dictionary will only fulfil its purpose if the following two conditions are met:

1. this information is provided in the correct place in the dictionary; and
2. the information about the specific and habitual relations between words is recognizable.

The use of such abstract and subjective terms as 'correct' and 'recognizable' allows some room for confusion. However, Herbst and Klotz explain that Hausmann's distinction between the base and the collocate can be of great help in identifying the correct place. So, from a learner's perspective, the knowledge of the base would evoke the need for a collocate which a language learner would be able to find in the dictionary under the entry of the base. However, unfortunately, as Herbst & Klotz (2003) stress, this approach is not widely used in dictionaries because one would often observe the following tendency: a collocation is either given under the entries of both the collocate and the base or under the collocate only. The alternating choice between the base and the collocate mirrors the two approaches towards writing a collocation entry (Svensén 2009). So, according to the semasiological perspective aimed at perception, the dictionary user has both parts of the collocation and needs the dictionary to decode its meaning. On the contrary, the onomaseological approach focusing on production, implies that the user first thinks about a concept which he or she wants to verbalise, and then searches for a correct expression in the dictionary. The latter approach is favoured in the present paper because decoding the meaning of a particular collocate should be possible with a general monolingual dictionary as well, let alone the fact that decoding of the base would not be necessary due to the semi-transparent nature of collocations; thus, in a collocation dictionary, it would be more appropriate to keep to the onomasiological approach.

As far as recognition of the information about collocation is concerned, one refers to presentation forms. To be more precise, in order the specific information about collocation becomes recognized by the language learner, this information needs to be (at least) noticeable, i. e. given in italics, bold, etc. Lea (2007) also agrees that one of the biggest challenges in compilation of a collocational dictionary consists in the right presentation of the material.

Another challenge is connected with selecting appropriate material for the dictionary. And given the numerous interpretations of collocations, the choice of material is undoubtedly one of the toughest decisions dictionary makers have to face. Lea (2007: 267) outlines three basic questions that could be asked to direct the process in search:

- Is this a typical language use?
- Might a learner want to express this idea?
- Would they look up this entry to find out how?

Without having to go further into details, regarding the question which collocations should be presented in a collocation dictionary, Herbst and Klotz summarize the answer as follows:

> Als Kollokationen einer Sprache können im sprachkontrastiven Sinne all diejenigen Kombinationen von Wörtern gesehen werden, bei denen die Gefahr einer falschen Übertragung besteht, da zumindest einer der Kombinationspartner im Kontext der Wendung anders übersetzt werden muss, als dies von den Benutzern typischerweise erwartet wird. (Herbst & Klotz 2003:138)

Basically, this quotation sufficiently explains that what a learner expects to find in the dictionary is probably something which is for some reason new to him/her, i. e. not fully transparent or predictable for the language learner.

Before proceeding to examine the target dictionaries, the following criteria are sought to be paid attention to in respect to the process of analysis (where the right column provides alternative realizations of the criterion):

A1 Macrostructure of the dictionary:	divided into subsections / solid
A2 Selection of items	idiomatic/frequent
B1 Choice of headword:	the base / the collocate
B2 Microstructure:	clearly structured / a block of text
B3 Meaning:	one / separate entries for different senses
B4 Selection of items:	collocates based/not based on frequency
B5 Recognizability of information:	varying fonts, indentation / no marking
B6 Style:	indicated / not indicated

2.1 LTP Dictionary of selected collocations

As for the criterion A1, the LTP *Dictionary of Selected Collocations* (LTP) (which is a dictionary of English collocations) consists of two sections – the noun and the adverb sections – listing 2000 nouns providing collocates accounting for 50,000 collocations, and 1200 verbs and adjectives collocating with 5000 adverbs. Although meant to make the search for collocations more user-friendly, the division of the dictionary into two sections appears to be a bit confusing. Since the headwords in the noun section are all nouns, the adverb section, as the name suggests, should contain adverbs. Instead, the headwords in this section are either verbs or adjectives. Such a division might be slightly misleading for the dictionary user since it uses different perspectives: the noun section has the noun as the starting point, and the adverb section refers to the adverb as the ending point. A possible improvement could consist in renaming the so-called adverb section into 'verb and adjective section'.

However, the division proved to be justified in respect to zo B1 because the headwords were always the bases of collocations. The only exception concerns the cases when the headword could appear in a phrase, e. g. *addiction to alcohol* and *freedom of action* (the headword is underlined). It is not so straightforward to judge what the bases in these cases are. On the contrary, it is easier to speculate on cases like *piece of cake* or *slice of bread*. While one could argue that in the latter cases, collocations consist of 'a measure' and 'something being measured', it is possible to account for *cake* and *bread* as the bases, because one rather thinks of the concept than the measure. On the other hand, grammatically, the word preceding *of* is to be seen as the independent head of the phrase followed by the postmodifier (or complement in the valency approach) which is dependent on it. Thus, the valency of the first noun

will determine the presence of the second noun. In the end, it could be argued that the distinction between a collocate and a base can be considered irrelevant for these phrasal collocations.

As for criterion A2, the LTP seems to largely follow the phraseological approach, which means it contains collocations which are not (fully) transparent. Concerning the structures listed in the dictionary, the most prevailing types are adjective+noun, verb+noun, noun+verb, adverb+adjective and verb+adverb. Thus, the LTP focuses on lexical classes rather than grammatical ones. However, in the cases which allow it, collocations in phrases are also provided, e. g. *in my humble opinion*, *the odds are that*..., *to show a certain reluctance to* and *heir apparent* (headwords underlined). As evident from these examples, the phrasal collocations are so heterogenous that it would be impossible to split them further into narrower classes without causing more complication for the dictionary user. However, not including them to the dictionary could be a sign of inconsistence since they are clearly not grammatical as, for example, *insist on* or *cater for*. Therefore, it can be claimed that the solution based on a category generalized as a 'phrase' is reasonable from the editors of the LTP.

