

Dorota OSUCHOWSKA

COMPILING A DICTIONARY OF COLLOCATION: PROBLEMS AND SOLUTIONS

Introduction

The aim of this paper is to describe some practical problems met and addressed during the production of *English at Work. An English-Polish Dictionary of Selected Collocations*. Started in the spring of 1999, the book is primarily aimed at the constantly growing number of people who use English for work purposes. It is the first and so far the only English-Polish dictionary of such a specific kind and scope.

For someone working in the English-speaking environment making mistakes is a luxury those professionals know they cannot afford. Firstly, it marks them out as language learners (and therefore incompetent) in confrontation with their foreign employers, the majority of whom happen to be native speakers of English. Secondly, it presents a sharp contrast with the good command of English they have declared. Thirdly, they have spent years learning English, invested a lot of their time and money in it, attended Business English classes; their accent and command of grammar may be irreproachable; and yet, without and sometimes in spite of long-time immersion in naturally occurring language, the English they produce differs substantially from English used by their foreign colleagues.

Collocability and a non-native language user

Although the theory of collocation has been dealt with in detail on numerous occasions¹ and lies outside the scope of the present paper, we should, perhaps, try to single out a couple of factors that are, on the one hand, closely connected

¹ For general discussion of collocations see Cowie (1981b), Mackin (1978), Mitchell (1971). For collocational restrictions, see Allerton (1984).

with the very nature of this linguistic phenomenon and the way it is represented in modern lexicography but which are, on the other hand, also responsible for the difficulties an average learner encounters in incorporating them into his everyday use.

Let us first look at the way collocational information is represented in pedagogical lexicography. It is true that, because of the growing awareness on the part of lexicographers and textbook authors of the potential problem collocability may pose to a non-native user, dictionary definitions in the new generation of dictionaries include typical collocates of a word, usually taken from the corpus of real language material. The introductory sections of learner-oriented monolingual dictionaries abound in statements emphasising the importance of, as Sinclair (1999:viii) puts it *the details of a natural use of a word*.

Due to space restrictions, however, the most such dictionaries can offer is one, or perhaps two examples of the headword, not necessarily connected with the particular context a user is interested in, to say the least. In other words, the user will only find a dictionary helpful in the task of interpreting an English text. Now, the sheer number and diversity of lexicographic publications appearing every year clearly shows the English language as the leader among languages used for international communication and we can therefore expect that more and more users will be interested in dictionaries helping them produce rather than interpret and, from the point of view of such a user, the information is simply not there.

Another factor responsible for the learner's difficulty in dealing with collocational information that can clearly be observed at both elementary and more advanced levels is what the late Dzierżanowska (1982:5) calls the *interference effect of the collocations in his own language*, a phenomenon meaning that collocational behaviour is carried over from the first language and translated into the target language. A Polish student, familiar with the phrase *commit murder/suicide* may be tempted to say that someone has *committed a book/a song/a poem* and see nothing wrong in it; *social care system* instead of *welfare system*; *agricultural reform* instead of *land reform* and other examples too numerous to be cited. An interesting observation frequently made and reported by colleagues with long-term classroom experience is that while most students gratefully and unquestioningly accept being corrected when the phrase in question is a grammatical collocation, especially one containing a preposition, they seem to instinctively reject being wrong in the case of lexical collocations. Phrases like *make an estimation* or *hearty/hot regards* sound perfectly normal to them and frequently become a bone of contention between the learner and the teacher, the former supporting his case with the fact that he had found it in a dictionary.² Moreover, a learner encouraged from an early stage to form

² Stanisławski lists *hearty* as one of the English translation equivalents of *serdeczny*, but does not inform the user about the adjective's combinatory potential.

sentences by analogy is not readily going to accept the fact that the same does not apply to phrases; if *smoking compartment* or *sleeping bag* are possible, why should not *teaching room* be acceptable?

The third important factor responsible for the learner's failure to successfully produce these, as Douglas Kozłowska (1998:13) puts it: *larger bricks of language* that collocations are seems to be closely connected with the very commonness of many of these phrases. The very notion of being common can be understood as referring to how frequently the phrase in question is represented in the learner's textbook and also to the fact whether they are capable of understanding the phrase without referring to a dictionary. The phrase *golden beaches* seems to be ideally suited to illustrate our point here since it appears in a couple of reading passages on tourism in intermediate level textbooks and yet a learner is almost sure to remember its first component as *gold*. Teachers confirm that learners display this tendency to dismiss such common phrases in favour of sometimes highly idiomatic, eye-catching expressions that, statistically, they are much less likely to need.

