COLLOCATIONS AND DISTINCTION OF SENSES IN PRINTED MONOLINGUAL LEARNERS' DICTIONARIES: THEORY AND PRACTICE

JEĻENA DOROŠENKO

University of Latvia, Latvia

Abstract. The paper discusses meaning presentation strategies in two general monolingual English learners' explanatory dictionaries against the background of the concept of meaning advocated in metalexicography. The analysis of theoretical views has revealed that the postulated primacy of context, and, more narrowly, of collocations, in defining meanings, backed by arguments from both linguistics and beyond, had an impact on processing word meanings, in terms of the resulting structure of dictionary entries. In the analysis of entries the texts of definitions have been juxtaposed to collocations given in the same entries in two editions of *Longman* and *COBUILD* dictionaries. Later editions not only provide considerably more collocations in both examples and definitions, but also more often split senses into subsenses, the decisive factor for splitting being different collocates of the head word, to the extent that definitions of some senses or subsenses are either identical or very similar. Meaning is seen as a fuzzy category with blurred and overlapping edges both in theory and in lexicographic practice.

Key words: meaning, context, collocations, monolingual learners' dictionaries, entry structure, number of senses, fuzzy categories

INTRODUCTION

The use of corpora made linguists at large, and lexicographers in particular, more alert to language in use, as opposed to the description of language units as isolated entities. One of the aspects of this new focus has been the search for words that, firstly, co-occur most frequently, and, secondly, for those patterns of co-occurrence which are not only regular, but to a large extent predictable (as in *commit + murder, crime,* plus a limited number of other noun collocates). The notion of collocation in both the first, i.e. the broad, and the second, i.e. the narrow, meaning of the term was increasingly in the centre of attention since mid-1980s at least. The 1990s were seen as the 'groundbreaking decade' in terms of 'attention to prefabricated chunks of language' (Fontenelle, 2002: 219). The impact of this shift was felt both in theoretical and applied linguistics, including language learning and lexicography. For lexicography this is a relatively recent development. Even though Saussure's view that syntax, or syntagmatic relations, are not part of the language system, but a feature of speech, had been

discarded long ago, it seems to have been implicitly followed by lexicographers for a long time (in this respect their reputation for disregard of developments in linguistic theory was justified). Thus, the recent decades of lexicographic theory focussing on collocations as patterns either mapped into meanings, or of meanings mapped into them, signified a turn of tide. The most obvious effect was the emergence of several dictionaries of collocations. First came the *BBI Combinatory Dictionary* and the *English Dictionary of English Word Combinations* (1986). This title was replaced by the *BBI Dictionary of English Word Combinations* in the revised edition in 1997, and in 2010 the expanded and updated version *BBI Combinatory Dictionary of English* came off print; its database used both the British National Corpus and Internet searches. The *Oxford Collocations Dictionary for students of English* was published in 2002, the second edition came off print in 2009. Both brands now have online downloadable versions. *Longman* and *Macmillan* later produced their own dictionaries of collocations. But perhaps less obviously, attention to collocations has also changed the approach to presenting and describing word senses in general explanatory monolingual English-English dictionaries, including learners' dictionaries. It is this change and, in particular, the very nature of theoretical arguments voiced soon after the 1990s in favour of the approach which made 'context' (or collocations) the decisive factor when grouping senses in explanatory dictionary entries, that will be discussed in the paper.

GOAL AND METHOD

The goal of the paper is twofold. Firstly, the shifting concept of word meaning as revealed in some discussions in theoretical, or meta-lexicography at the turn of the century, is analysed and summarised. The focus is on the range and nature of arguments put forward to substantiate the new theory of meaning, especially since the arguments come from both linguistics and other domains of knowledge. The second goal is to show the impact of the new concept of word meaning on practical lexicography in terms of how word senses and subsenses are distinguished in monolingual dictionary entries and on what grounds. For this purpose the diachronic comparison of the structure of the entries of two common polysemantic words has been made: *taut* in two editions of the *Longman Dictionary* [*LDOCE1* and 5] (1978 and 2009, i.e. some time before the extensive use of corpora in dictionary making and some time after the 'groundbreaking decade') and *calm* in two editions of *COBUILD* (1987 and 1995, to see if any dynamics can also be observed practically within the decade). This involves relating definitions of word meanings in the entries and collocations used as examples in the same entries to the number of senses in the entry. The small size of the sample makes it a purely qualitative study with a merely illustrative role. However, it is backed by relevant quantitative data based on a broader study of *LDOCE1* and 5 concerning the overall increase in the presence of collocations in entries, carried out in 2012 under my supervision (MA paper by

Kokareviča, 2012). In other words, the policies of splitting senses are revealed on the background of the analysis of theoretical views voiced by major lexicographers in the collection of papers *Lexicography and Natural Language Processing* (2002) which systematized the lexicographic experience of the 1990s and outlined its theoretical implications.