Regarding the same criterion, A2, it is perfectly clear that the English language consists of endless combinations that fit the criteria of being not (fully) transparent and belonging to lexical classes. Probably to be able to provide collocations which do not appear just once in the language, one would have to base this selection procedure on corpus data.

So, three lemma frequency lists for the first 150 nouns, verbs and adjectives have been generated in the BNC. It was decided to consider the first 150 lemmas due to the Zipfian distribution (Baroni 2008) of the BNC according to which the distribution of the highest ranks is much sharper than of the lower ones, which would target the most frequent lemmas. The lists were then tested against the headwords of the LTP, and afterwards, each of the lemmas missing in the dictionary was manually checked in the BNC on the matter of possible and appropriate collocates. The following lemmas were not found in the LTP (see Table 1):

Table 1: Words from the BNC frequency lists (top 150) missing in the LTP.

Rank	Noun	Rank	Verb	Rank	Adjective
57	eye	100	draw	77	special
61	council	143	represent	83	short
83	minister			92	common
87	hour			101	private
				108	wide
				126	individual

Whereas this list does not claim to be exhaustive in itself, it illustrates how comprehensive the coverage of the LTP is in comparison with the most frequent lemmas of the BNC. Also, although it might be quite arguable whether certain items from this list should be included in the list or not, parallels were drawn to establish the appropriateness of the items in terms of those which are already in the dictionary. So, with respect to nouns, the first 150 frequent lemmas in the BNC contained *eye* (missing in the LTP) collocating with *naked* (149 hits), *watchful* (106 hits) and *private* (158 hits), etc. Regarding the verbs, *draw* is not represented in the LTP although it could collocate with *heavily* (64 hits) or *largely* (28 hits). And finally, the LTP does not include adjectives *private* collocating with *purely* (23 hits), *strictly* (20 hits) and *intensely* (14 hits), or *common* which collocates with *increasingly* (92 hits), *particularly* (46 hits), *extremely* (31 hits) and *relatively* (31 hits). Finally, a simple calculation shows that from 450 most frequent nouns, verbs and adjectives from the BNC, 97.3 % of them were covered by the LTP, which can be considered a good index.

Criterion B2 is fulfilled, which can be supported by Figure 2 illustrating the entry under the headword DAYS:

DAYS (period with *the*)

V: bring to mind, look back on, recall, recollect, remember, yearn for ~

V: ~ are gone/numbered/over/past

A: bad old, carefree, dark, difficult, early, good old, halcyon, last, latter, old, olden ~

P: end (your) ~; in ~ gone by

Figure 2: A dictionary entry for DAYS from the LTP.

One of the advantages of such a presentation concerns the fact that the headword DAYS implies that the collocates given are only to be used with this very word form rather than *day* generally. In fact, concerning dictionary headwords, the LTP explicitly distinguishes between possible lexical units of a lexeme (e. g. OPERATION (military, etc.) and OPERATION (surgical) as well as different word forms (e. g. DAY and DAYS), which is extremely helpful for the dictionary user to avoid misinterpretation and confusion. Second, the subdivision of the entry into V (verb), A (adjective) and P (phrase) enables one to quickly look for the collocates (ordered alphabetically) of the necessary word class without having to look through a solid text block. Finally, the tilde sign (~) helps to learn where exactly the headword is to be put (i. e. before/in the middle/after the collocate), which fulfils criterion B5. On the whole, the dictionary entry looks very user-friendly since it seems to provide all the necessary information and still looks not overloaded.

However, some weaknesses are worth mentioning. Although the headword selection seems to be corpus-based, the choice of the collocates, i. e. criterion B4, might not

be following the same scenario. For example, as seen from Figure 2, the verb section for the entry DAYS lists the following collocates: *bring to mind, look back on, recall, recollect, remember* and *yearn for*. All of them were searched for in the BNC and no occurrences of *bring to mind, recollect* and *yearn for* collocating with *days* were found at all. This certainly does not mean that *recollect*, for example, should not be listed in the dictionary as a collocate, since it seems to be an unfortunate coincidence that it is not in the BNC, but this situation begs the question why *reminisce* is not listed in the dictionary although it has the highest mutual information score (MI, a measure of statistical association) in combination with *days* (5.9). Regarding the adjective section, it is not clear why *last* and *later* should be listed in the LTP (which do not seem highly idiomatic), but *bygone* and *palmy* – not. Checking other headwords resulted in such inconsistences as well: the headword END was not shown to collocate with *abrupt* (40 hits), *bitter* (39 hits) or d*ead* (123 hits) (the latter reflecting the change of meaning for the base). Thus, the selection criteria for the collocates do not appear to be fully transparent.

Second, the B2 criterion is sometimes fulfilled with occasional inconsistency. Although the editors of the LTP have very wisely divided the entry into sections in terms of word classes, some deviations are present, e. g. in the entry DAYS under the verb section, one can find *bring to mind ~* which seems to be more appropriate for the phrase section rather than the verb where *recollect, recall* and *yearn for* (prepositional verb but still verb) finely belong. Next, the LTP does not meet criterion B6. One could probably claim that collocation dictionaries are primarily made for language learners to help them write essays using natural English. So, if one presupposes that essays normally include formal language, there will be no need in indicating different styles in the dictionary since it would logically not contain any informal uses. However, the LTP is aimed at those who want 'to write, translate or speak English accurately' (Hill & Lewis 1997: 6), so it is supposed to distinguish between different situations of use. However, the following collocations, for example, were found in the LTP without any indication of them belonging to the informal style: *get mixed up in an affair, wriggle out of a deal* and *damn fool*.

Finally, the fulfilment of criterion B2 leads to conflicting judgements. The LTP does not contain any example sentences with collocations, which on the one hand saves space and makes the collocates visible enough without having to mark them in italics or bold, but on the other hand could confuse the learner since they might not know whether *get mixed up in ~* (for the headword AFFAIR) needs an article or not.

