Ton van der Wouden, Vakgroep Nederlands, Rijksuniversiteit Groningen

Prolegomena to a Multilingual Description of Collocations

Abstract

We discuss collocation from the perspective of the non-native speaker. Bilingual dictionaries fail to pay due attention to the phenomenon in a systematic way. A broader definition of collocation is argued for and it is shown that, from such a broader perspective, the very same semantic content may be expressed alternatively, but still collocationally, with syntactical, morphological or lexical means. Thus, a treatment of collocations based on conceptual rather than grammatical principles is called for.

1.   Introduction

Collocation is a pervasive phenomenon in natural language[1]. Between command of a given language and fluency in that same language is a large gap, a yawning abyss formed by idioms, collocations and other fixed combinations that have to be learned and mastered. English learners' dictionaries (such as LDOCE, Cobuild) and specialized dictionaries of collocations and idioms (e.g. BBI, 1986, Cowie and Mackin, 1975, Cowie et al., 1983, Lipton, 1991), offer a large amount of collocational information. Although they never reach the level of completeness or exhaustiveness Mel'cuk et al. (1984) aim for, they are extremely helpful both for the student of English as a foreign language and for the translator into English. Thus, they help us e.g. in selecting the typical activity verb with OATH (to administer, to break), the correct intensifying adjective for TEA (strong), the collective noun with CROWS (murder) or OWLS (parliament), etc.

In the remainder, we use the term COLLOCATION as the generic term for idiosyncratic restrictions on the distribution of lexical material.[2] We will show below that, from the perspective of such a broad definition, dictionaries in general are far less detailed as regards collocational combinations as one would wish, if any systematic attention is paid to them in the first place.[3] General and learners' dictionaries may help the second language learner in some parts of the collocational area, such as verb complements or prepositional selection, but they are totally
incomplete in other parts. The same holds, to only a slightly less extent, for the specialized collocational dictionaries just mentioned. Bilingual dictionaries hardly ever give collocational information in a systematic way.

If we compare the way collocations 'work', so to speak, in different, although closely

related languages, various suggestions with respect to the optimal treatment of collocations, especially in bilingual dictionaries, dictate themselves. In the next section we will take a brief look at collocations from the perspective of the foreigner learning English (or translating into English), who has to cope with the problem of using the right combination of words in the right place. In the following section some brief attention will be paid to Mel'cuk's theory of (or descriptive framework for) collocations. In the fourth section, we discuss some of the differences that become manifest if one compares collocations in various languages. In the final section, conclusions are drawn with respect to the inclusion of collocational data in bilingual and other dictionaries.

2.   The relevance of the study of collocations

The proper use of the right collocation in the right place is an indispensable aspect of flawless command of a language. To quote from Greenbaum (1970: 83 ff.):

"From his experience of the language, the native speaker of English becomes aware whether a collocation he produces or encounters is habitual or occasional or unique and deviating from everyday language. The non-native speaker needs to learn in the first place the habitual collocations. [...] He may easily come to form aberrant collocations by limiting himself to one degree intensifier, perhaps MUCH, with or without the premodifier VERY. Behre notes that in Agatha Christie's writings, MUCH is frequently used in dialogues by foreign characters, 'who anxiously stick to it' [...]. It may be supposed that Agatha Christie is imitating the unidiomatic use of MUCH by foreign speakers of English."

If an English student or translator doesn't want to sound like a foreigner, wishing to avoid falling in the trap described by Greenbaum, he'd better know his fixed combinations. For this purpose, he may try and look them up in the kind of dictionaries referred to in the introduction, or learn them by heart from specialized drilling books. However, much of the information he is looking for is just not in these books. Bilingual dictionaries aren't of to much help either, as they likewise too often fail to supply all the information our hypothetical student is looking for. Moreover, the collocational combinations that are listed may be hard to find for the foreign user. Still, the desires the non-native user may entertain towards dictionaries, especially bilingual ones, with respect to a more systematic treatment of collocations are justified. Moreover, in our view there is no reason for despair: in principle, theories and formal means do exist that allow for a much more systematic and comprehensive treatment of collocations.[4] The only sacrifices one has to make are financial on the one hand, and in one's conception of the way a (bilingual) dictionary should look on the other.