THEORETICAL BACKGROUND

Until recently most research focussed on the role of collocations as examples in showing more patterns of word use than definitions can possibly do (see, e.g., Cowie's historical review of collocations used as 'skeleton examples' in English and French dictionaries (Cowie, 2002: 79–80). Referred to by this author also as 'lexical collocations' or 'minimal lexicalised patterns' (ibid.: 78), they consist of two or more open-class words in a specific syntactic pattern.

In the 1990s and early 2000s research was also devoted to the selection of collocations in monolingual dictionaries, and of the headwords under which they appear (the base/node or the collocate, or both). The issue of selection always involves the discussion of the concept of collocation as a type of conventionalised word combination/semi-preconstructed phrases distinct (not always clearly) from idioms/frozen/bound collocations (the terms vary for both categories) on the one hand, and free word combinations, on the other, as in (Heid, 1994: 226–257; Bentivogli and Pianta, 2000: 663–670). Another problem discussed at the turn of the centuries is the user's access to collocations, or the ways of presenting collocations in the microstructure of entries, as in (Heid, 1994: 240–241; Van Der Meer, 1998: 313–322; Heid, 2004: 729–739).

The interest in these issues was triggered by the steadily increasing presence of collocations in explanatory monolingual learners' dictionaries. A close-up on any major mainstream monolingual explanatory dictionary provides evidence of the increasing number of collocations in the overall structure of entries. E.g., the comparison of two editions of the *Longman Dictionary of Contemporary English*: 1978 and 2009, analysing 50 random entries for nouns and adjectives, revealed a significant increase in the number collocations used both as part of definitions for word senses and subsenses, and in examples for these entries (Kokareviča, 2012). All in all, in the analysed sample *LDOCE5* (2009) contains 267 collocations in both noun and adjective entries, while *LDOCE1* (1978) contains twice less – 131 collocations (Kokareviča, 2012: 44). The number of entries containing no collocations at all decreased in *LDOCE5* twice as compared to the same sample in *LDOCE1* (6 vs. 13). Similarly, in the same sample the number of collocations in examples almost doubled in *LDOCE5* – from 123 to 232 (ibid.: 6). While in *LDOCE1* 'collocations can be found as senses or subsenses in bold 8 times in noun entries and zero times in adjective entries, in *LDOCE5* collocations can be found as senses or subsenses in bold 18 times in noun entries and 16 times in adjective entries (ibid.: 47). The example

below shows collocations of **tangent** used both in the definition (in bold) and as an example in *LDOCE5*, in contrast to *LDOCE1* where for the same sense of the word no example is given (ibid.: 48):

> *LDOCE1*
> **tangent 3 <u>go/fly off at a tangent</u>** *infml* to change suddenly from one course of action, thought, etc., to another.
> *LDOCE5*
> **tangent** [C] **1 <u>go off at a tangent</u>** *BrE*, **<u>go off on a tangent</u>** *AmE*
> *informal* to suddenly start thinking or talking about a subject that is only slightly related, or not related at all, to the original subject: *Let's stay with the topic and not <u>go off at a tangent</u>.*

The account above makes it obvious that in absolute numbers the bulk of the increase of the presence of collocations in explanatory dictionaries is due to collocations used as examples. However, increase in the sheer number of collocations, in the number of entries using collocations as part of definitions, as well as increase in the number of collocations per entry used as examples, is not as yet evidence of structural changes in dictionary entries; namely, of changes in the amount of senses and sub-senses seen by dictionary makers as worth distinguishing. The question is, are collocations increasingly sense discriminators in explanatory dictionary entries, or not: do they encourage splitting of senses and subsenses in dictionary entries? In other words, how are collocations and word meanings related in English explanatory dictionaries? At the turn of the centuries collocations were revisited in metalexicographic research largely with these questions in mind.