Thus far, the LTP seems to have taken the phraseological approach towards presentation of collocations in the dictionary and provided a good coverage of the most frequent words in the BNC. Yet, sometimes fully transparent collocations of the kind *well-known author* or *tasty food* can be found. Therefore, two questions in this respect should be kept apart for the editors of LTP: instead of asking oneself what can generally collocate with *food*, it is better to ask what can collocate with

food that one would not necessarily know how to verbally express, thus, resulting in having *delicious, plain* or *nourishing food* in the dictionary instead of *tasty* or *good food*.

2.2 Wörterbuch der Kollokationen im Deutschen

The dictionary of German collocations, *Wörterbuch der Kollokationen im Deutschen* (WKD) claims to contain over 3200 headwords for the most important collocations (nouns (2346), verbs (617) and adjectives (290). Altogether, these headwords provide collocates accounting for 192,000 collocations in 54,000 collocation groups. Criterion A1 is not met since the dictionary has a fairly simple structure constituting a list of entries ordered alphabetically. This, however, can be treated as a positive feature as well because a dictionary user would not have to do extra work guessing in which part of the dictionary a certain collocation can be found. Alphabetical order, on the contrary, is more familiar to basic users. Also, according to the foreword of the WKD, one of those providing support throughout the work on the dictionary was Franz-Josef Hausmann, which could mean, following his criticism of Sinclarian node (Hausmann 1984), that the headwords in the dictionary would always reflect the base of a collocation (thus, criterion B1 met). Also, interestingly, if both parts of the collocation are present as headwords in the WKD, the rule of the base as the headword still applies to all of them, which is certainly wise enough as it saves space in such a costly product as a dictionary. For example, once can find both *Sinn* and *ergeben* in the dictionary, however the collocation *Sinn ergeben* can only be found under the entry of SINN.

Although it is hard to say which approach prevails without having to manually divide 192,000 collocations into two groups, it can be claimed that the selection of collocations in the dictionary represents a blend of the phraseological and the statistically-based approach. Thus, along with unpredictable cases like *Atem holen* and *starke Aufsicht* one would find *vom Dach springen* and *Behinderte unterrichten*. Yet, whereas the latter cases can probably not be called idiomatic, it is rather questionable whether they are statistically significant either. A search in the DWDS-Kernkorpus[2] does not show a single occurrence of them. However, surprisingly, the DWDS-Referenz- und Zeitungskorpora, which is a part of the whole DWDS corpus, shows 73 hits for *vom Dach springen* and 10 hits for *Behinderte unterrichten*, which seems to justify their presence in the dictionary. These findings support the claim made in the foreword of WKD, that it is based on corpus data, which supports criterion A2. Hausmann's possible influence can also be traced in the grammatical types of items that the dictionary has provided, or in other words, in the variety of struc-

2 The balanced corpus of the German language from 1900–1999 consists of 121397601 tokens.

tures. And whereas at first sight they might seem very similar to those in LTP, there is one particular structural aspect that distinguishes the WKD from the LTP. While for a noun entry, the LTP would list the verb and adjective collocates as well as a category generalized as 'phrases', the WKD takes a narrower perspective and rejects generalizations. Thus, one would come across lists of specific usages involving nouns in prepositional complements (marked in italics), for example, standing out from other classes of collocates but having no names of their own. Thus, the range of possible structures is subdivided depending on the base:

1. noun:	adjective+noun (bissig+Bemerkung)
	(subject)noun+verb (Elfmeter+gelingen)
	(object)noun+verb (Beruf+anstreben)
	noun (in a prepositional complement)+verb (auf Details+achten)
2. verb:	adverb+verb (freudig+begrüßen)
3. adjective:	adverb+adjective (eindeutig+erkennbar)

Returning to criterion A2, a test against the DWDS-Kernkorpus has been attempted to reveal the WKD's coverage of the most frequent lemmas. Again, three lists, each consisting of 150 most frequent nouns, verbs and adjectives were examined in the DWDS core corpus and tested against the WKD[3]. Table 2 below shows the results of this comparison[4].

Table 2: Words from the DWDS-Kernkorpus frequency lists (top 150) missing in the WKD.

Rank	Noun	Rank	Verb	Rank	Adjective
18	Teil	29	zeigen	14	lang
		46	erscheinen	22	verschieden
		103	wenden	59	ähnlich
		149	hängen	77	eng
		150	fällen	78	rot
				120	breit

3 The following query was used for generating the frequency lists: count($p=NN #sep) #by[$l] #desc_count. The NN was substituted with V* for verb lists and ADJ* for adjective lists.

4 Some words from the frequency lists were missing in the WKD, but were still not included to Table 2 since they either did not have any specific collocates or were too general to collocate with anything, e. g. *Mutter, sein*, modal verbs, *haben, tatsächlich, einzeln,* etc.

Generally speaking, the coverage of the WKD seems to be almost identical to that of the LTP since it lists 97-5 % of the 450 most frequent nouns, verbs and adjectives of the DWDS-Kernkorpus. This is, however, not surprising taking into account the fact that it is based on a corpus consisting of more than 50 million sentences (Quasthoff 2011). Nonetheless, several gaps were revealed concerning the inclusion of high-frequency headwords to the dictionary. For example, given that the ranking of *Teil* is rather high due to the number of its occurrences (49,052), it is somewhat surprising that it has not been found in the WKD. Yet, it could allow for essential collocations for a language learner by combining with *groß* (5137), *wesentlich* (415), *erheblich* (264), *überwiegend* (222) and *beträchtlich* (154). Also, the WKD does not contain such verbs as *erscheinen*, which is in the top 50 of the most frequent verbs in the DWDS-Kernkorpus, collocating with *wichtig* (198), *zweifelhaft* (124), *fraglich* (83), *sinnvoll* (55) and *unwahrscheinlich* (35), and *fällen* which collocates with *schwer* (510) and *leicht* (183). Finally, the following could serve as examples for missing adjective headwords in the WKD: whereas the dictionary contains *weit* collocating with *möglichst, ziemlich, vergleichsweise, extrem* and *ungewöhnlich*, it does not have the headwords *eng* and *breit* which have exactly the same collocates. Thus, while the approach based on the selection of statistically significant collocations for the dictionary can definitely be traced, it remains unclear why certain high-frequency words are not found in the WKD whereas words of similar frequency can. Perhaps, one explanation for such a situation can lie in differing frequency lists among various corpora. So, given that the prevailing majority overlaps, a small amount of words would still be different based on the selection of texts used for the compilation of a certain corpus.