3. Semantic typing

Recently, a whole new type of dictionaries has been presented to the world. The ECD (Explanatory Combinatory Dictionary) designed by Mel'cuk is set up as a monolingual dictionary, be it of a rather onorthodox type. A central feature of this dictionary is the usage of so-called 'lexical functions', a finite set of functions in the mathematical sense, representing certain extremely general ideas, such as 'very', 'begin' or 'implement', or else certain semantico-syntactical roles (Mel'cuk and Zolkovsky, 1984). The lexical functions serve two uses within the Meaning-Text Theory (Mel'cuk et al., 1981, Mel'cuk et al., 1984, Mel'cuk and Polguère, 1987). One is to control the proper choice of lexical items, whereas the other is to describe sentence synonymy. The latter is done by describing a number of equivalences in terms of lexical functions. Such paraphrasing rules account for the equivalence of e.g. ''Mary answered my question to satisfaction'' and ''Mary gave a satisfactory answer to my question''.

In Heylen et al. (1991) the following is claimed concerning Mel'cuk style lexical functions:

''these functions are impressively successful in covering a large number of candidate collocations. [...] What they offer is not JUST a notation for collocations. In effect, they imply that there are a relatively restricted number of central collocation types (currently around 15) and hence the collocations a given word enters into can be more-or-less exhaustively typed. Note that in ordinary dictionaries collocations are NOT typed.''

Collocations can very often be identified with semantic operations (possibly via Mel'cuk-style functions), which is of great significance for the structure of the dictionary indeed. These lexical functions moreover apply very fruitfully to many more cases of collocation than is commonly seen, if only we define collocation in the broad sense we did in section 2.

The simplest method to translate collocations as collocations is to lexicalize them directly in the bilingual dictionary:

English: pull the trigger =
French: appuyer sur la détente =
Dutch: de trekker overhalen

English: sleep deep =
French: dormir profondement =

3

Dutch: diep slapen

English: stark naked =
French: nu comme la main =
Dutch: poedelnaakt

Under assumption of this direct treatment, successful translation does not presuppose a typing of collocations. However, such a typing is desirable; it opens up the way to a systematic treatment of collocations and hence to a substantial reduction in the number of collocations explicitly handled in the dictionary coding (a costly task, both in terms of time and of money). In some way or another, the existence of a collocational relation (a collocation function) between a head and its collocands is established during analysis; in some way or another this information is used to synthesize the correct translation in the target language.

English string: the whole truth
English semantic representation: truth + VER
French semantic representation: vérité + VER
French string: toute la vérité

In this example, the VER diacritic is cashed out by the French ''toute'' (which is not, of course, the default translation for ''whole'').

Dutch string: poedelnaakt
Dutch semantic representation: naakt + MAGN
French semantic representation: nu + MAGN
French string: nu comme la main

In this more complicated example, the MAGN diacritic is expressed morphologically in the compound POEDELNAAKT in Dutch, 'poodle-naked', whereas it is expressed by syntactic means in a prepositional modifier COMME LA MAIN, 'like the hand' in French. This is, to our knowledge, the only case where Dutch ''poedel'' is properly translated as ''comme la main''.

Given that lexical functions can be composed, some of these compositions can turn out to be equivalent in meaning to other functions or compositions. These equivalences may also be used in generating paraphrases, or in generating equivalent translations.

4.   Contrastive study of collocation

As Zgusta (1971: 294) writes,

"The basic purpose of a bilingual dictionary is to coordinate with the lexical units of one language those lexical units of another language which are equivalent to their lexical meaning. [...] The fundamental difficulty of such a coordination of lexical units is caused by the anisomorphism of languages, i.e. by the differences in the organization of designate in the individual languages and by other differences between languages."

This anisomorphism problem manifests itself most strikingly on the borderline between grammar and lexicon: languages do not divide equally the information between basic expressions and rules of composition. A certain meaning may be expressed alternatively as a single lexical item or as a compound (Dutch ''schimmel'' vs. English ''white horse''), or as a monomorphemic element or an idiom (English: ''to die'' = ''to kick the bucket''). To be more specific, sometimes a lexical function is expressed as a free combination, sometimes as a collocational combination of morphemes, words and/or phrases.

Stated in different terms: it is, given their very general nature indeed, intuitively plausible that the lexical functions Mel'cuk and his group have been studying are universal cross-linguistically. However, the way they are implemented across the various languages of the world may differ dramatically.

Ttraditional lexicographic methodology, using an orthodox, narrow definition of collocation, has, as far as we know, failed to see that collocations almost always translate into collocations, if only the scope of the definition of 'collocation' is extended. If one adopts e.g. a definition such as the one given in section 2. above, it is very easy to see that a lexical function that is expressed collocationally as a word combination in one language may be expressed collocationally with morphological or syntactic means in a language that is very closely related typologically.