RESULTS AND DISCUSSION

The collection of papers *Lexicography and Natural Language Processing* published by Euralex in 2002 can be viewed as one of the first consolidated reflections on the 1990s or the 'groundbreaking decade' in terms of attention of dictionary makers to collocations. Most of the authors of its contributions were involved in the 1990s in major lexicographic projects, which makes the book a broadly summarizing statement on both their actual dictionary making experience and on its theoretical implications projected, in their turn, on the future. Even though only one of the papers used the term 'collocation' in the title, and one – 'word groups', almost all papers had something to say on collocations. The most striking aspect of the book is the general overt scepticism voiced by major mainstream lexicographers about word meanings, or senses. The very existence of word senses (as existing prior to contexts) is problematised, challenged or even denied:

'Word meaning (if such a thing exists at all)…' (Rundell, 2002: 147). Discussing word sense disambiguations, the same author notes 'increasing doubts among lexicographers … as to whether there is anything to disambiguate' (ibid.). Atkins' statement 'I don't believe in word senses', voiced by her much earlier in

a discussion at 'The Future of the Dictionary' workshop in Uriage-les-Bains in 1994, which became part of lexicographic folklore, is referred to in the collection. It was also used by Kilgariff (one of the contributors to the collection) as the title of his later paper claiming that '…word senses are abstractions from clusters of corpus citations, in accordance with current lexicographic practice. The corpus citations, not the word senses, are the basic objects in the ontology' (Kilgariff, 2008: 20). Thus, '[…] "word sense" or "lexical unit" is not a basic unit. Rather, the basic units are occurrences of the word in context (operationalised as corpus citations' (ibid.). He suggests 'an alternative conception of the word sense, in which it corresponds to a cluster of citations for a word' (ibid.).

While not everybody subscribes to this radical stand and some believe that word senses do exist, they are seen as 'extremely vague and unstable' (Hanks, 2002:159), 'opaque', and 'the perceived meaning is private' (Hanks, 2002: 181). Again in the words of Kilgariff, 'most dictionaries encode a variety of relations in the grey area between "same sense" and "different sense" (Kilgariff, 2008: 11).

The arguments backing this scepticism are drawn from both within linguistics and beyond. The simplest and oldest linguistic one is that sense distinctions made by lexicographers are subjective. This idea is restated by Hanks a number of times, e.g.: 'A word may have about as many senses as a lexicographer cares to perceive … it is often impossible to map the semantic distinctions made by one dictionary onto those of another' (Hanks, 2002: 159). Context, on the contrary, is seen by him as objectively observable and measurable: '…syntagmatics requires working with palpable material, which can be measured and objectively evaluated' (ibid.: 181).

However, 'vagueness' or 'subjectivity' do not lead anywhere in terms of practical guidelines on how to tackle word senses in dictionary making. Thus, a more specific and potentially a more operational term has been suggested**.** Rundell discusses word meaning as 'the fuzziest category of all', fuzzy categories being those with blurred edges, or graded boundaries. The concept had been first introduced in 1965 in mathematics in the terms 'fuzzy sets', used by Lotfi A. Zadeh for sets whose elements have degrees of membership instead of either belonging or not belonging to a set, and 'fuzzy concepts' or 'fuzzy categories'. While fuzzy sets had been first studied in mathematics and logic, Rundell points out that fuzziness (also referred to as 'gradience') was recognized also in language studies as an 'endemic' feature for word-classes and for types of texts discussed in terms of modes of discourse (Rundell, 2002: 147). For example, some nouns are more central (or prototypical) than others (only countable nouns normally have grammatical number and case, while uncountables mostly have none); or online messages exchanged in real time admittedly have the features of both written and spoken communication modes. Rundell concludes: 'word meaning can be regarded as (at best) yet another form of prototype' (ibid.: 147). It should be noted that the idea of subjectivity of word meaning is implied by the very notion of fuzziness which in cognitive semantics presupposes that categories are learnt and rooted in experience, so meaning, also mostly learnt from experience, is not

an objective, but a subjective construct: '…"furniture" is a fuzzy category in that while "table" and "chair" are clearly members, some people judge artefacts such as "picture" and "carpet" as belonging to this category while for others such objects are better thought of as belonging to a related category such as "furnishings"' (Evans, 2007: 88).