Taking a closer look at the entry in Figure 3, one feature is particularly remarkable and typical of dictionaries, and that is the two-column presentation which visually makes the page less crammed with too much information. Thus, so far, one could claim that the organization of entries in the LTP deviates from the convention.

There are a few more positive things to be mentioned regarding the WKD which, however, cannot be assigned to one of the criteria listed above due to their language specificity, but which still reflect the content of an entry. Although this information is probably to be taken for granted, it is encouraging to see the article *die* next to the headword, which indicates its feminine gender. Another aspect driven by language specificity is the indication of cases. Thus, since verbs in German require a specific case for the noun following it, just mentioning a collocate would be of little use for a language learner (unless the case information is already known to them). So, for example, the WKD states that *Kreativität* can collocate with the verb *entspringen* which in turn would require the noun to be in the dative case. This indicates that the dictionary aims at language learners as their primary target user.

Then, in terms of criterion B2, the LTP and the WKD were compared on the number of entries. While both dictionaries have approximately the same amount of entries, the latter provides almost four times as many collocations than the LTP (192,000 compared to 55,000). Logically, for the WKD, this was possible to achieve

Kreativität die

V: NOM. entstehen · gedeihen ■ DAT. entspringen ■ AKK. anregen · fordern · fördern ■ beweisen · entwickeln · zeigen ■ ausleben · entfalten · freisetzen · mitbringen ■ entdecken ■ vermissen ■ ersticken · hemmen ■ DAT. *jmdm. zu (mehr) K.* verhelfen ■ *an K.* hapern · mangeln
A: außergewöhnlich · bemerkenswert · besonders · groß ■ enorm · erstaunlich · überschäumend · unerschöpflich · ungeahnt ■ ungebremst · ungebrochen ■ sprühend ■ fehlend ■ göttlich · kindlich · menschlich ■ geistig · künstlerisch · schöpferisch ■ politisch · sprachlich · wissenschaftlich ■ ökonomisch · technisch · unternehmerisch

Figure 3: A dictionary entry for KREATIVITÄT from the WKD.

by providing more collocates for an entry. However, simply listing all the collocates alphabetically as was done in the LTP, would not be appropriate for the WKD since the average number of collocates for one entry is about 60. So, what one can find in the WKD is an attempt to sort certain semantic groups of collocates, e. g. based on the strength of evaluation for adjectives, i. e. from strong/positive to negative/weak; and life cycle for verbs, e. g.: for *Denkmal* one would find *errichten – enthüllen – anschauen – sanieren – zerstören*. Thus, what one could observe in the entry KREATIVITÄT is first listing adjective collocates denoting positive and strong evaluation, e. g. *außergewöhnlich, bemerkenswert* or *enorm*, followed by negative evaluation, e. g. *fehlend*, interrupted by a black square separator (■).

Nonetheless, the fulfilment of criterion B2 at times leads to vague subdivisions. It is somewhat unclear why the first two groups of adjective collocations – *außergewöhnlich – bemerkenswert – besonders – groß* and *enorm – erstaunlich – überschäumend – unerschöpflich – ungeahnt* – are not listed under one group since they both seem to have strong and positive evaluation and belong to evaluative adjectives. This, however, is not the case with the last two groups of adjective collocates in one of which one can find *politisch* and *wissenschaftlich* and in the other – *ökonomisch* and *technisch*. So, it is quite straightforward that these collocates are not evaluative, however, they could have been put into one group as they encompass various fields; instead, they are split for an unknown reason. Next, as far as B3 is concerned, a native speaker expertise questioned the possibility of *göttlich* and *kindlich* as collocates for *Kreativität*. The two collocates were checked in the whole DWDS corpus which showed no occurrences of these collocations. The question is why certain collocates being neither frequent nor idiomatic are listed under the entry. Therefore, the WKD tried to combine both the phraseological and the frequency approaches to select the items for the headwords, which resulted in a slightly better coverage (0.2 % difference) of the corpus as com-

pared with the LTP. However, occasionally items being neither idiomatic nor frequent can be found in the dictionary. Finally, language-specific information, i. e. articles and cases, was paid enough attention to, so as to help the language learner to make a better use of the dictionary.

2.3 Dictionary of Russian word combinations

In comparison with the LTP (1997) and the WKD (2011), the *Dictionary of Russian Word Combinations*[5] (DRWC) is the oldest one (1983). Logically so, the dictionary does not meet the A2 criterion since it is not corpus-based and the Russian National Corpus (RNC) was only introduced 20 years later in 2003. However, pre-revolutionary and Soviet monolingual dictionaries of the Russian language as well as other lexicographic sources are claimed to have been used to compile the DRWC (Denisov & Morkovkin 1983). The dictionary encompasses fewer entries in comparison with the English and German ones described earlier (2506 headwords): 1255 nouns, 727 verbs and 524 adjectives. However, it is not possible to say how many collocations these headwords provide, as the DRWC does not explicitly state it (though it has about 700 pages of small-size font). Also, similarly to the WKD, its simple alphabetically ordered structure (criterion A1 similar to the WKD) makes the search for the necessary headword as uncomplicated as possible.

