The Function of Collocations in Dictionaries

Margaret Cop

An increasing number of recent publications in English, German and Russian devoted to word combinations in dictionaries of several languages bear witness to the importance of these units in the lexicography of all languages (e.g. Cowie 1981 and 1986, Mel’čuk/Zolkošvijk 1984, Mel’čuk 1984, Kromann 1979, 1982, 1984, 1985 and 1988, Ivir 1988, Morkovkin 1984,); and the lively controversy which The BBI Combinatory Dictionary of English has stirred up amongst lexicographers and metalexicographers alike (e.g. in recent issues of the EURALEX Bulletin) shows that the differing roles which collocations can play in a dictionary still seem far from clear. The function of collocations in a dictionary is determined by the target group, the activity for which the dictionary is supposed to be used and the type of dictionary in which such collocations appear: mono- or bilingual, general or specialized.

What are collocations?

The definition of collocation underlying this paper is a hyponym of the more commonly known one and it is therefore more specific. However, for the sake of orientation, here are some approximate equivalents (superordinates) in French, German and Russian:

- English: collocation, lexical combination, lexical cooccurrence
- French: collocation, cooccurrence lexicale
- German: Kollokation, Wortfügung, Wortverbindung
- Russian: slovosocetanie¹

Collocations in a narrower sense will be seen as affinitive, bipartite combinations such as:

- Engl.: dispel FEAR, sweeping CONCLUSION, vitally IMPORTANT
- Fr.: rentrer sa COLERE, un PRIX dérisoire, REFUSER net
- Ger.: VERDACHT schöpfen, herbe KRITIK, peinlich GENAU
- Rus.: podnjat' VOPROS, sočnye KRASKI, žiznennno VAŽNYJ

These idiomatic combinations are semi-finished products of ‘langue’ („Halb fertigprodukte der Sprache“ — Hausmann 1984 and 1985; Bolinger 1966 speaks of “prefabs”) which the speaker need not create himself but which he retrieves from memory. Collocations thus play an economising role in speech production (Peters 1983: 85-86). Because the receiver can “identify” with these combinations, he accepts and immediately understands them. An infelicitous combination will cause discordance (e.g. *scatter fear) and may irritate the receiver (Korosadowicz-Strużyn’ska 1980: 115).
This is why collocations are important learning units. Collocation thus belongs to "langue" as a norm (as opposed to langue as a system; Hausmann 1984: 398—99 and Hausmann 1985: 118) and is consequently determined by usage. This will be illustrated with three different kinds of word combination.

What types of word combinations are there?

1. True or "significant" collocations

Marilyn Martin has shown that even advanced "quasi fluent" foreign learners frequently produce unidiomatic collocations such as *I was a large smoker instead of I was a heavy smoker. She says, "Like two identical electric charges, the members of an infelicitous collocation repel each other." (132). In the same vein of thought, we might say that the members of a correct collocation are attracted to each other like two opposite electric charges: +A \rightarrow -B. This type of combination has also been referred to in English as being a closed or strong collocation (Sinclair 1987), a settled combination (Cowie 1986), a fixed or recurrent combination (BBI: 4), a restricted combination, in French, une cooccurrence lexicale restreinte (Mel'cuk 1984: 4), and in German, Kollokation (Hausmann) or usuelle Kollokation (Kromann 1989).

Collocations consist of a base or a key word, for example, FEAR, COLERE, VERDACHT, VOPROS and a collocator, e.g. dispel, rentrer, schöpfen, podnjat'. They are thus bipartite units. However, they need not necessarily be made up of two words: for lexicographical purposes it seems useful to include compounds and multiword lexical units having an isolatable base or key word which can be smaller than the orthographic word: they must thus be partially transparent. E.g. WAGE freeze but not "snow job", RAUCHschwaden ("SMOKEcloud") but not "Purzelbaum" ("somersault" — literally "tumbling down-TREE"); CRY one's eyes out, grant DIPLOMATIC IMMUNITY, but not "eat one's heart out". Some collocations are used only in combination with others so that they "overlap", for example, fluent COMMAND and command of a LANGUAGE combine to form fluent command of a LANGUAGE. I shall discuss the notion of base and collocator in greater detail below.

