

Carole Tiberius (University of Leiden) & Lut Colman

Title

Corpus Pattern Analysis for Dutch in *Woordcombinaties*

Abstract

In this presentation, we describe the application of Corpus Pattern Analysis (CPA) to Dutch in the context of [Woordcombinaties](#) (Word Combinations). *Woordcombinaties* is an online phraseological lexicographic resource combining information on collocations, idioms, conversational routines and patterns in one tool. For collocations, the example of Sketch Engine for Language Learning (SkeLL) is followed, whereas the patterns are encoded according to Hanks' CPA technique (Hanks 2004, 2013). In the context of the project *Woordcombinaties* this technique has been tailored to the needs of the target audience (advanced learners of Dutch as a second or foreign language). Sentences are sorted before annotation, using a specially developed GDEX configuration in Sketch Engine to enable the output of short, comprehensible and yet informative sentences. Furthermore, arguments are represented by indefinite pronouns *iemand* 'someone', *iets* 'something', *ergens* 'somewhere', *zo* 'such' (or combinations thereof) in the pattern slots where this is possible for increased readability and the corresponding semantic types and syntactic functions are accessible through a tool tip. The individual slots are also enriched with collocations offering a kind of advanced word sketch in the patterns. Idioms and conversational routines are currently encoded as special instances among the collocations and the patterns. However, a more elaborate treatment is foreseen to enable specific search options, such as the possibility to search for idioms based on image categories, such as 'food' for *een vinger in de pap hebben* 'have a finger in the pie'.

References

- Hanks, P. (2004). Corpus pattern analysis. In Proceedings of the 11th EURALEX International Congress, pp. 87-98.
- Hanks, P. (2013). Lexical analysis: Norms and exploitations. Cambridge, MA: The MIT Press.