With a prevailing proportion, the collocations in this dictionary can be better characterised as frequent ones rather than idiomatic (criterion A2) since more often both parts of a collocation are fully transparent and rather represent habitual co-occurrences. The editors of the dictionary in their turn claim to have based their selection of items on educational and methodological expediency including both objective (frequency lists) and subjective (collective evaluation of a given word considering its thematic, situational and combinatorial value) criteria (Denisov & Morkovkin 1983: 8).

Another similarity which is common for all of the three dictionaries concerns the headwords, i. e. the headwords are always either nouns, verbs or adjectives. However, unlike the LTP and the WKD, the headwords in the DRWC are not always the bases (B1 partially met). Deviations from Hausmann's bases can be often observed. This will be discussed in more detail while analysing the dictionary entry ПРОДАВАТЬ 'to sell' later in this section.

The headwords, in their turn, provide a ground for a wide range of various structures. For example, for a noun entry, one could find the following structures: adjective+noun, noun+noun (standing in case relationships), verb+noun (e. g. in genitive, dative or accusative cases), preposition+noun and noun+noun+short adjective+-noun, etc. So, the number of structures listed is wide, from simple binary ones till

5 In the original: 'Словарь сочетаемости слов русского языка'.

extensive phrases, which provides a full and comprehensive description of the collocational possibilities of a word.

Although the DRWC is not a corpus-based dictionary, it was supposed to be tested against the RNC which appeared later and find out how representative the selection of its items was. However, the RNC does not contain the feature of extracting any significant frequency lists. The only ones it provides are bi-grams and n-grams which generate combinations of the kind потому что 'because' or не было 'was not', which we are not interested in. Thus, a decision was made to use the frequency lists generated with the help of the Sketch Engine, a corpus manager and text analysis software containing more than 500 different corpora in over 90 languages. One argument in favour of this decision consists in the fact that while working on the German frequency lists in the DWDS-Kernkorpus, the Sketch Engine-generated frequency lists were compared to those of the DWDS corpus, and it turned out that they were largely the same, which proves the validity of those Russian frequency lists used for the testing of the DRWC. So, once again, three lists each containing 150 most frequent nouns, verbs and adjectives were generated in the Sketch Engine to allow further analysis[6]. Table 3 illustrates the results of this comparison.

Table3: Words from the Sketch Engine-generated frequency lists (top 150) missing in the DRWC.

Rank	Noun	Rank	Verb	Rank	Adjective
36	информация 'information'	93	бывать 'be'	12	различный 'different'
45	рынок 'market'	104	обладать 'possess'	97	открытый 'open'
64	услуга 'service, favour'	107	предоставлять 'provide'	121	индивидуальный 'individual'
66	данные 'data'	123	приобретать 'acquire'	132	уникальный 'unique'
75	ситуация 'situation'	127	учитывать 'consider'	142	доступный 'available'
91	счёт 'bill, score'	129	поставлять 'deliver'	148	традиционный 'traditional'
102	автомобиль 'automobile'	130	утверждать 'confirm'		
109	технология 'technology'				
115	сеть 'network'				

6 The precise name of the corpus used in the Sketch Engine is Russian Web 2011 (ruTenTen11) consisting of about 18 billion tokens.

The comparison has revealed interesting results: although the dictionary was compiled using resources other than electronic corpora, it still covers 94.2% of the corpus-generated frequency lists, which can be considered a rather positive result given the circumstances (criterion A2). Also, a closer look at the list of the missing words shows that those words are mostly missing which diachronically received more importance only later in the XXI century, e. g. nouns denoting market, service, data, technology or network. However, this explanation can certainly not account for all the missing words, since there are quite a few words with meanings not related to the historical development. For example, although the high-frequency noun *выбор* 'choice' occurs 40,466 times in the RNC, it is not present in the DRWC. So, a learner of Russian would not be able to find the following collocations in the dictionary: *сделать выбор* ('make a choice', 174 hits), *правильный выбор* ('right choice', 134 hits), *окончательный выбор* ('final choice', 68 hits), *широкий выбор* ('wide choice', 66 hits) or *осознанный выбор* ('conscious choice', 27 hits). As for verbs, not many of them in Table 3 can be thought of possessing specific relations with other words. On the other hand, the DRWC proved to contain combinations of both kinds – idiomatic and simply frequent but not idiomatic. Yet, the verb *учитывать* 'consider' which has 10,097 hits in the RNC, could serve as the base for such collocations as *учитывать интересы* ('consider interests', 34 hits) or *учитывать специфику* ('consider specificities', 28 hits), and the highly-frequent adjective *открытый* ('open', 52471) building the collocations *на открытом воздухе* ('outdoors' (literally: 'in the open air'), 381 hits), *с открытым ртом* ('with one's mouth open/surprised', 328 hits), *широко открытый* ('wide open', 289 hits) and *открытое лицо* ('open face', 84 hits).

Figure 4 below illustrates the fulfilment of criterion B2 for the verb entry ПРОДАВАТЬ 'sell' from the DRWC.

So, the entry starts with the headword in bold capital letters which additionally shows the stress with an apostrophe. Such presentation will help the dictionary user to quickly recognize the word on the page (criterion B5 met). It is then followed by the various grammatical forms that the lexeme, in this case the verb, can have, i. e. word forms with different inflections denoting the grammatical aspect of person.

Interestingly, whenever deviations from the regular form in terms of spelling or pronunciation are possible, the irregularities are indicated in the list of word forms. Another positive aspect concerns the specification of word forms belonging to different grammatical aspects (i. e. imperfective and perfective), which is one of the biggest challenges for learners of Russian. It must be mentioned, however, that whenever both aspects exist for a given verb, it is the infinitive form of the imperfective aspect that appears as the headword. Thus, in the entry above, the imperfective *продавать* (which is the headword) means 'to sell' and the perfective *продать* signals the result, i. e. 'to have sold'.