A case in point form some subgroups of the collocations consisting of an adjective plus an intensifying adverb in English (the L6 type of Benson et al. (1986)), that are often best translated into Dutch by means a noun-adjective compound. I.e., the same lexical function MAGN (the general function of intensifying, 'high degree') is expressed syntactically in English but morphologically in Dutch. To wit:

"Dutch is [...] not alone in having adjective-specific intensifiers; it is however unique in several aspects of their formation and use. Particularly striking to me is their pervasiveness in the language. For many adjectives, the specific intensifier is used almost exclusively, so that, idiomatically speaking, HEEL or ERG [the unmarked adjective intensifiers in Dutch. TvdW] are virtually wrong. For example, I recall one incident in which I was recounting

something to a native speaker of Dutch; as I told my story, he was echoing me, unconsciously adapting my wordings into a more normal variety of Dutch. Adjectives which I intensified with HEEL or ERG frequently emerged with specific intensifiers in his version; ERG DUUR 'very expensive' he echoed with PEPERDUUR 'pepper-expensive'." (Fletcher 1980, 447)

Below, we list some examples of this semi-productive word formation process, in which the irregularities in the English translations are noteworthy.

Dutch	'English'	English
bloedheet	'blood-hot'	(sweltering) hot
ijskoud	'ice-cold'	icy cold
peperduur	'pepper-expensive'	very expensive
poedelnaakt	'poodle-naked'	stark naked
spinnijdig	'spider-angry'	(as) cross as two sticks
stekeblind	'pinch-blind'	stone-blind
stokdoof	'stick-deaf'	stone-deaf
stomdronken	'mute drunk'	dead drunk
stomtoevallig	'mute accidental'	accidental
stomverbaasd	'mute amazed'	astonished
stomvervelend	'mute boring'	deadly dull

This very same lexical function MAGN of Mel'cuk et al. (1984) may, on the other hand, be expressed in various types of verbal collocations both in English and in Dutch. This is illustrated by the following listing of collocations given in the lemmata LIEGEN ('to lie') and REGENEN ('to rain') of Martin et al. (1986).[5]

	Dutch			English	
hij loog	dat hij scheelzag		he lied	in his teeth
'he lied'	that he squinted'			in his throat
	(als)of het gedrukt stond			
	'as if it were printed'			
	dat hij blauw zag			till he was blue in the face
	'that he looked blue'			
	dat hij zwart werd			till he was black in the face
	'that he became black'			
	dat hij barstte			
	'that he cracked'			

het regent	dat het giet	it's raining	cats and dogs
'it's raining that it poors'			
het regent	pijpestelen	the rain is coming down in buckets	
	'pipestems'		

Whereas 'collocation' is commonly interpreted as involving idiosyncratic restrictions on the combination of words or groups of words, examples such as those given above show that the restriction to the word level is too narrow: we find collocational relations between elements both below the word (the case of POEDELNAAKT and PEPERDUUR) and beyond that level (sentences and subclauses as in the table above).

For a language learner (or a translator, for that matter), organization of collocation types according to the grammatical categories involved (as chosen by e.g. BBI (1986)) is not very useful, as the user will often not know where to look for the information needed. Because of the anisomorphism even of the closely related languages English and Dutch, the way a given concept (lexical function) is implemented in the language learners' native tongue (or in the source language in the case of translators) is not of great help in finding the collocational item looked for in the target language, because of the fact that it may belong to a different grammatical category.

On the other hand, a semantic TYPING of collocations in the dictionary seems to fit more closely with the strategies the translator or foreigner use. The most well-known example of such a semantic organization may of course be found in ECD like dictionaries such as Mel'cuk et al. (1981). This type of book of reference seems to be more useful IN PRINCIPLE for this specific type of dictionary user.

Until now, however, bilingual dictionaries of the ECD-type do, to the best of our knowledge, not exist; only tiny fragments of languages have been described monolingually by the Mel'cuk group. Moreover, given the enormous amount of time it takes to compile dictionaries of this level of exhaustiveness, we may not expect that a complete bilingual dictionary of this type will be published in our lifetimes; perhaps the investigations into the possibilities of building a specialized editor for this type of dictionaries (as described in Décary and Lapalme (1990)) might speed things up considerably. If such a bilingual dictionary would be finished and published, it would be extremely expensive and of an unmanageable size. New media for data storage, such as CD-Rom, come to mind for solving the size problem.

A final question one should raise concerning bilingual dictionaries structured according to the ECD principles is, whether an ordinary user will be able to fruitfully use such dictionary? It seems to call for a level of semantic analysis that t be too high for a non-linguist to be mastered. It is to be expected, however, that the complete formal apparatus of the ECD-dictionaries in their current form is not needed for a popular dictionary. Furthermore, the new media for data storage hinted at in the paragraph above allow both for user-friendly and flexible access, such as menus and graphical user interfaces on top of complex data base management systems.