2. Idiosynchratic or literary combinations

Let us consider the following passages from the novels Stone Angel and Surfacing by Margaret Laurence and Margaret Atwood respectively.

Outside the bedroom window a maple grew, the leaves a golden green as the sunlight seeped through them, and in the early mornings the sparrows congregated there to argue, splattering their insults in voices brassy as Mammon, and I'd hear them laugh, liking their spit and fire. (Laurence, p.70)

I thought it was a bad thing to grow up in a house with never a framed picture to tame the walls. (Laurence, p.72)

I swiveled the caps back onto the paint tubes, I had no intention of working. (Atwood, p.120)
You do not usually **swivel** a cap onto a paint tube, you screw it back on, and pictures do not usually **tame** walls — they may grace or decorate them or simply hang on them. Insults usually are not **splattered** but flung or hurled. The reasons for the choice of these collocates lie on the level of pragmatics and the text and not on the lexical-syntactic level. They thus go beyond the level of collocation. In the first passage, the choice of **splatter** is not primarily influenced by “insults”; the collocates **seep** (sunlight) and **splatter** (insults) and the expression “spit and fire” support each other in communicating an impression of fluidity to the reader of the text. In the idiosynchratic or literary combinations there is lexical resistance between the two member elements. The relationship can be seen as two elements carrying the same electrical charge: \(-A \leftrightarrow B\). It has thus been referred to by Hausmann (1984: 399) as a “counter-creation” (“Konter-Kreation”), and by Kromann as a “non-usual collocation” (“nicht-usuelle Kollokation”) because it violates the lexical norm. This is not to say that affinity is totally absent — it is found on a semantic (metaphoric) and textual level.

The privilege of using this type of combination, however, is reserved to the native speaker of a language. A foreign learner using “I swiveled the caps back onto the paint tubes” will be called to order. We see that the line between the idiosynchratic combination and the “discordant”, incorrect collocation is very thin indeed: both violate the norm, but with varying effects.

### 3. Free combinations

Free combinations can be said to consist of two neutrally charged elements. As a result there is neither attraction nor repulsion: \(A \leftrightarrow B\). Free combinations can be created anew at will and do not function as “stage props” of language. They therefore pose no problem to the foreign language learner. Examples of such combinations are:

- **English**: see a house, work quickly
- **French**: voir une maison, travailler rapidement
- **Russian**: videt’ dom, bystro rabotat’

These combinations reflect the system of the language rather than the norm. This type of combination has also been called **open** (Cowie 1981:226), **Ko-Kreation** (Hausmann 1984: 398—99), and **regelrechte Wortverbindung** (“true word combination”; Kromann 1989).

It must be said that what is “free” and what is “restricted” in a foreign language will also be determined by contrasts between a learner’s native language and the foreign language; “a hot day” can be a significant collocation to the beginning French learner of English because in French one does not usually say “c’était une chaude journée”.

The following diagram summarizes the types of lexical combination:
The Component Parts of a Collocation

Since the individual elements of a collocation fulfil diverse functions in a dictionary it will be useful to distinguish between them. This has been done in most of the literature on the topic by distinguishing between a node and a collocate; according to British linguistics, in dispel fear, both dispel and fear can be the node, depending on which word is the object of attention (Sinclair 1974:16—17 and 1988). This does not, however, do justice to the fixed role which each member can play in one and the same collocation: Hausmann as well as Mel’čuk et al. (also Leed/Nakhimovsky 1979, esp. 110 ff. and Martin, both based on Mel’čuk et al.) distinguish between the BASE ("Basis") or KEY WORD ("ključevoe slovo"), e.g. FEAR⁴ and the collocator ("Kollokator"), which corresponds to the value of a particular lexical function in Mel’čuk’s terminology; that is, the morpheme, word or phrase which expresses an aspect of what can be done with the base, or which more closely qualifies it; e.g. dispel, but also arouse, express, mortal, grave, a wave of FEAR, etc.⁵ This orientation of BASE and collocator does not exactly coincide with the hierarchy of word classes as one may assume from reading Pätzold 1987 (cf. p.154). Pätzold criticizes the practice in BBI of lemmatizing to cry one’s eyes out under "cry", and argues that it should be listed under "eyes". But the noun "eyes" is not the base word here — even if it is at the top of the grammatical hierarchy. The word which is more closely qualified, i.e. the base, is the verb CRY. I would like to illustrate the concepts of BASE and collocator further by looking at two approaches to collocations, the semasiological and the onomasiological one.