Although this is not typical of a collocation dictionary, word forms are followed by a definition of the headword. If the lemma consists of more than one lexical unit, then

Figure 4: A dictionary entry for ПРОДАВАТЬ 'sell' from the DRWC.

they are marked by Arabic numbers under the same entry (similarly to the WKD). Thus, criterion B3 is met.

With respect to criteria B2 and B5, collocates of various structures are listed in separate paragraphs. Their presentation deserves some explanation. As indicated previously, verbs in the DRWC are not only followed by nouns or pronouns in the accusative case. To be able to signal a certain grammatical case, question words which reflect the valency slots and indicate the case are used. The groups themselves are introduced by the headword followed by a question signalling the case or another construction (e. g. infinitive construction: *решить продавать* 'decide to sell'); so, *продавать кого-что* 'sell who/what' tells the dictionary user that the following collocates will have to be used in the accusative case[7], *продавать кого-что-л. кому* 'sell who/what to whom' signals the dative case, and *продавать что-л. где* 'sell what where' signals the prepositional case. In addition, the introduction to the collocate group is marked in bold with more attraction given to the signalling questions (e. g. w h e r e instead of where). Finally, other traditional conventions are maintained in the DRWC, such as the tilde sign (~), example sentences and indication of style in

7 In Russian secondary schools, students are taught the 6-case paradigm system using signalling questions as the basis for this distinction. That is why signalling questions in the DRWC could be of great help for any language learner.

parenthesis (the latter being only present in the DRWC out of the three dictionaries analysed so far).

Regarding possible drawbacks, to begin with, it was mentioned earlier that criterion B1 was partially met. As a matter of fact, the headwords of the DRWC inconsistently provide both collocates and bases as a headword. To be more precise, the list of collocates under the entry can both contain a collocate and a base in Hausmann's sense. Thus, under the entry ПРОДАВАТЬ 'sell', one can find both *продавать газеты* 'sell newspapers' and *продавать выгодно* 'sell profitably'. Interestingly enough, the dictionary also has the entry ГАЗЕТА 'newspaper' which also lists *продавать газеты* 'sell newspapers' as one of its possible collocations. And although such overlaps double the chances of a dictionary user to find the necessary collocation, such intersections are or should probably be unacceptable in the business of printed dictionary making nowadays.

Next, within criterion B2, some of the collocate groups seem to provide such a list of lexical units which are hard to imagine as anything specific or characteristic of the headword. Thus, in the entry above, the places where one could generally sell are specified as a separate collocate group; so, one gets the following collocates: *в магазине* 'in the shop', *в киоске* 'in a kiosk', *на рынке* 'in the market', *на почте* 'at the post office', etc. Moreover, the places are specifically introduced with the prepositions *в* 'in' and *на* 'on' followed by some examples and then elision marks (...). This, however, appears to be a very unfortunate decision since it has just brought into another specificity of the Russian language. In fact, there is no rule that would clearly define the use of these two prepositions with places. Therefore, a language learner might end up making further mistakes with prepositional uses due to this confusion. A better solution could consist in not including the category of spatial collocates at all.

Then, the example sentences were previously referred to as a positive feature of the DRWC. Yet, these examples are scarce since they normally list just a few uses but not all of the collocates. In this respect, one can recall overlapping collocations, which might have influenced the lack of sufficient examples for each pattern.

As for criterion B4, it is somewhat challenging to provide any kind of justification to collocates listed. So, the first collocate group can be logically roughly divided into four categories: pets, food, clothes and equipment. But there are certainly many more things that can be sold and are expected to have a higher frequency, e. g. a car or a house (which are not to be seen in the lists). Also, the lists of collocates are not ordered alphabetically. One would reasonably suppose that they might be listed based on the frequency ranks. So, if one checks the first collocation in the list, i. e. *продавать собаку* 'sell a dog', in the RNC, one would find 3 occurrences in a corpus consisting of about 300 million words. The second collocation, *продавать попугая* 'sell a parrot', does not occur in the RNC at all. The RNC, however, shows that the most frequent collocations of this headword with a noun are *продавать хлеб* ('sell bread', 35 hits), which is present in the DRWC, and *продавать землю* ('sell land', 19 hits). An elaborate solution might have been made by the editors of the DRWC:

the lists of collocations are predominantly followed by elision marks (...) meaning that there are other collocates that can also combine with the headword. However, whereas it might be an easy way out for the dictionary makers, a learner of the language would not find such a scenario particularly helpful or pleasant, as he or she will either end up thinking that the collocates which are listed are the most representative ones, or mistakenly using another word not present in the list by means of simple generalization.

Summing up, the DRWC stands out from the three dictionaries being analysed by having a mixed methodology of headword representation, detailed information on the headwords (word forms, definition), a heterogenous nature of collocate structures and example sentences. Also, there is one main difference that one can put one's finger on: whereas the LTP and the WKD put more focus on the item-specificity in lexical terms, the DRWC seems to concentrate more on the grammatical relations by listing various patterns and items which could fill these slots. Yet, it can be argued that it justifies its title which is a *Dictionary of Russian Word Combinations*, not of collocations exclusively. Also, of the three dictionaries, the DRWC seems to be the one which is the most focused on language learners since it describes basic information for every headword such as stress, conjugation, gender, etc.

3 Conclusion

Thus far, it has been attempted to analyse one of the existing collocation dictionaries in each of the languages under question, i. e. English, German and Russian, by applying a set of outlined criteria which are by no means comprehensive and just served as a set of criteria to simplify the process of analysis. However, the list can also be used as a methodology for analyzing other collocation dictionaries. So, to this point, the outcome of the analysis is the following: given the flexible nature of a collocation, which according to many researchers represents gradience on an idiomatic scale, the three dictionaries reflect a similar type of gradience. So, whereas the LTP in English seems to be the most phraseological dictionary since it predominantly consists of idiomatic and item-specific collocations, in the DRWC free combinations (being not necessarily statistically significant) prevail over the idiomatic uses. The German WKD in its turn appears to take the middle position on this scale consisting of combinations of both kinds. However, it is important to admit that the distribution of dictionaries on this scale should not be regarded as language-specific because it is certainly not due to the assumption that the Russian language does not have so many idiomatic combinations that they are not to be seen in the DRWC. Thus, the idea of endless potential of language creativity is to be taken for granted.