These three above mentioned factors, that is: a) lack of consequence or often the utter failure in representing a lexeme's combinatory potential in the workplace context in both general-purpose and specialised monolingual dictionaries; b) the difficulty in producing collocations due to the interference effect of one's own language; c) the fact that a prevailing number of collocations are easily understood by the learners and, as a result, dismissed by them as not worth remembering, were the most important reasons for writing the book.

Collection of data

The work started in June 1999 and was completed in December 2000. All headword data comes from materials privately collected by the author: authentic work resumes, letters of application, handouts, guidelines and brochures written for people first entering or re-entering the job market, notes of recordings of counselling sessions and workshops created for people temporarily out-of-work, transcripts of radio and television programmes as well as newspaper articles dealing with various issues connected with employment and finally, notes of recordings of actual job interviews. Thus, the underlying text corpus consists of materials created not with a learner in mind but by and for people whose first language is English and as such ensures the user's immersion in a language variety which is probably less patterned than the type found in the usual class situation.

The endeavour was not at any stage supported by any institution or individual and the techniques of material collecting and recording available to the author must appear crude when compared with those available to the

established names in dictionary making these days. Nor is the size of the book comparable with unabridged learner's dictionaries like *Cobuild Collins English Dictionary* or *LDOCE*. No attempt was ever made to critically look at the data being collected in order to detect and eventually correct potential faults in the data. On the contrary, it was a conscious decision from the very beginning that the original data was not to be in the slightest way distorted, adapted or tampered with, which includes spelling adoption or drawing a demarcation line between technical terms, idioms, jargon or clichés and collocations. I decided on a pragmatic approach as opposed to a theoretical approach which meant that every phrase encountered was to be recorded in the same manner one would present it to a learner, although some of them were not collocations but free word combinations as is the case of *aware of* and *be aware of*.

Editorial process

The headwords were recorded together with their collocates and arranged in alphabetical order. At this stage it became clear that, at least in the case of the source material gathered, certain verbs showed a greater collocational potential than other parts of speech and were probably going to constitute the core of the future dictionary. For verbs with high collocational potential their noun and adjective counterparts (if existent) were added. It should be noted that all collocates that accompany those two latter categories came from the source material and **not** from other dictionaries; they were all at some stage encountered in the materials and added later.

Simultaneously, part-of-speech labels and English synonyms were being appended to the headwords. At this point the material was still being collected for the author's private use and adding the synonyms was primarily seen as introducing, as Clark (1988:x) puts it: *some variety to relieve what would otherwise be a limited and ultimately boring vocabulary*.

In the spring of 2000 the material, which now constitutes the first part of the book, namely an alphabetical listing of key English words, identified as parts of speech, and their synonyms together with a list of their collocates was presented to *Wydawnictwo Naukowe PWN* (henceforth *PWN*), who expressed their interest in publishing it as a bilingual English-Polish dictionary of workplace-related collocations. From that moment on, all steps undertaken resulted from the co-operation between the author and the publisher. *PWN* suggestions included: 1. supplying the collocations with a Polish translation, since most of the English words have a number of meanings; 2. supplying the dictionary with a Polish index for ease and speed of access; 3. adding an English index to make it a bidirectional dictionary and thus potentially useful for native speakers of English.

A new user profile thus emerged: a relatively advanced learner of English, familiar with practically all the English words in the dictionary but unsure about their behaviour in the workplace context; someone either currently enrolled in a Business English course and wanting to extend the vocabulary offered by their regular coursebook; Polish teachers of BE for the reason given above; native speakers who teach BE and may need some guidance in terms of what poses the greatest problem for their students; people preparing for a career with an English/American company, both in Poland and abroad and finally, professional people already working in an English-speaking environment. This new user profile was instrumental in the author's decision to include some extra information, for instance concerning certain aspects of CV writing or his contacts with a potential employer, in form of the Appendix.

A look at the book's structure

The book, in its final form consists of four parts:

Part I: an A-Z bilingual dictionary, with entries organised in the following manner:

1. headword;
2. part-of-speech label;
3. synonyms;
4. list of collocates with the Polish translation accompanying each phrase.

The following example from the dictionary may serve to illustrate this:

acquire verb syn. get, obtain, achieve, secure, amass

zdobyć, nabyć, zyskać

~ experience – *zdobyć doświadczenie*

~ qualifications – *zdobyć kwalifikacje*

~ skills – *zdobyć umiejętności*

~ material – *zdobyć materiał*

~ property – *nabyć/zakupić nieruchomość*

~ a reputation – *zyskać reputację (np. o producencie)*

~ value – *nabrać/zyskać na wartości*

Part II: an A-Z English index. The way of presenting single lexical items (many words may appear as a collocate only and not be re-listed as independent entries) called for a dependable cross-reference index. Every important term in every phrase is cross-referenced, except for prepositions, pronouns and other less useful items. Verbs and derived nouns like *open* and *opening*, or verbal adjectives like *finished* constitute separate entries but homonyms like *open*, verb and *open*, adjective do not. This system allows for maximum clarity and ease of access; for the reader who wants to make sure about for example, the

grammatical category the word in question represents, this information is included in Part I.