The Semasiological Approach to Collocations

The semasiological approach to collocations focusses on the collocator (e.g. dispel) and shows how corresponding bases (e.g. FEAR) function to help explain the meaning of collocators. Supplying base words in collocator dictionary entries is thus useful for meaning discrimination in cases of polysemy and synonymy. On the other hand, giving collocators such as dispel in the base articles for DOUBT, FEAR, and CLOUDS does not significantly contribute to the explanation of the words concerned. Semantically speaking, the base is more autonomous than the collocator. It is also less likely to pose meaning problems than the collocator. It can
be learned on its own while the collocator must be learned in collocation. The definition of the collocator is thus incomplete without the syntagmatic dimension of collocation (Hausmann 1979: 192).

Polysemy in the monolingual dictionary
The admirable Slovar Slovenskega Knjižnega Jezika exemplifies the semasiological approach very well. Brief explanations of individual word meanings are supplemented and discriminated from other senses by a wealth of contexts, mainly collocations and free combinations. The verb izpolniti, here a collocator, is defined as follows:

1. See to it that something promised, announced becomes reality, fact. Bases given are: THREAT, ONE'S WORD, A PROMISE, REQUEST, A WISH, ONE'S DUTY, AN ORDER, FATHER'S WILL, A NORM, A PLAN, A TASK, CONDITIONS, EXPECTATIONS, COMPULSORY SCHOOL ATTENDANCE.

2. Write required information, usually in a fixed place, with the bases: REGISTRATION FORM, POSTAL ORDER.

3. See to it that a place is no longer empty, with the bases: A GAP, A POSTING, LIFE.

4. To live up to a certain age, with the base: SIXTY YEARS. 6

I think the bases supplied in the article have illustrated the meaning of izpolniti very well even to those who do not understand Slovene.

Synonymy in the monolingual dictionary
On the advanced and intermediate level of foreign language teaching, words are typically explained by their synonyms. This is also done in monolingual dictionaries for the learner (cf. Jain 1981: 277 ff.). This practice is dangerous in that the student will tend to transfer the typical contextual partners of a word to its synonyms. E.g. redress is a synonym of remedy — you can remedy a SITUATION but you cannot redress it, and rectify is a synonym of correct — you can correct a PERSON but you cannot rectify him (Martin: 133). For effective meaning discrimination, collocator glosses should always support necessary synonym glosses and are to be preferred to synonym glosses in the dictionary.

Accessibility of collocations in dictionaries
An analysis of collocations listed in some 270 articles of 3 English learners' dictionaries carried out by Andrea Arican has shown that they are most often found in collocator entries (e.g. pursue STUDIES under pursue). We have seen that this practice explains the meaning of the collocator rather than the use of the base in context. It seems then, that meaning and the reception of texts (i.e., the semasiological approach) is still given priority over the production of texts.

Questions such as “what can you do to STUDIES?” are not systematically answered. So if foreign learners do not know that “pursue” is the partner needed here, they will have to look it up in an L1—L2 bilingual dictionary where they are faced with the problem of choosing the correct equivalent of the source language collocator which they look up.

Native speakers cannot resort to a bilingual dictionary. They will ask: what can you do to a CAREER?, what do you do to a CAVEAT? how do you express the
notion of a QUESTION which is disputable? They will need to find pursue in the entry for CAREER, enter and insert in the entry for CAVEAT, and moot in the entries for POINT and QUESTION. This belongs to the onomasiological approach to collocations.