Moving on now to consider the influence of language specificity on the compilation of dictionaries, it can be definitely observed in each of the cases. So, whenever a

characteristic feature was present in the language and could serve as a reason for mistakes by language learners, these cases were paid due attention. First, two out of the three languages possess the distinction of grammatical gender in nouns; the German dictionary indicates it with the help of the articles (*der, die, das*), and the Russian dictionary indicates it with words – masculine, feminine, neutral – since the language does not have articles. Second, the same languages display a case paradigm system: German distinguishes between 4 cases (nominative, genitive, dative, accusative) and Russian – between 6 (the additional ones are instrumental and prepositional). That is why the case was indicated with the case abbreviation (e. g. G for genitive) in the WKD and with signal questions in the DRWC after each verb listed as a collocate. Finally, the DRWC provided distinctions between the imperfective and perfective aspects of the verbs and the varying word forms to predict the possibility of potential common mistakes. As for the LTP, it did not have to resort to any of these methods since it was irrelevant for the English language.

Having discussed language specificity, one could turn to the consistency of presentation methods in the dictionaries. All dictionaries made use of bold and italic fonts to make the target information noticeable. Also, various groups of information were structurally separated so as to make the search for the necessary structure easier. However, it is vital to state that sentences with example collocations can be regarded one of the weakest parts for all the three dictionaries. So, examples were either not present at all (LTP, WKD) or they were very sketchy covering few collocates (DRWC).

Another point concerns the description of the headword. Only the DRWC provided information about its meaning, the stress and the word forms. Whereas the latter can be justified since the Russian language has rich declination and conjugation paradigms for nouns, verbs and adjectives with a plenty of irregularities which concern both spelling and pronunciation, the necessity of having the definition and the stress of the headword in a collocation dictionary (or a dictionary of word combinations) is rather questionable. It is undoubtedly wrong to claim that this information is absolutely useless because it clearly carries informational value for a language learner. On the other hand, one (printed) dictionary cannot entail information of all kinds, i. e. definition, stress, word forms, grammatical and lexical patterns in which the headword occurs, possible idioms, irregularities, lexical relations with other words (synonyms, antonyms), etc. There is simply not enough space.

Yet, one should also admit that the amount and scope of information in these dictionaries is not only explained by language-specificity issues. In other words, there seems to be a direct relationship between the scope of the dictionary and the target groups. This means that the more detailed the information around the headword is, the more it is likely to be directed at language learners. In this respect, the LTP seems to be focusing on various potential users from language learners to native speakers since it does not provide any other information apart from the collocate lists, whereas the DRWC provides exhaustive information including such trivial features as

word stress. Finally, the WKD finds itself in between these two orientations, although slightly closer to language learners as their most potential target group.

In conclusion, returning to the state of affairs, it is straightforward to admit that most problems in dictionary making are directly related to the necessity to save space in these dictionaries, which leads to the conclusion that having endless amounts of space could result in not having to sacrifice important aspects in favour of the more important ones. This, however, is not a dream because online and electronic dictionaries have already been introduced in the XXI century and proven themselves to be a better and more comprehensive reference tool. Regarding general advanced learners' dictionaries, (Handl 2009) also concludes that the principles of presenting collocations would have two major drawbacks in a printed dictionary: first, a strict separation of the denotational information and the collocation boxes in the entry; second, absence of context for collocations. As it turns out, the latter problem seems to be topical for collocation dictionaries as well. These drawbacks point in the direction of an electronic dictionary, and thus a collocational web where cross-references are possible, and one should not rely on alphabetical ordering.

In the end, although it is not particularly new to state that online and electronic collocation dictionaries are a better alternative and provide more space for information and make it possible to include all kinds of combinations (both idiomatic and frequent), it was still useful to see in what way collocations tend to be presented in printed dictionaries, because often the same printed dictionary entry is just transferred to the online dictionary with a minimum of changes. For example, the entry DECISION in the 7th edition of the printed *Oxford Advanced Learner's Dictionary* (OALD7) is almost identical with the one that can be found in the online version of the *Oxford Learner's Dictionaries* (OLD). The changes concern the addition of pronunciation in American English as well as information about the word origin, word family and extra examples in the end of the main entry. However, the idea of an online or electronic collocation dictionary could be more convincing, useful and user-friendly if more interactivity was brought into the entry, e. g. cross-references, word clouds, frequency information, moving windows, other relevant information grouped logically, etc. It is certainly much easier to provide suggestions for improvement rather than to actively work on it and implement it, however, the results of such work are definitely rewarding and would lead to a step forward in lexicography and language learning.

Bibliography

Dictionaries

DRWC = *Dictionary of Russian Word Combinations* [Словарь сочетаемости слов русского языка]. Ed. by Denisov, Petr N./Morkovkin, Valeriy V. Moscow: Russkiy Yazik, 2nd edn. 1983 [1st edn. 1978 Comp. Denisov, Petr N./Morkovkin, Valeriy V.].
FWD = *Feste Wortverbindungen des Deutschen. Kollokationenwörterbuch für den Alltag*. Narr Dr. Gunter, 2014.
LCD = *Longman Collocations Dictionary*. Pearson Longman, 2013.
LTP = *LTP Dictionary of Selected Collocations*. Ed. by Hill, Jimmie/Lewis, Michael. Hove: Language Teaching Publications, 1997.
MCD = *Macmillan Collocations Dictionary*. Macmillan Education, 2010.
OALD7 = *Oxford Advanced Learner's Dictionary*. Ed. by McIntosh, Colin/Turnbull. Joanna. Oxford: Oxford University Press, 7th edn. 2005 [1st edn. 1948 Comp. A.S. Hornby].
OCD = *Oxford Collocations Dictionary for Students of English*. Oxford: Oxford University Press, 2002.
OLD = *Oxford Learner's Dictionaries*. Oxford: Oxford University Press.
WKD = *Wörterbuch der Kollokationen im Deutschen*. Ed. by Quasthoff, Uwe. Berlin/New York: De Gruyter, 2011.