The concept of fuzziness of meaning, according to Rundell, gets support not only from logic and mathematics, but also from natural sciences: parallels are drawn with the classification of species in the natural world. Thus, Rundell quotes the geneticist Jones' book *Almost Like a Whale: The Origin of Species*: 'species can […] no longer be seen as absolutes […] they are not fixed. Instead, their boundaries blend before our eyes […] differences blend into one another in an insensible series' (Jones, 1999, cited by Rundell, 2002: 147). Indeed, if birds are defined as having feathers, beaks and the ability to fly, then a nightingale will be a more 'prototypical' bird than, for example, a penguin which does not fly. These observations are seen by Rundell as a close parallel to features of meaning: '…it would be difficult to find a better description of how word meaning works' (ibid.: 147).

It should be noted that appeals of linguistics to arguments from natural sciences were quite common in the second half of the 19th century within the so-called biological paradigm in comparative-historical linguistics based on Darwin's evolution theory and on the then popular view that the study of language is, or should be, a natural science. It followed that its method was on the whole (or should be) the same as that of the other natural sciences. The reference to Jones with its title echoing Darwin's book is a telling one: for some time after the publication of Darwin's *The Origin of Species* in 1959 linguistics saw itself as a subfield of natural sciences.

The approach 'continued up to the end of the 19th century when it had been seriously called in question by Saussure (1857–1913), whose posthumously published *Cours de linguistique generale* (1916) launched 20th-century structuralism on its course' (Harris and Taylor, 2005: xix). Structuralism had various denominations, but they all shared the idea that language studies should not use analogies with other domains of knowledge or rely on arguments from them. It seems that the 21st century linguistics, following about a century of emancipation from the postulates of natural sciences, feels free again to revisit the biological paradigm, not putting the already firmly established autonomy of language studies at risk.

Since many explanatory monolingual dictionaries brand themselves as targeted at learners, the attack on meaning as a self-contained entity is also linked specifically to the perceptual perspective, relevant in language acquisition and language learning studies: 'Human beings have a natural tendency to define the context […] rather than focusing on particular contribution of the word to the contexts in which it occurs' (Hanks, 2002: 159). In a similar vein, Fontenelle claims that '[r]esearch in applied linguistics and language learning has shown that words are best learnt and retained if they are presented […] in context and

more specifically together with the other items with which they are most likely to appear' (Fontenelle, 2002: 219–220). To reformulate, the justification for the increased presence of context in entries and its role in staking out senses is dictionary users-learners' perceptual ease.

The arguments listed above resulted in methodological considerations on the desired sequence of analytical procedures when distinguishing senses in a dictionary entry. The idea is that corpus evidence for each word should be grouped by contexts, and only then sense distinctions can be deduced: '…the lexicographer must first group the corpus evidence for each word according to the contexts in which it occurs, and then decide to what extent it is possible to group different contexts together (on the grounds that they express essentially the same meaning), and to what extent it is necessary to make distinctions' (Hanks, 2002: 159). Summarising, Hanks writes: 'Lexicographers should think first in terms of syntax and context […] rather than directly in terms of semantics. They can thus approach meaning indirectly, through syntagmatic analysis' (Hanks, 2002: 159–160). Indeed, if context is observable and measurable, while meaning is not, the sequence of analysis becomes crucial: it is logical to proceed from the known to the unknown: from context to meaning, and not vice versa, for: 'If perceived meaning is the organizing principle of a dictionary, examples can be found to illustrate the perceptions of the writer, but that does not mean that the writer has achieved the appropriate level of generalization' (Hanks, 2002: 181). The resulting guidelines for dictionary makers are as follows: '… syntagmatics, rather than (or, rather, in tandem with) perceived meaning, should be the organizing principle of the dictionary entry' (Hanks, 2002: 181).

Is the acknowledgement of fuzziness of meaning and the priority of context (therefore, of collocations) which follows from it, reflected in any way in dictionary entries' structures? If so, what are these structures like? Fontenelle has predicted the emergence of 'a new generation of dictionaries which take the collocational dimension as a central axis.' (Fontenelle, 2002: 220).

