# Anna Dziemianko. *User-friendliness of verb syntax in pedagogical dictionaries of English*

## 1. Introduction

During the planning stages of the second edition of the *Longman Dictionary of Contemporary English* (LDOCE2, 1987) – the first dictionary I edited – one of the main questions under discussion was what to do about syntax. About 15 years later, when the *Macmillan English Dictionary* (MED, 2002) was at a similar stage, syntax had almost ceased to be an issue. By the late 1990s, we were able to conclude that approaches to describing syntactic behaviour in the various monolingual learners' dictionaries of English (MLDs) had reached a natural end-point: they had coalesced around a limited range of fairly simple options, and we took the view that there was not a great deal more to be done in this area. Having read Anna Dziemianko's excellent book, I am not so sure.

*User-friendliness of verb syntax in pedagogical dictionaries of English* reports on a large-scale, rigorously-designed experiment which the author conducted in order to assess the usefulness and usability of the various systems used in MLDs for describing the syntactic behaviour of verbs. This forms the heart of the book, but Dziemianko kicks off with a well-researched survey of the field. She follows the trajectory of syntax-coding systems, from the 'verb patterns' introduced in Palmer's *Grammar of English Words* (1938) to the (supposedly) transparent approaches of the present day, and she reviews relevant user-research along the way.

For a long time, the choice was between two equally arcane (and mutually incompatible) coding systems, as found in LDOCE1 (1978) and OALD3 (1974). The descriptive power of these systems was never in doubt: they enabled lexicographers to provide a delicate and fine-grained account of most syntactic patterns. For this reason, they were popular in the NLP community – I was almost lynched at a computational linguistics conference in the US when word got out that I was 'the man who removed the codes from LDOCE'. But by the early 1980s, it was becoming clear that the average dictionary user got very little benefit from these codes. This must be what Tony Cowie had in mind when he referred – in his Introduction to a special issue of *Applied Linguistics* on pedagogical dictionaries – to "the gap that is known to exist between the sophistication of some features of dictionary design and the user's often rudimentary reference skills" (Cowie

1981. 206). In his classic user-study in the same volume, Béjoint reports that "their [the dictionaries'] introductions are not commonly referred to, and neither are the coding systems for syntactic patterns" (Béjoint 1981. 219). Extensive market research at Longman similarly revealed that "although grammatical information is sometimes sought, most users found mnemonic codes offputting and impenetrable" (Summers 1987. F8). For the heavily-coded systems, the game was clearly up, and we then entered a period in which the two (then three, then four, then five) competing MLDs experimented with alternatives to OALD's 51 'verb patterns' and LDOCE's almost infinite alphanumeric combinations (like I5, L8, and X7). The trend was towards simplification and – to a degree – standardization, and a contemporary student who switches from one dictionary to another no longer has to relearn an elaborate inventory of symbols and codes.

But there have been two other big changes since the 1980s, and both have implications for descriptions of verb syntax. First, changes in defining styles. On the one hand, 'full-sentence definitions' (FSDs) were introduced in COBUILD1 (1987), and have since been taken up (in varying degrees) by the other MLDs (Rundell 2006). As Dziemianko shows, "the left-hand part of a full-sentence definition is a reflection of the characteristic syntactic patterns in which the verbs occur" (37). Thus the definition of *hope* ("If you hope that something is true, or if you hope for something…") tells the reader – without the need for codes – that the verb can be used in a *that*-clause or in a PP with *for*. On the other hand, the move away from 'lexicographese' meant that even 'traditional' definitions now dispensed with the brackets used (*inter alia*) for showing typical objects. This entails some loss of precision with regard to syntax. When *assassinate* is defined (without brackets) as:

to murder an important or famous person, especially for political reasons [OALD7]

it is no longer clear from the definition wording alone whether the verb is transitive or not.

