

Multi word patterns and networks. How corpus-driven approaches have changed our description of language use

Kathrin Steyer

Due to the rise of corpus linguistics and the feasibility of studying language data in new quantitative dimensions, it became more and more evident that language use is fundamentally made up by fixed lexical chunks, set phrases, long distance word groups and multiword expressions (MWE). Sinclair's inductively reconstructed collocations (cf. 1991) and Hausmann's collocation pairs (cf. 2004) are the two leading concepts in collocation research. Basically, they are merely different ways of looking at the same fundamental principle of language, namely linguistic frozenness and fixedness. Compositional collocations and idioms differ in their degree of lexical fixedness and semantic opacity, their recognisability and prototypicality (Moon 1998, Burger et al. 2007). But they all share the most important characteristic: They are congealed into autonomous units in order to fill a specific role in communication. All these fragments are fixed patterns of language use (cf. Hunston/Francis 2000; cf. Hanks 2013). There is no core and no periphery. The difference is only in the degree of conspicuousness for the observer. These word clusters did not become fixed expressions by chance, but because there was a need of speakers for an economic way to communicate (cf. Steyer 2013).

Two assumptions constitute the basis of my talk:

- MWEs usually have multiple entries in the mental lexicon: on the one hand as more or less specified lexical units (lexemes) and on the other hand as (proto)typical realisations of a more abstract multiword pattern (MW pattern): for example [*für* ADJ *Ohren klingen*] (lit. 'to sound for ADJ ears) ADJ fillers: *deutsche* ('German') / *westliche* ('Western') / *heutige* ('contemporary') / *europäische* ('European') / *ungeübte* ('untrained').
- Independent of their lexical fixedness or variability, MWEs possess a holistic quality in the sense that they fulfil a specific role in communication as autonomous language units. This does not mean that they necessarily have an idiomatic meaning – sometimes they are completely transparent and compositional. The holistic quality can be attached to an abstract pattern and be functional in nature.

MWEs and MW patterns are not clear-cut and distinct entities. On the contrary, fragments and overlapping elements with fuzzy borders are typical for real language use. This means that there really are no MWEs as such. In real communicative situations, some components are focused while others fade into the background.

In my talk, I first discuss the nature of MW patterns that are reconstructed with complex corpus-driven methods. The examples are all taken from the *German Reference Corpus* (*Deutsches Referenzkorpus*) (cf. DeReKo) (located at the Institute for the German Language in Mannheim, IDS).

I show how we use an iterative methodology (quantitative - qualitative) to detect the nature of lexical fillers of pattern gaps and to visualize MWE hierarchies and networks. This

methodology includes complex phrase searches and reciprocal analysis with *COSMAS II* (The IDS Corpus Search, Management and Analysis System); collocation analysis (cf. Belica 1995) that not only detects significant word pairs, but also significant syntagmatic cotext patterns; and slot analysis with the help of our *UWV Tool* that allows us to bundle KWICs. At the end, I will present a vision of a pattern-based lexicographic representation for humans (“MWE fields”) (Steyer et al. 2013).

References

- Belica, C. 1995. *Statistische Kollokationsanalyse und Clustering. Korpuslinguistische Analysemethoden*. Mannheim: Institut für Deutsche Sprache.
- Burger, H., D. Dobrovolskij, P. Kühn and Neal R. Norrick. 2007 (eds.). *Phraseologie. Ein internationales Handbuch zeitgenössischer Forschung/Phraseology. An international Handbook of Contemporary Research*. (2 Editions). (= HSK 28, 1/2). Berlin/New York: de Gruyter.
- DeReKo 2014. *Deutsches Referenzkorpus. / Archiv der Korpora geschriebener Gegenwartssprache 2014-II* (Release 11.09.2014). Mannheim: Institut für Deutsche Sprache. Accessed on 17 March 2015. <http://www.ids-mannheim.de/DeReKo>.
- Hanks, P. 2013. *Lexical Analysis. Norms and Exploitations*. Cambridge, Massachusetts/London: The MIT Press.
- Hausmann, F. J. 2004. ‘Was sind eigentlich Kollokationen?’ In K. Steyer (ed.), *Wortverbindungen – mehr oder weniger fest. Jahrbuch des Instituts für Deutsche Sprache 2003*. Berlin/New York: de Gruyter, 309-334.
- Hunston, S. and G. Francis. 2000. *Pattern Grammar. A corpus-driven approach to the lexical grammar of English*. Amsterdam/Philadelphia: John Benjamins.
- Moon, R. 1998. *Fixed Expressions and idioms in English. A Corpus-Based Approach*. Oxford: Clarendon Press.
- Sinclair, J. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Steyer, K. 2013. *Usuelle Wortverbindungen. Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht*. (Studien zur Deutschen Sprache 60). Tübingen: Narr.
- Steyer, K., A. Brunner and C. Zimmermann 2013. *Wortverbindungsfelder Version 3: Grund*. Accessed on 31 March 2015. <http://wvonline.ids-mannheim.de/wvfelder-v3/index.html>.