Sabine Schulte im Walde & Maximilian Köper (University of Stuttgart):

The role of ambiguity and abstractness in multi-modal models of German noun compounds and particle verbs

A standard multi-modal model integrating

(a) corpus-based textual co-occurrence features from a large German web corpus, and

(b) images downloaded from bing.de

is exploited to automatically predict the degree of compositionality for German noun-noun compounds (such as *Feuerwerk 'fireworks'*) and German particle verbs (such as *anstrahlen 'beam/smile at'*). Our models confirm previous insights that combining linguistic and perceptual data outperforms the usage of the individual information (Andrews et al., 2009), and that the imageability of the target multi-word expressions plays an important role for the quality of the predictions (Kiela et al., 2014). At the same time, we were surprised that the model obtains better results for the particle verbs, which are more ambiguous and less concrete than the noun compounds.

Our contribution to the symposium will explore the lexical, empirical and perceptual properties of the German particle verbs (PVs) that might play a role in predicting the degrees of compositionality:

(i) corpus frequencies of the PVs;

(ii) the degree of PV ambiguity, and the influence of the predominant sense;

(iii) the concreteness vs. abstractness of associations to PV images; and

(iv) the impact of imageability.

Our analyses rely on large-scale web corpus data, dictionary entries of the number and definition of senses, a new collection of associations to verb images, and a semi-automatic resource of affective norms (Köper & Schulte im Walde, 2016). We capture particle verbs across four selected particles: *ab, an, auf* and *aus*.

References:

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review,* 116(3):463-498.

Kiela, D., Hill, F., Korhonen, A., & Clark, S. (2014). Improving multi-modal representations using image dispersion: Why less is sometimes more. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics,* pp. 835-841, Baltimore, Maryland, USA.

Köper, M., & Schulte im Walde, S. (2016). Automatically generated affective norms of abstractness, arousal, imageability and valence for 350.000 German lemmas. In: *Proceedings of the 10th International Conference on Language Resources and Evaluation,* pp. 2595–2598, Portoroz, Slovenia.