The second major change has, of course, been the arrival of corpora. With large amounts of language data at their disposal, lexicographers have been able to focus more systematically on what
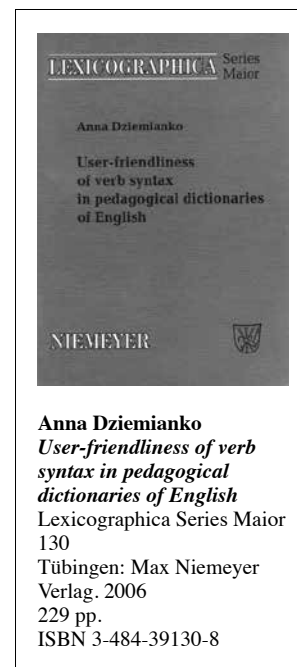
Patrick Hanks calls "the probable not the possible" (Hanks 2001) – and this has implications for syntax as well as for meaning and phraseology. LDOCE1 and OALD3 aimed to give a complete account of the *possible* (as opposed to regularly-occurring) syntactic behaviour of verbs, and their coding systems provided the tools for doing this. Thus at the second meaning of *suppose* ('to believe'), LDOCE1 has no fewer than six codes, including [X1] (=verb+object+adjective complement: *they supposed him dead*) and [X9] (=verb+object+adverbial: *they supposed him somewhere else*). Most users of English (native or otherwise) could get by pretty well without knowing about either of these patterns. Yet it was common in both dictionaries for a verb entry to start by reeling off a list of codes, with only a subset of these actually illustrated by examples – for the very good reason that the non-illustrated patterns were (like these for *suppose*) almost never used in normal discourse. So the trend away from opaque coding entails not only simplification, but some loss of information too – albeit a loss that most of us would not mourn.

A key theme, then, as Dziemianko observes, is this tension between complete and accurate description on the one hand, and user-friendliness on the other: "the ease of accessibility is difficult to reconcile with the accuracy of description" (5). (An interesting question is whether or not this amounts to a fundamental incompatibility.) She mentions the familiar case of verbs whose surface pattern is verb+noun/pronoun+*to*-infinitive, and notes the technical distinction between *We want you to leave* (where 'you' is a direct object) and *We advise you to leave* (where it is an indirect object): they look identical, but the underlying differences emerge when you try a passive transformation. Older coding systems could (and did) account for this distinction, but contemporary MLDs tend to stick to surface grammar. This is an issue that no doubt has resonance in the more bracing academic climate of Dziemianko's native Poland, but I suspect it would mean very little to the average UK-educated teacher of EFL. At any rate, the author—rightly, I think—concludes that this development "should be assessed positively" (16). Dziemianko's admirably thorough opening chapter takes us through all these developments and sets the scene for her research project.

## 2. Design of experiment

In Chapter 2, Dziemianko describes the design of her experiment and the thinking behind it. In brief, she identifies a number of variables that affect the usability of the syntactic information supplied in MLDs. These are:

ᵢ definition style: the choice here is between what she calls 'analytical' and 'contextual' definitions (or, if you prefer, conventional definitions and FSDs);

ᵢ type of explicit syntactic information: 'formal' codes (such as Vn), 'functional' codes (like T+obj+*to*-inf), and 'pattern illustrations' or 'PIs' (like *want sb to do sth*);

ᵢ location of codes: these can appear either in the entry's example text (where a code or PI precedes an example that instantiates it) or outside the entry in an 'Extra Column'.

Dziemianko creates 10 different mini-dictionaries, each of which contains entries for the same 15 verbs, with every entry in a given dictionary exhibiting the same combination of the variables described above. This minimizes variation among the 10 different versions, to ensure that the effects of each variable can be individually assessed (70). The 15 verbs used in the study are all of low frequency (and therefore unlikely to be familiar to the testees), and cover a range of syntactic behaviours from the simple (like *haemorrhage*) to the complex (like *jolt, yank*, and *subpoena*). The dictionary entries are designed to look as 'real' as possible, and they assemble material from a range of MLDs in various permutations, including definitions, example sentences, IPA pronunciations, part-of-speech labels, and of course the various forms of syntactic code. Following a cleverly-designed pre-test, subjects complete a multiple-choice test relating to each of the 15 verb entries in their mini-dictionary. Additionally, they are asked to underline any part (or parts) of the entry in which they located the information they needed to perform the test. Two large groups of subjects took the test: about 300 high school students and a similar number of students from Dziemianko's own university in Poznan. This adds the further dimension of language proficiency, so any differences in dictionary-use strategies between these two cohorts can also be observed.