Corpora

BNC = *British National Corpus*
DWDS = *Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart*
RNS = *Russian National Corpus* [Национальный корпус русского языка]
ruTenTen11 = *Russian Web 2011* (part of the Sketch Engine, corpus manager and text analysis software)

Articles and monographs

Baroni, Marco. 2008. Distributions in text. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics. An international handbook*, 803–821. Berlin: Mouton de Gruyter.
Bartsch, Sabine. 2004. *Structural and functional properties of collocations in English: A corpus study of lexical and pragmatic constraints on lexical co-occurrence*. Tübingen: Gunter Narr Verlag.
Behrens, Heike. 2009. Usage-based and emergenist approaches to language acquisition. *Linguistics* 47(2). 383–411. doi: 10.1515/LING.2009.014.
Boyd, Jeremy K. & Adele E. Goldberg. 2009. Input effects within a constructionist framework. *The Modern Language Journal* 93(3). 418–429.
Chen, Hao-Jan Howard. 2011. Developing and evaluating a web-based collocation retrieval tool for EFL students and teachers. *Computer Assisted Language Learning* 24(1). 59–76.
Choi, Sungmook. 2017. Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research* 21(3). doi: 10.1177/1362168816653271.

Cowie, Anthony P. 1981. The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics* 2. 223–235.
Dabrowska, Ewa. 2014. Words that go together: Measuring individual differences in native speakers' knowledge of collocations. *The Mental Lexicon* 9(3). 401–418.
Denisov, Petr N. & Valeriy V. Morkovkin (eds.). 1983. *Slovar' Sochetayemosti Slov Russkogo Yazyka [transl.: Dictionary of Russian Word Combinations]*. Moscow: Russkiy Yazik 2nd edn.
Ellis, Nick C. 2002. Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24(2). 143–188. doi: 10.1017.S0272263102002024.
Ellis, Nick C. & Laura Collins. 2009. Input and second language acquisition: The roles of frequency, form, and function. Indtroduction to the special issue. *The Modern Language Journal* 93(3). 329–336.
Goldberg, Adele E. 2006. *Construction at work: The nature of generalization in language*. Oxford: Oxford University Press.
Greenbaum, Sidney. 1970. *Verb-intensifier collocations in English: an experimental approach*. The Hague: Mouton.
Halliday, Michael A.K. & Ruqaiya Hasan. 1976. *Cohesion in English*. London: Routledge.
Handl, Susanne. 2009. Towards collocational webs for presenting collocations in learners' dictionaries. In Andy Barfield & Gyllstad Henrik (eds.), *Researching Collocations in Another Language. Multiple Interpretations*, 69–85. Hampshire: Palgrave Macmillan.
Hausmann, Franz Josef. 1984. Wortschatzlernen ist Kollokationslernen. *Praxis des neusprachlichen Unterrichts* 31. 395–406.
Hausmann, Franz Josef. 1985. Kollokationen im deutschen Wörterbuch. In Henning Bergenholtz & Joachim Mugdan (eds.), *Grammatik im Wörterbuch*, 118–29. Tübingen: Max Niemeyer Verlag.
Herbst, Thomas. 1996. What are collocations: Sandy beaches or false teeth? *English Studies* 77(4). 379–393.
Herbst, Thomas & Michael Klotz. 2003. *Lexikografie*. Paderborn: Ferdinand Schöningh.
Herbst, Thomas & Brigitta Mittmann. 2008. Collocation in English dictionaries at the beginning of the twenty-first century. *Lexicographica*, vol. 24. 103–119.
Hill, Jimmie & Michael Lewis (eds.). 1997. *Dictionary of selected collocations*. Hove: Language Teaching Publications.
Jiang, Nan & Tatiana M. Nekrasova. 2007. The processing of formulaic sequences by second language speakers. *The Modern Language Journal* 91(3). 433–445.
Lea, Diana. 2007. Making a collocations dictionary. *Zeitschrift für Anglistik und Amerikanistik* 55(3). 261–271. doi: https://doi.org/10.1515/zaa.2007.55.3.261.
Matsuno, Kazuko. 2017. Processing collocations: Do native speakers and second language learners simultaneously access prefabricated patterns and each single word? *Journal of the European Second Language Association* 1(1). 61–72. doi: 10.22599/jesla.17.
Orlandi, Adriana & Laura Giacomini (eds.). 2016. *Monolingual collocation lexicography: State of art and new perspectives* (Linguistic Insights 219). Bern: Peter Lang.
Palmer, Frank R. (ed.). 1968. *Selected papers of J.R. Firth 1952–59*. London and Harlow: Longmans.
Quasthoff, Uwe. 2011. *Wörterbuch der Kollokationen im Deutschen*. Berlin/New York: De Gruyter.
Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
Svensén, Bo. 2009. *A handbook of lexicography. The theory and practice of dictionary-making*. Cambridge: Cambridge University Press.
Tarp, Sven. 2008. *Lexicography in the borderland between knowledge and non-knowledge: General lexicographical theory with particular focus on learner's lexicography*. Tübingen: Max Niemeyer Verlag.

Tomasello, Michael. 2003. *Constructing a language*. Boston: Harvard Umiversity Press.
Wolter, Brent & Henrik Gyllstad. 2013. Frequency of input and L2 collocational processing. *Studies in Second Language Acquisition* 35. 451–482. doi: 10.1017/S0272263113000107.
Yamashita, Junko & Nan Jiang. 2010. L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly* 44(4). 647–668.