The Onomasiological Approach to Collocations: The Monolingual Dictionary

This approach serves text production and has the base as its starting point rather than the collocator. The base represents what a person is writing or talking about. Someone writing a text about a QUESTION or POINT about which there can be disagreement may be looking for a way to elegantly contextualize QUESTION — he or she might find disputable on his or her own and be seeking a source of inspiration for other possibilities, or — and this does occur more often than one may wish when one is writing a text, especially in a foreign language — he or she may need a list of contextual partners to jog his or her memory or spark his or her imagination; to this aim, users will need to find collocators in the base entries QUESTION and POINT, for example, moot. The collocation moot QUESTION entered under moot is hidden from the writer of a text seeking a suitable partner for QUESTION. The text writer may require further collocators at some other point in his or her text such as raise a QUESTION, ponder a QUESTION, a QUESTION pops up, etc. He or she must be able to find these all under QUESTION and related words and not scattered amongst individual collocators.

Consequences of the Two Types of Approach for the Learner

Learners acquire vocabulary both semasiologically — e.g. by receiving texts — and onomasiologically — e.g. by producing texts. This is why learners can benefit from both types of access to collocations. In systematic learning, however, it seems less complicated to learn a series of collocators which can be used with one or more bases than to try and learn which bases can be used with a certain collocator: this would involve learning the entire polysemy structure of a word which amounts to systematically ploughing through the dictionary article for the word. The result is a series of collocations the bases of which diverge too widely in meaning to be effectively anchored in the learner’s memory; the collocations which the article ‘pursue’ may yield could look like this:

pursue 1 ———> A FUGITIVE
pursue 2 ———> A PERSON
pursue 3 ———> A DESIRE
pursue 4 ———> A PLAN, POLICY
pursue 5 ———> STUDIES, A HOBBY
pursue 6 ———> A MAN
pursue 7 ———> A POINT

ILL HEALTH/BAD LUCK
There is no constant here to facilitate learning. Also, questions may arise as to additional bases which can be used with the collocator. E.g. if I can pursue a plan, can I also pursue a plot or a scheme?

There is no doubt that it is important to learn which bases can be paired with a range of collocators and when collocations overlap, but this is only helpful when the number of collocations involved is fairly limited. You can, for instance perform or conduct a TEST, and an AUTOPSY, and you can conduct but not perform a SURVEY. This approach can quickly become somewhat bewildering as the following diagram taken from Cowie 1986 shows:

![Diagram showing collocations and bases]

It would seem more productive to learn what you can do with words than what you cannot do with them. The advantage of learning a series of collocators which can be used with a particular base is that the base is a lexical-semantic constant, for example QUESTION:

<table>
<thead>
<tr>
<th>moot</th>
<th>question pops up</th>
</tr>
</thead>
<tbody>
<tr>
<td>raise</td>
<td>a</td>
</tr>
<tr>
<td>ponder</td>
<td>a</td>
</tr>
</tbody>
</table>

The onomasiological approach can also begin from a particular lexical function. For example, how do you express the concept 'INCEP', i.e. the beginning of an action, with a number of related bases. This too, may become complex because the constant is an abstract lexical function which could be realized in a particular language in a multitude of ways. This is more complicated to learn and more difficult to present satisfactorily in a dictionary:
The Semasiological and Onomasiological Approaches Combined: The Case of the Bilingual Dictionary

As we have seen, it is useful to list collocations in the collocator entry for text reception and in the base entry for text production. This applies to the monolingual dictionary where access to collocations is only possible via the target language. This means that at least one of the elements of the collocation must be known.

Monolingual dictionary — text reception: entry dispel — dispel FEAR
— text production: entry fear — dispel FEAR

In the bilingual active dictionary (L1—L2) the situation is somewhat more complex because access is only possible via the system of the source language, and because two language systems are matched, causing problems of anisomorphism.