This is at best a cursory overview of a meticulously planned piece of research, which (to my knowledge, anyway) is on a larger scale, and covers a wider range of variables, than anything attempted so far in this area. What is so impressive here is Dziemianko's terrier-like determination to identify any non-relevant factors that might vitiate her results, and then make appropriate adjustments to minimize the

risk. I'm not qualified to comment on the soundness of her statistical methods (described in some detail on pp72-82), but by the time I got to this point I had seen enough to take this section on trust.

## 3. Findings and implications

The immense care taken over the design of the experiment pays off handsomely in the breadth and depth of the data it delivers. A short review can't do justice to the 50-odd pages of analysis in Chapter 3, in which numerous hypotheses are tested against the experiment's results, so a few highlights will have to do. In no particular order:

ᴸ subjects with higher language proficiency were much more likely to get their syntactic information from *multiple* sources of information within the entry, whereas the high-school students tended to focus on just one or two entry components;

ᴸ examples were the favourite source of syntactic information in most cases, particularly among the high-school students;

ᴸ definitions were in general the least favoured source of syntactic information, but *contextual* definitions (or FSDs) were resorted to more often than analytical ones;

ᴸ the positioning of codes (whether in a side column or in the body of the entry) did not seem to make much difference to the frequency with which they were consulted;

ᴸ where codes were used, functional codes – perhaps surprisingly – were preferred to formal ones. For the university students especially, coded syntactic information was still quite frequently used (and successfully, on the whole);

ᴸ (but) PIs were generally preferred to codes of either type. They were consulted "much more frequently …than any codes in entries with analytical definitions, and even than codes and contextual definitions taken together in the others" (154). Where PIs appear in the entry, the resort to examples is sharply reduced (152). And (somewhat counterintuitively) PIs were used more often by university students than by the less proficient high-school students.

Where does this leave us? Dziemianko concludes (188) that "as far as syntactic information is concerned, a user-friendly verb entry should contain examples, a contextual definition [FSD] and functional codes interspersed among examples". But she concedes that the jury is out on "those conclusions which pertain to codes and pattern illustrations". In most respects, this looks like sensible advice. As far as the use of contextual definitions goes, my own view (Rundell 2006) is that these work best when the syntax is

straightforward and there is a dominant syntactic preference – thus verbs typically used reflexively, intransitively, or with a simple PP tend to fit this model well. But the format is less successful with verbs whose syntactic behaviour shows a range of equally valid possibilities. In cases like this, you either have to commit to just one of several structures (thus apparently downgrading other possibilities), or to create a cumbersome definition that attempts to account for them all.

## 4. Some concluding remarks

Most writers who have carried out research in this area have ended with a plea for more teaching of dictionary skills, and Dziemianko is no exception (190-191). This is understandable enough – it is obviously frustrating if users are unaware of, or unable to use, all the riches their dictionaries provide. Desirable though this may be, I suspect it is not the answer. For the generation now using MLDs (typically, people in the age range 16-24), complete transparency is the default expectation. The iPod comes with almost no instructions – you just have to figure it out, and most people under 30 have no problem with this. So it is incumbent on designers of dictionaries to create systems that users don't have to *learn* and that don't require elaborate explanatory material.

On the other hand electronic media open up new opportunities. Users could choose from several levels and several types of syntactic information to suit their individual needs, skills, and preferences – from the minimal to the complex, from pattern illustrations to descriptively powerful codes. We also need to think about the many areas of grammar which *none* of the current systems deals with adequately. MLDs are still relatively superficial when it comes to explaining issues such as whether a complement or pattern is optional or obligatory; in what circumstances the object of a transitive verb can safely be omitted; whether an obligatory adverbial (for verbs like *put*) has an endless range of exponents; and so on. To give a single example: you can *prevent someone leaving* or *prevent someone from leaving*: the *from* appears to be optional – but it isn't optional when the verb is passivized. This is hardly an obscure fact of grammar, but you won't find it in any of the current MLDs. Colligation, too – the preferences some verbs have for appearing in the passive or in a progressive form or infinitive, for example – is at best covered patchily. The description of syntactic behaviour is far from complete, and better ways of presenting that description can still

be discovered.

