Improved metaphor identification via (contextual) distributional semantic representations.

Gerold Schneider[1] , Johannes Fröh[2] & Janis Goldzycher[1]
[1]University of Zurich, [2]University of Bonn
gschneid@ifi.uzh,ch, jofroe@uni-bonn.de, goldzycher@cl.uzh.ch

A solid agreement exists on Firth's quote "You shall know a word by the company that it keeps". Yet, metaphors, defined by Gerard Steen as "[linguistic, cognitive, and communicative] cross-domain mapping in conceptual structure" (Steen, 2011), create meaning through the collocation of words that, otherwise, keep quite contrary company. We present an approach for automatic metaphor identification using metaphor's unique semantic cross-domain mapping as an identification feature through the combination of collocation measures with distributional semantics as a starting point for semi-supervised detection.

Based on recent advancements in distributional semantics (e.g., Baroni & Lenci, 2010), we hypothesize that metaphors are non-compositional collocations, as their lexical participants are far away in semantic space (Ricoeur 1977). We combine traditional frequentist approaches with distributional semantics measures using *word2vec* for improved idiom and metaphor identification using the cosine distance in the semantic space between the participants, and replaceability. We evaluate the performance of this approach on four constructions:

(1) Verb-preposition collocations: we report increases in F-score by 10%; (2) Verb-object constructions: we report increases in F-score by 80%; (3) Idiomatic noun compounds: we show that two thirds of the gold standard ranking in Reddy et al. (2011) can be predicted by the cosine distance; (4) Metaphors, by means of a gold standard from salient biblical references in contemporary social media.

Our second research focus is the identification of new metaphors. We hypothesize that a given metaphor's cross-domain mapping is expressed via the vectorial relation $\vec{v}$ between the participating words *w1* and *w2* (Steen, 2011). New metaphors are identifiable as the translation of $\vec{v}$ across the semantic space, to related words *w3* and *w4*. We test this hypothesis utilizing a religious communication dataset, as religion heavily draws on metaphorical language that is translated into contemporary contexts and wordings (Ricoeur, 1977). In the process, we extract religious quotes that are particularly salient.

References: • Baroni, M. & A. Lenci (2010). "Distributional Memory: A general framework for corpus-based semantics". Computational Linguistics, 36, 4, 673-721. • Reddy, S., D. McCarthy & S. Manandhar (2011). "An Empirical Study on Compositionality in Compound Nouns". Proceedings of 5th International Joint Conference on Natural Language Processing, 210-218. • Ricoeur, P. (1977). The Rule of Metaphor. Translated by R. Czerny, K. McLaughlin & J. Costello. New York, London: Routledge. • Steen, G. (2011). The

contemporary theory of metaphor - now new and improved!. Review of Cognitive Linguistics, 9(1), 26-64.