A corpus annotated for metaphor in German

Markus Egg
Humboldt-Universität zu Berlin
markus.egg@hu-berlin.de

Valia Kordoni
Humboldt-Universität zu Berlin
evangelia.kordoni@hu-berlin.de

We present work on a German corpus annotated for metaphor with material from different registers (Halliday and Hasan, 1985), like in the VU Amsterdam Metaphor Corpus (Steen et al., 2010): speeches of the Parlamentsreden-Korpus (Blaette 2017), newspaper commentaries (including the Potsdam Commentary Corpus, Stede 2004), sermons, light fiction, and debates from competitions of 'Jugend debattiert'. In text parsed by the Stanza package (Qi et al 2020) we annotate in the INCEpTION tool (Klie et al. 2018) the metaphors and the contexts that trigger their metaphorical interpretation, a strategy also pursued in Dipper et al. (2021). For the annotation, we developed extended guidelines on the basis of the 'Metaphor Identification Procedure-VU' (Steen et al., 2010; Herrmann et al., 2019).

Non-conventionalised and extended metaphors show up predominantly in highly persuasive registers. The expectation that oral and fictional discourse are low on metaphor (Steen et al. 2010) could only be confirmed for fiction. From the viewpoint of tenor (author-recipient relations), the closeness of the interlocutors correlates inversely with overall metaphoricity. Also, non-conventionalised and extended metaphors occur more frequently in registers with a hierarchical difference between authors and recipients. Sermons had the highest degree of non-conventional metaphors, with extended metaphor emerging as a clear register marker for sermons.

References: Blaette, A. (2017). GermaParl. Corpus of plenary protocols of the German Bundestag. • Dipper, S., Ehlert, A., Scholz, D., Theodor, F., and Weber, L. (2021). Recognizing deliberate metaphors. DGfS-CL 2021. • Halliday, M. and Hasan, R. (1985). *Language, context and text*. • Herrmann, B., Woll, K., and Dorst, A. (2019). Linguistic metaphor identification in German. Susan Nacey et al. (eds), *Metaphor identification in multiple languages*. 113-135. • Klie, J.-C., Bugert, M., Boullosa, B., de Castilho, R. E., and Gurevych, I. (2018). The inception platform: Machine- assisted and knowledge-oriented interactive annotation. *COLING 2018: system demonstrations*. • Qi, P., Zhang, Y., Zhang, Y., Bolton, J., and Manning, C. (2020). Stanza: A Python natural language processing toolkit for many human languages. *ACL 2020: system demonstrations*. • Stede, M. (2004). The Potsdam Commentary Corpus. Bonnie Webber et al. (eds), *ACL 2004 Workshop on Discourse Annotation*, 96-102. • Steen, G., Dorst, A., Herrmann, B., Kaal, A., Krennmayr, T., and Pasma, T. (2010). *A method for linguistic metaphor identification: from MIP to MIPVU*.