The trend has an obvious impact on the handling of senses in explanatory monolingual dictionaries. Two headwords have been chosen to show it: **taut** and **calm** as common-core English words listed in all dictionaries, polysemantic, but with a relatively simple semantic structure. Entries of **taut** have been compared in *LDOCE1* and *LDOCE5* (1978 and 2009). Collocates are given below in bold italics.

> *LDOCE1*
> **taut1** tightly drawn; stretched tight: *Pull the **string** taut!| taut muscles*
> **2** showing signs of worry or anxiety: *a taut **expression** on her **face***
> *LDOCE5*
> **taut1** stretched tight *The **rope** was stretched taut.* **2** showing signs of worry, anger etc. and not relaxed | *a taut **smile** | Catherine looked upset, her **face** taut.* **3** having firm **muscles**: *her taut brown **body*** **4** a taut ***book, film, or play*** is exciting and does not have any unnecessary parts: *a taut thriller*

The *LDOCE5* entry has more examples (full sentences or collocations: 5 versus 3 in *LDOCE1*), but taking into account the definitions in *LDOCE5*, the number of collocations in its entry is 9. *LDOCE1* does not use collocations in definitions at all. More importantly, in *LDOCE5* the amount of senses of **taut** has doubled. Sense 1 of **taut** in *LDOCE1* is split into senses 1 and 3 in *LDOCE5* where sense 1 relates to an inanimate object – *rope*, while sense 3 – to the physical state of either *muscles* or, generally, of human *body* (undifferentiated in *LDOCE1*). Sense 2 has *face* and *smile* as collocates, sense 4 – *book, film, play, thriller* for texts or films. Collocate nouns in *LDOCE5* are decisive: they split one of the old senses and add one new sense. Each sense is matched with a set of collocations, the definitions of senses do not overlap in wording, even though the similarity of 'not relaxed' and 'firm' in senses 2 and 3 is evident. To generalise, the impact is the tendency to split senses more, the justification for splitting being different nouns collocating with the defined word.

An even more radical impact of the approach emerges in the *Collins Cobuild English Dictionary* (1995). The entry **calm** (adjective, verb and noun) has, all in all, 11 senses, only 4 of which are given below to illustrate the approach. Examples have been omitted here, since definitions always incorporate collocations (collocates are in bold italics).

> **5** If someone or something **calms** a ***situation***, they reduce the amount of trouble…
> **6** If a ***sea or lake*** is **calm**, the water is not moving…
> **7 Calm** ***weather*** is pleasant weather with little or no wind.
> **10** To **calm** ***a pain or an itch*** means to reduce it or to get rid of it.

Thus, **calm** as a verb is defined in both senses 5 and 10 by 'reduce' in identical syntactic frames (verb + direct object), the difference being only in the collocate noun naming the object. Senses 6 and 7 of **calm** as an adjective, while not defined by the same word, can be easily generalised as 'not moving', but the collocate nouns differ again. Obviously, also in this case it is the collocate nouns that determine the splitting of senses in the entry. In senses 6 and 7 *Cobuild* (1995) follows the tracks of its predecessor *COBUILD* (1987) where the number of senses differs from the later edition (only 6 senses for **calm** adjective, noun and verb), but the approach has been already shaped, e.g.:

> **3** A <u>sea or lake</u> that is **calm** does not have any waves because there is ***no strong wind***.
> **4** <u>Weather</u> that is **calm** is very still ***without any wind***.

The approach in *COBUILD* 1995 is more radical than in the later *LDOCE5* in one aspect at least: no effort is taken even to vary the wording of definitions: 'reduce' is used as the defining word for the verb **calm** with different collocates producing distinct senses. Otherwise, *LDOCE5* and *COBUILD* (1995) are similar in offering 'gradient' definitions for distinct senses of **taut** ('not relaxed' and 'firm') and of **calm** ('no strong wind' and 'without any wind'). This fits Rundell's idea of

'a mode where a word does not have separate meanings but rather a set of meaning potentials each of which may be activated in particular contexts' (Rundell, 2002: 147), also referred to by him as 'much fuzzier meaning – clusters, where a basic semantic core is elaborated, in real text, in a variety of ways.' (ibid.: 148).