5.   Conclusion

From the comparison of two closely related languages discussed in this paper the following conclusions with respect to the treatment of collocations in dictionaries may be drawn:
- there are more collocations than fit into the traditional lexicographer's philosophy;
- in bilingual dictionaries, more attention should be paid to collocational information;
- organization of collocational dictionary information according to conceptual principles is more useful than organization according to grammatical principles;
- in a bilingual dictionary that pays attention to collocational dependencies, the word level is not the only relevant level: both morphological and syntactical information will have to be included;
- new media for data storage will be more useful for comprehensive user-friendly and user-manageable bilingual dictionaries than the classical paper form.

Bibliography

BBI (1986): BENSON, M., E. BENSON, and R. ILSON (1986): The BBI Combinatory Dictionary of English: A Guide to Word Combinations. Benjamins, Amsterdam/Philadelphia.

Cobuild: Collins Cobuild English Dictionary (1987). Collins, London.

COWIE, A.P., and R. MACKIN (1975): Oxford Dictionary of Current Idiomatic English. Vol. 1. Oxford University Press, London.

COWIE, A.P., R. MACKIN and R. MCCAIG (1983): Oxford Dictionary of Current Idiomatic English. Vol. 2. Oxford University Press, London.

DÉCARY, M., and G. LAPALME (1990): "An editor for the explanatory and combinatory dictionary of contemporary french (DECFC)." Computational Linguistics 16, 145-154.

FLETCHER, W.H. (1980): "'BLOOD-HOT', 'STONE-GOOD': a preliminary report on adjective-specific intensifiers in Dutch." Leuvense Bijdragen 69, 445-472.

GREENBAUM, S. (1970): Verb-intensifier collocations in English: an experimental approach. Mouton, The Hague [etc.].

HEYLEN, D., L. BLOKSMA, T. van der WOUDEN, S. WARWICK-ARMSTRONG, R.L. HUMPHRIES, and S. MURISON-BOWIE (1991): "Collocations and the Lexicalisation of Semantic Operations." Ms. Eurotra Utrecht.

LÉPINETTE, B. (1989): "Vers un dictionnaire explicatif et combinatoire bilingue: bases théorique et élaboration de quelques articles." Cahiers de Lexicologie LIV, 105-162.

LDOCE (1987): Longman Dictionary of Contemporary English. Second edition. Longman House,

LIPTON, J. (1991): An Exaltation of Larks. The Ultimate Edition. Viking.

MARTIN, W., G.A.J. TOPS, A.P.A. BROEDERS, P.H. van GELDEREN, L. ROOS, and M.H.M. SCHRAMA (1986): Van Dale Groot Woordenboek Nederlands-Engels. Van Dale Lexicografie, Utrecht/Antwerpen.

MEL'CUK, I.A., and A. POLGUÈRE (1987) "A formal lexicon in the meaning-text theory." Computational Linguistics 13.

MEL'CUK, I.A., and A.K. ZOLKOVSKY (1984): Explanatory Combinatorial Dictionary of Modern Russian: Semantico-Syntactic Studies of Russian Vocabulary. Wiener Slawistischer Almanach, Sonderband 14. Vienna 1984.

MEL'CUK, I.A., L. IORDANSKAYA, and N. ARBATSCHEVSKY-JUMARIE (1981): "Un nouveau type de dictionnaire: Le dictionnaire explicatif et combinatoire du français contemporain." Cahiers de Lexicologie, 38, 3-43.

MEL'CUK, I.A., N. ARBATSCHEVSKY-JUMARIE, L. ELNITSKY, L. IORDANSKAYA, and A. LESSARD (1984): Dictionnaire explicatif et combinatoire du français contemporain. Les Presses de l'Université de Montréal, Montréal.

van der WOUDEN, T. (1991): "Restricties op de distributie van lexicale elementen", Ms. OTS Utrecht.

ZGUSTA, L. (1971): Manual of lexicography. Academia and Mouton, Prague/The Hague - Paris.

Keywords

collocations, cross-categoriality, second language acquisition, Mel'cuk

Endnotes

1. The research described here is partly based on joint work with the authors of Heylen et al. (1991); thanks are due to them and to Martin Everaert and Henk Verkuyl. All errors are my own. The financial support of the European Community and the NBBI through the EUROTRA project, and of NWO, the Dutch organisation for scientific research, is gratefully acknowledged.

2. Cf. van der Wouden (1991).

3. Cf. Lépinette (1989).

4. Cf. Lépinette, o.c.

5. Horizontal alignment of Dutch and English collocations is not intended to suggest that these collocations are each other's optimal translation.