

## A model of corpus-based co-occurrence contrastive analysis: Case study of light verb construction in German and Polish

**Research question.** The objective of the proposed model is to establish the equivalence of phrases from two language systems (German and Polish) based on their setting, i.e. lexical setting, syntax, morphology and usage – “functional equivalence” (Dobrovolskij 2015, 277). The model allows determining syntagmatic patterns and patterns of use of analyzed units, which on the one hand serve to establish functional equivalence, and on the other hand, they constitute complete linguistic material to be used in teaching and lexicography.

**Method and results.** The model (Taborek 2018, cf. fig. 1) involves three principal stages of analysis:

- a) The first stage involves the analysis of source units (light verb constructions) based on data from monolingual corpora. The analysis based on finding co-occurrences with the highest frequency rate (cf. Steyer 2013). The list of co-occurrences is subjected to linguistic analysis and categorization, which involve word classes, syntactic functions, and semantic distinction. For the German light verb construction *in Vergessenheit geraten* ‘fall into oblivion’ the most frequent co-occurrences – *fast* ‘almost’, *schnell* ‘quickly’, *zu Unrecht* ‘wrongly’, *Brauch* ‘custom’ are arranged morphologically and syntactically. Then within syntactic functions, items are semantically classified, in which adverbial co-occurrences express time (*schnell*), degree (*fast*) and valuation (*zu Unrecht*). Arranging co-occurrences according to their syntactic function makes it possible to establish the “syntagmatic pattern” (Steyer 2013), e.g. SUB + *gerät* + (ADV) + *in Vergessenheit*. The next stage of generalization consists in replacing syntactic functions with variables X, Y, Z, etc., which makes it possible to establish three patterns (Hunston/Francis 1999) X *gerät* ADV *in Vergessenheit*, X *gerät* ADV *in Vergessenheit und wird wieder entdeckt*, and X *droht in Vergessenheit zu geraten*.
- b) The second stage of analysis consists in the identification of **potential equivalents** based on bilingual dictionaries, lexical bases, bilingual corpora, translation support tools and translation memories (cf. Mellado Blanco 2019). The equivalents are subjected to verification based on objective, empirical and corpus-based methods.
- c) The third stage is a corpus-based analysis of **co-occurrence of potential equivalents** conducted in a similar way to the analysis of a source unit in a). Based on the data from monolingual corpora, a list of the most frequent co-occurrences is created. These co-occurrences are arranged morphologically, syntactically, semantically, and pragmatically. Further steps include the determination of syntagmatic patterns and patterns of use of potential equivalents in a similar manner to the analysis of source language units.

Then a source unit is compared and contrasted with potential equivalents of (i) co-occurrences, (ii) syntagmatic patterns, and (iii) patterns of a source unit with co-occurrence, syntagmatic patterns and patterns of a potential equivalent. If co-occurrences, patterns and schemes display similarity, then it is possible to speak of functional equivalence, as defined by Dobrovolskij (2015, 277).

**Short conclusions.** A sample corpus-based analysis of the example *in Vergessenheit geraten* has found that its functional equivalent is *odchodzić w niepamięć/zapomnienie* while lexicographic sources provide different equivalents, such as *iść* (pf. *pójść*)/*popadać* (pf. *popaść*) *w niepamięć/zapomnienie*. The results of sample analyses of light verb constructions show the need for applying co-occurrence that reflects the context of use as a **tertium comparationis** in order to determine functional equivalents, which, besides light verb constructions, can include both verbs, e.g. *in Frage stellen* – *kwestionować* and prepositional phrases, e.g. *in Not geraten(e)* – *w potrzebie*.

## References

- Dobrovól'skij, D. 2015. "Introduction." *International Journal of Lexicography* 28(3), 275–278.
- Hunston, S./Francis, G. 2000. *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam/Philadelphia: Benjamins
- Mellado Blanco, C. 2019. "Phrasem-Konstruktionen kontrastiv Deutsch–Spanisch: ein korpusbasiertes Beschreibungsmodell anhand ironischer Vergleiche." *Yearbook of Phraseology* 10: 65–88.
- Steyer, K. 2013. *Usuelle Wortverbindungen. Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht*. Tübingen: Narr.
- Taborek, J. 2018. Korpusbasiertes kontrastives Beschreibungsmodell für Funktionsverbgefüge. In: Schmale, Günter (ed.): *Lexematische und polylexematische Einheiten des Deutschen*, Tübingen: Stauffenburg, 135-154.

**Fig. 1. A model of corpus-based co-occurrence contrastive analysis**