Dziemianko's research (even if this was not the primary intention) makes a strong case for dictionary designers to revisit the area of syntactic description, and provides a great deal of valuable data to inform this debate. The book isn't always an easy read, and Dziemianko occasionally gets bogged down in debates that aren't strictly relevant: for example, there is a lengthy discussion (22-28) on the relative merits of 'made-up' versus 'real' examples – which doesn't add much to Dziemianko's argument, and is a rather overblown topic anyway. One might question, too, how far her subjects are typical of the whole community of MLD users. Her university cohort had an average of ten years' English instruction, and had attended courses in linguistics and English grammar – which must put them at the higher end of the skills spectrum. One other minor complaint: it was a little surprising to find no index, though perhaps that's more of a problem for a reviewer than for a 'normal' reader. But these are very small blemishes. This is an exemplary study and a valuable contribution to the body of user-research.

**References**

Béjoint H. 1981. The foreign student's use of monolingual English dictionaries, *Applied Linguistics* II.3.207-222.

Cowie A.P. 1981. Introduction to special issue on pedagogical dictionaries, *Applied Linguistics* II.3.

Hanks P. 2001. The probable and the possible: lexicography in the age of the Internet, in Sangsup Lee (ed.), *Asialex 2001 Proceedings*. Seoul: Yonsei University. 1-15.

Rundell M. 2006. More than one way to skin a cat: why full-sentence definitions have not been universally adopted, in Elisa Corino, Carla Marello, Cristina Onesti (eds.), *Proceedings of the XII Euralex International Congress, 2006*. Torino: Edizioni dell'Orso. 323-338.

Summers D. 1987. Introduction to *LDOCE2*. Harlow: Longman.

**Michael Rundell**
**Lexicography MasterClass**
**michael.rundell@lexmasterclass.com**

# Włodzimierz Sobkowiak. *Phonetics of EFL Dictionary Definitions*
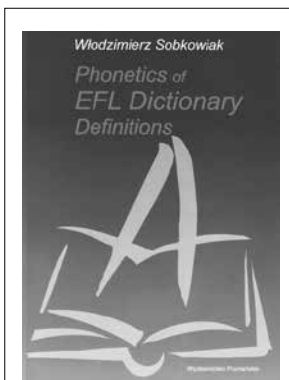
**Włodzimierz Sobkowiak**
*Phonetics of EFL Dictionary Definitions*
Anglistyka
Poznań: Wydawnictwo Poznańskie. 2006
249 pp.
ISBN 83-7177-450-8

The rationale of *Phonetics of EFL Dictionary Definitions* is to provide lexicographers with phonetically-based insights into their choice of words in dictionary definitions so that these definitions can be more easily understood by second language learners.

The book concentrates on a rather neglected area of lexicography, namely, the application of phonetic principles to dictionary writing. Why is this type of research relatively rare? The answer is found when one considers the widespread knowledge needed in such disparate and, to many, largely inaccessible areas in order to tackle this issue. The areas of expertise include: phonetics, computational linguistics, statistics, corpus linguistics, contrastive phonology, natural language processing (NLP), etc. It is unusual to find one person who can enter such a large arena of events, and be capable of handling such immense diversity. Sobkowiak is an exception. His knowledge of all of these spheres is impressive, and his ability to integrate these outlying strands into one woven piece of lexicographic cloth is indeed admirable. In fact, looking at Sobkowiak's work over the last years indicates a nearly one-man crusade for the inclusion of phonetic analyses in lexicographic research (Sobkowiak 2002, 2003, 2004).

The same reasons that make this work truly notable, namely its breadth and attention to detailed analyses, also provide the major obstacles to its wider acceptance and fuller understanding. To comprehend the book, you must be familiar with concepts as diverse as statistical frequency analyses and N-grams from the field of corpus linguistics to more esoteric and specialized notions from the field of phonetics. Concepts such as phonological interference, sandhi phenomena, as well as the various phonetic terms that are used in the book (e.g. devoicing, syllabic sonorants, palatalization, overnasalization, just to name a few) may be rather obscure to someone from a purely lexicographic background. Meanwhile, the detailed tables and charts sometimes bog down the reader with so many intricacies that one is often trying to look for the forest while navigating the many trees.

One suggestion that could be helpful