CONCLUSIONS

The use of corpora in lexicography has highlighted the role of context, and, more narrowly, of collocations, in defining and distinguishing word senses in explanatory monolingual dictionary entries. The claim of the primacy of context in defining word senses leads to postulating a particular order of analytical procedures employed when constructing an entry: first contexts are grouped, and then word senses are deduced. Different collocational patterns justify the splitting of senses even when their definitions are identical or differ only in terms of gradience. The structure of dictionary entries, indeed, displays that meaning is treated as a category with blurred edges. Thus, collocations become the decisive factor in outlining sense distinctions. The advocates of the new hierarchy of context and meaning are the people directly involved in dictionary making – the authors cited above are all lexicographers who have to 'present' meanings to dictionary users. Faced with the enormous volume of observable quantifiable data, they felt compelled to revise the concept of meaning in favour of 'objective' (context) versus 'subjective' (word senses) factors, in order to make it operational for practical purposes. Thus, the pressure of raw data was the stimulus for redefining semantic analysis and the very concept of meaning for the needs of applied lexicographic description of language. The range of arguments supporting the revision of the concept of meaning is broad: from cognitive semantics (meanings are fuzzy concepts) to language acquisition studies (the perceptual perspective) within linguistics, and to natural sciences (classification of species in the natural world) beyond it.

REFERENCES

Bentivogli, L. and Pianta, E. (2000) *Looking for Lexical Gaps.* Euralex, vol. 2 (pp. 663–670).

Cowie, A. P. (2002) Examples and Collocations in the French 'Dictionnaire de langue'. In M.-H. Correard (ed.) *Lexicography and Natural Language Processing* (pp. 73–90). Euralex.

Evans, V. (2007) *A Glossary of Cognitive Linguistics.* Edinburgh, Edinburgh University Press.

Fontenelle, T. (2002) Lexical knowledge and natural language processing. In M.-H. Correard (ed.) *Lexicography and Natural Language Processing* (pp. 216–229). Euralex.

Hanks, P. (2002) Mapping meaning on to use. In M.-H. Correard (ed.) *Lexicography and Natural Language Processing* (pp. 156–198). Euralex.

Harris, R. and Taylor, T. J. (2005) *Landmarks in Linguistic Thought I: The Western Tradition from Socrates to Saussure*. London: Routledge.

Heid, U. (1994) On ways words work together – topics in lexical combinatorics. In *Euralex Proceedings* (pp. 226–257). Euralex.

Heid, U. (2004) On the presentation of collocations in monolingual dictionaries In *Euralex Proceedings*, vol. 2 (pp. 729–739). Euralex.

Kilgariff, A. (2002) Sketching words. In M.-H. Correard (ed.) *Lexicography and Natural Language Processing* (pp. 125–137). Euralex.

Kilgariff, A. ([1997] 2008) 'I don't believe in word senses' Computer and the Humanities 31: 91–113 the Netherlands: Kluwer Academic Publishers. Available from: http://arxiv.org/pdf/cmp-lg/9712006.pdf [Accessed on 31 August 2015].

Kokareviča, J. (2012) *Sense Organisation in Dictionary Entries in Two Editions of the 'Longman Dictionary of Contemporary English'*. MA thesis. Riga: UL.

Rundell, M. (2002) Good old-fashioned lexicography: Human judgment and the limits of automation. In M.-H. Correard (ed.) *Lexicography and Natural Language Processing* (pp. 138–155). Euralex.

Van Der Meer, G. (1998) Collocations as one particular type of conventional word combinations. Their definition and character. In *Euralex Proceedings*, vol. 1 (pp. 313–322).

DICTIONARIES ANALYSED

BBI Combinatory Dictionary of English Word Combinations (1986) Amsterdam/ Philadephia: John Benjamins.

[*COBUILD*] *Collins Cobuild English Dictionary* (1987) London: Harper Collins Publishers.

[*COBUILD*] *Collins Cobuild English Dictionary* (1995) London: Harper Collins Publishers.

[*LDOCE1*] *Longman Dictionary of Contemporary English* (1978) Essex: Longman.

[*LDOCE5*] *Longman Dictionary of Contemporary English* (2009) Essex: Longman.

Oxford Collocations Dictionary for Students of English (2002) Oxford: Oxford University Press.

Jeļena Dorošenko (Dr. Philol., Assoc. Prof.) is currently working at the University of Latvia. Her research interests include lexicography, semiotics and history of linguistics. Email: yelena-d@inbox.lv