Part III: A totally different strategy was suggested by the publisher for The Polish index. It contains all the significant Polish translation equivalents of the English nouns, grouped in niches. To save space, niche lemmata are represented by a swung dash (~), cf. the following entry from the Polish index:

prawo: przestrzegać abide; obey; observe starać się o ~ transmisji bid surowe; twarde; restrykcyjne; popytu i podaży law ustanowić pass istniejące rule mądre; uniwersalne; leżące u podstaw principle szanować respect przestrzegać poszanowania; wprowadzić w życie enforce ~ wyłączności claim dać ~ delegate

Part IV: Appendix. Part IV includes a lot of culture-dependent information both new to the Polish users and yet crucial to their success in the original goal they have set for themselves: functioning in the English-speaking working environment which on one hand entails performing flawlessly in English in a workplace context and on the other, understanding the reality underlying that context. Thus, the inclusion of the material, the preponderance of which shows a strong tendency to attract a range of collocates, comes from the firm belief on the part of the author that studying a foreign language should never be limited to the study of lexical items and the rules of grammar alone and a good learner-oriented dictionary must occasionally include information that we traditionally associate with other types of publications, for example course books of a different kind.

Final remarks

Compiling a dictionary is a mammoth task; long-term; always a team effort, with the participants having access not only to large computerised corpora but assisted by sophisticated computer programs helping them select and organise the material in question. None of these assets were the author's: it was a one-person project with the author determined to have the complete product ready in a little more than a year and a computer was only used to type the material. It was, however, clear from the very beginning that since, as Sinclair (1987:143) points out, *the normal use of language is to select more than one word at a time and to blend such selections with each other*, it was to be a dictionary of collocations. It is only a matter of time before Poland becomes a member of the European Union, with a huge number of Poles – administrative workers at different levels, companies, community representatives, professional people from various fields, government officials to name just a few –

communicating in English on an everyday basis being a norm rather than the exception. The only way to prepare them for the task must be to start with the re-evaluation of existing curricula and, if need be, devising a range of teaching materials aimed at this specific audience – materials organised in such a way that (especially with the opportunities of today's technology) not only a chosen word form but foremost the context that is currently of interest for the language user is just a click of the mouse away.

It is hoped that this book is a small step towards this goal.

References

- Allerton, D.J.** 1984. "Three (or four) levels of word co-occurrence restrictions". *Lingua*. 63. 17-40.
- Cowie, A.P.** 1981b. "The treatment of collocations and idioms in learners' dictionaries". *Applied Linguistics*. 2. 223-35.
- Mackin, R.** 1978. "On collocations: Words shall be known by the company they keep". In *Honour of A.S. Hornby*, ed. by P. Strevens. Oxford-New York: Oxford University Press. 149-65.
- Mitchell, T.F.** 1971. "Linguistic *goings-on*: Collocations and other lexical matters arising on the syntagmatic record". *Archivum Linguisticum*. 2. 35-69.
- Sinclair, J. and A. Renouf.** 1991. "Collocational frameworks in English". *English Corpora Linguistics: Studies in Honour of Ian Svartvik*, **K. Aijmer, B. Altenberg (eds)**. London & New York: Longman. 128-142.

Lexicographic publications

- (The) BBI Combinatory Dictionary of English*. 1986. **Benson, M., E. Benson & R. Ison**. Amsterdam – Philadelphia: John Benjamins.
- Collins Cobuild English Language Dictionary*. 1995. **J.M. Sinclair (ed.)**. London & Glasgow: Collins.
- English at Work. An English-Polish Dictionary of Selected Collocations*. 2001. **Osuchowska, D.** Warszawa: Wydawnictwo Naukowe PWN.
- Oxford Advanced Learner's Dictionary of Current English*. 1980. **A.S. Hornby (ed.)**. Oxford: Oxford University Press.
- Routledge Dictionary of Language and Linguistics*. 1996. **H. Bussman (ed.)**. Translated by K. Kazzazi and G.P. Trauth. London & New York: Routledge.
- Selected English Collocations*. 1982. **H. Dzierżanowska and C. Douglas Kozłowska**. Warszawa: Państwowe Wydawnictwo Naukowe.
- Word For Word. A Dictionary of Synonyms*. 1988. **J.O.E. Clark**. London: Clark Robinson Limited.