

Paolo DiMuccio-Failla

## DISAMBIGUATING WORD SENSES THROUGH SEMANTIC CONDITIONS

### A project in learner's lexicography

**Keywords** Learner's lexicography; semantic conditions; phraseology; word sense disambiguation

We propose a novel type of learner's dictionary in the COBUILD tradition, based on the following general theoretical assumptions:

- 1) there are such things as basic units of word meaning, at least in the case of common words in normal every-day usage;
- 2) basic word senses are phraseological in nature, since they are not features of words in isolation but of their contextual patterns of typical usage – this hypothesis has been famously first formulated by John Sinclair (cf. Sinclair 1991);
- 3) such extended units of word meaning can be described by canonical forms of distinctive observable patterns, determined by collocation (the co-occurrence of particular words with the given word), colligation (the co-occurrence of particular grammatical patterns), semantic preference (the co-occurrence of words with particular meanings), and semantic prosody (a particular connotation of the described state of affairs or a particular attitude of the speaker, cf. definition by Louw 1993) – we call this *Sinclair's thesis* about lexical items (cf. Sinclair 1998; Sinclair 2004; Sinclair/Jones/Daley 2004).

On the other hand, we think that there is another independent aspect of word usage that contributes to distinguishing and disambiguating different senses: the *contextual scene construal* (cf. Langacker 1987), which is not always collocational in nature. For example, in disambiguating the fundamental (core) meaning of the verb *to follow*, the scene is usually so construed as to make clear that the person being followed was already in the act of going to a different place. To the best of our knowledge, no collocation corresponds to this state of affairs.

Hanks (2004) has noticed that semantic types alone do not suffice to provide clear distinctions between word senses, and even adding semantic roles to the picture does not solve the problem. In a previous publication (DiMuccio-Failla/Giacomini 2017) we pointed out one of the culprits: semantic types are not always linguistically collocated (i. e. they do not participate in collocations), even if *cognitively typical*. Take, for example, the verb *to toast*: there is no semantic type representing the statistically formed lexical set (cf. Jezek/Hanks 2010) of direct objects (*bread, sandwich, muffin, marshmallow, walnut, ...*) found in a corpus, yet it is hard to dispute that *breadstuff nut seed marshmallow* is the right semantic preference. So we all know that one usually toasts breadstuff, even if one does not usually say to toast breadstuff.

However, we are convinced that the main problem in finding the right semantic preferences is the need to add not only semantic roles but also *semantic conditions* to the picture. These conditions apply to the elements of a described scene in the *contextual scene construal*, giv-

ing a very robust contribution to word sense disambiguation (cf. Mennes/van der Waart van Gulik 2020 for critical discussion of state-of-the-art WSD).

Our semantic types are based on a growing ontology, linguistic in nature, very similar in flavor to WordNet but with distinctions which are obtained through phraseology and not through synsets. Patterns are identified by first determining colligations (valency distinctions as they are typically found in the dictionaries), semantic types, semantic roles, and conditions imposed on the participants in the scene. Finally, all relevant collocations are added. For the verb *agree*, for instance, one of the patterns of usage, corresponding to a single sense of the word, is the following:

**to AGREE** **⟨WITH a gv. person⟩** ...

■ **⟨THAT a ct. case is given⟩** [when knowing that this gv. person thinks so] **OR**

■ **⟨ON a gv. SUBJECT/MATTER/TOPIC/ISSUE OR ABOUT a pt. entity⟩** [when knowing this gv. person's opinion]

In this example, arguments are introduced by prepositions in bold characters and conditions are indicated in square brackets.

We tested our hypothesis by trying to match all usage examples found in selected public repositories (e.g. examples in the UK Dictionary at Lexico.com) with the previously defined patterns of usage. Our test was performed on the following verbs: one with very high semantic polysemy (*to follow*), one with medium degree polysemy (*to agree*), and verbs indicated by Jezek/Hanks (2010) as difficult to disambiguate neatly, e.g. *attend* and *finish*. We managed to attribute around 95% of the example sentences to the corresponding patterns. Our model seems to be able to resolve minimal distinctions in meaning while keeping ambiguous what is intrinsically ambiguous. In the framework of our lexicographic project, we intend to apply our approach to NLP word sense disambiguation.

## References

- DiMuccio, P./Giacomini, L. (2017): Designing an Italian learner's dictionary with phraseological disambiguators. In: Mitkov, R. (ed.): Computational and Corpus-Based Phraseology. Second International Conference. Europhras 2017, LNAI 10596. Cham, pp. 290–305.
- Hanks, P. (2004): Corpus pattern analysis. In: Proceedings of the XI EURALEX International Congress. Volume 1. Lorient, pp. 87–98.
- Jezek, E./Hanks, P. (2010): What lexical sets tell us about conceptual categories. In: Lexis 4, pp. 7–22.
- Langacker, R. W. (1987): Foundations of cognitive grammar. Stanford.
- Louw, B. (1993): Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In: Baker, M./Francis, G./Tognini-Bonelli, E. (eds.): Text and technology. In Honour of John Sinclair. Amsterdam, pp. 157–176.
- Mennes, J./van der Waart van Gulik, S. (2020): A critical analysis and explication of word sense disambiguation as approached by natural language processing. In: Lingua 243, pp. 1–14.
- Sinclair, J. M. (1991): Corpus, concordance, collocation. Oxford.
- Sinclair, J. M. (1998): The lexical item. In: Weigand, E. (ed.): Contrastive lexical semantics. (= Current Issues in Linguistics Theory 171). Amsterdam, pp. 1–24.
- Sinclair, J. M. (2004): New evidence, new priorities, new attitudes. In: Sinclair, J. M. (ed.): How to use corpora in language teaching. Amsterdam, pp. 271–299.

Sinclair, J. M./Jones, S./Daley, R. (2004): English collocation studies: The OSTI Report. London/New York, pp. xvii–xxix.

UK Dictionary (2022): <https://www.lexico.com> (last access: 06-04-2022).

## Contact information

**Paolo DiMuccio-Failla**  
University of Hildesheim  
[muccio@uni-hildesheim.de](mailto:muccio@uni-hildesheim.de)