For text production the foreign user (e.g. with French as L1 and English as L2) seeking access to the English collocation 'dispel fear' could look it up under either of the two elements — the base ('peur') or the collocator ('dissiper') in the L1—L2 side of their dictionary. The problem of presenting collocations in bilingual dictionaries has been discussed in detail by Hausmann (1988: 148—151) and Kromann 1989 so I shall not go into that here. I would like to look at one specific problem for the active dictionary that arises from an overlap of semasiology and onomasiology.

Suppose a German-speaking user would like to express "einen gemütlichen Schwatz" in English and looks it up in an L1—L2 dictionary under gemütlich. He or she may find the information given in ex. 1 below:

1. gemütlich (Schwatz, Beisammensein, etc.) cosy
2. gemütlich cosy (Schwatz, Beisammensein)
3. gemütlich (bequem, behaglich) comfortable
4. gemütlich (Schwatz, Beisammensein) cosy (chat, get-together)
5. gemütlich (chat, get-together) cosy
6. Schwatz chat, chinwag (no listing of collocations)

The L1 bases SCHWATZ and BEISAMMENSEIN are given in the collocator article, gemütlich to show the German-speaking producer of a text which sense of German 'gemütlich' is covered by the corresponding equivalent, 'cosy'. This exemplifies the semasiological function.

This procedure is also safer than giving synonym glosses (see ex. 3). One can say 'eine behagliche (gemütliche) Atmosphäre' but not "a comfortable atmosphere". Looking back at ex. 1, the bases, SCHWATZ and BEISAMMENSEIN serve not only as meaning discriminators but also as context partners. They thus serve a dual function in the bilingual active dictionary. Only they are supplied in the wrong language to adequately fulfil their function as context partners: The collocation is neither 'a Schwatz cosy' (ex. 1) nor 'a cosy Schwatz' (ex. 2)! Looking under SCHWATZ in the German-English side of the dictionary (see ex. 6) will not necessarily help in choosing the correct base to collocate with cosy. This is because in German and English bilingual dictionaries collocations are often listed only in collocator articles (as in ex. 1), less often in base articles (e.g. Schwatz) and only rarely in both for obvious space-saving reasons. The choice must be made at
Schwatz between chat and chinwag, but the more usual combination is ‘a cosy chat’. How is the learner to know this? The entire collocation should have been listed in the relevant language in the first place (as in ex. 4 and 5).

Anisomorphism of Languages

What further complicates the accessibility of collocations in bilingual dictionaries is the anisomorphism between languages. A collocation in one language may not correspond to a collocation in another. This means that in a bilingual active dictionary only such collocations of the target language are accessible as have a corresponding source language expression. In English, a building is ablaze or in flames. In German, it is possible to say, das Gebäude brennt, or das Gebäude steht in Flammen. But much more idiomatic and thus more colorful is das Gebäude brennt lichterloh or das Gebäude steht in hellen Flammen; there is no English word corresponding to lichterloh and hell under which one could list these collocations in an English-German dictionary. These collocations are thus not accessible via the net of the English language. We are confronted with the age-old problem resulting from filtering the target language through the lenses of the source language.

This does not mean that the “necessary evil” — the bilingual active dictionary should be done away with completely. It simply means that users must be sensitized to the capabilities and limits of this type of dictionary. They should also be introduced to the usefulness of the specialized collocation dictionary with L2 base entries — especially for text production. The type of strategy which the user must learn is described by Leed and Nakhimovsky as follows:

Take the problem [. . .]: how do you translate ‘place an order’ into Russian? The proper question to ask is not “What is the Russian for ‘place’?” but “Given that ‘order’ is zakaz in Russian, what do you typically do to a zakaz (110)

Concluding Remarks

To sum up, the following must be considered when compiling dictionaries of collocations:

Only “relevant” or “true” collocations should be included. Whether or not a collocation is relevant will be determined by the target group.

Target group

Native speakers are likely to search for stylistically elegant collocations such as enter or insert a CAVEAT, allay or dispel FEAR or a moot point. This justifies the presence in the BBI Combinatory Dictionary of deliver o.s. of an OPINION which has been judged by one reviewer as “too pompous” for the learner. Native speakers do not need combinations such as “a hot day” and “make a mistake”. They also may wish to avoid overused cliché-type collocations such as “to garner compliments”.

In addition to stylistically elegant collocations, foreign learners will require such collocations as differ from the ones in their own language: This could involve such simple combinations as make a mistake, commit an error and not *perform a
mistake as would be possible in Russian (soveršat' ošibku). Collocational dictionaries should thus ideally take into consideration the native language, and the foreign language competence of the target group.

Activity

For reception purposes only non-transparent collocations should be entered and listed under the collocator; therefore, 'make a mistake' would not need to be entered, and 'moot question' would be listed under moot. For text production it is advisable to list relevant collocations under the base (i.e. under question). Because listing collocations doubly can be space-consuming one could consider providing access via collocators by means of cross-references (e.g. from moot to question).

Type of dictionary

The L1—L2 translation of collocations should be left to general, active bilingual dictionaries. Defining collocators could be left to the dictionary of definitions. Only the specialized, monolingual or passive bilingual collocation dictionary can deal with collocations for text production and for learning in a satisfactory manner. To avoid complication in bilingual dictionaries such as the anisomorphism of languages and the dual function of the collocator there should be an L2 macrostructure and the lemmata should be bases; in this way collocations are supplied instead of being translated and they are readily accessible. Two examples of such existing dictionaries are The BBI Combinatory Dictionary of English and the series of Soviet dictionaries of Russian word-collocations which have been published in Russian-Czech, Russian-English, and Russian-German. These dictionaries could be improved by taking the language of the target group into consideration and by making the collocation accessible via the collocator by means of cross-references. In this way they could better serve the needs of the learner who learns collocations by both the semasiological and the onomasiological approach.

References

Cited Dictionaries


Denisov, P. N. and V. V. Morkovkin (eds.). 1983. SLOVAR’ SOCETAEMOSTI SLOV RUSSKOGO JAZYKA. Moskva: Russkij jazyk.

DUDEN. STILWÖRTERBUCH DER DEUTSCHEN SPRACHE. Die Verwendung der Wörter im Satz. 1970. Mannheim; Wien; Zürich: Bibliographisches Institut. (Der groe Duden; Bd. 2).


Other Literature


Bolinger, Dwight. 1966. 'Meaning and memory' in Forum linguisticum 1; 1—14.


Korosadowicz-Struzyn'ska, Maria. 1980. 'Word collocations in FL vocabulary instruction' in *Studia Anglica Posnaniensia* 12: 109—120.


**Notes**

1 Various aspects of this broad term (which means "word combination") have been described by Apresjan 1969: 81 and Morkovkin 1984. Ter-Minasova 1981: 10 refers to "leksiko-frazeologičeska sočetaemost' slov" or "kollokacija".

2 Ter-Minasova 1981: 7 speaks here of "metašemiotičeskij charakter slovosočetanija" or the meta-semiotic character of word combinations.

3 In fact, even only to the "authors" of that language: an interesting phenomenon of the past is the great popularity amongst native speakers of collocation dictionaries based on word combinations used by respected authors. An example is Maurice de la Porte's *Les Epithèthes* of 1572, based on an established Latin tradition (cf. Hausmann 1982: 187ff.).

4 In Russian literature one also finds the terms "glavnoe slovo" (main word), "oporne slovo" (supporting word) and "glavnij c)len" (main member).

5 This has also been called "leksiceskoe napolnenie" or lexical complement .Galisson (1981:14) speaks of "terme-noyau" for the base and "terme-satellite" for the collocator.

6 The structure and content of the article have been simplified and translated into English.

7 Hausmann (1988:151) shows how this problem could be solved by using a cross-referencing system.

8 This explains the title „Stilwörterbuch“ for the Duden dictionary containing word combinations and idioms, and which has nothing in common with the English dictionaries of style.