

Predicting the interpretational diversity of novel compounds in an exploratory-confirmatory study

Fritz Günther¹, Melanie Bell², Martin Schäfer³

¹ Humboldt-Universität, Berlin, Germany

² Anglia Ruskin University, United Kingdom

³ Leipzig University, Germany

Background: The meaning of all linguistic expressions is ambiguous to some extent and can be interpreted in various different ways. The vast majority of words are at least somewhat polysemous and express more than one meaning. In compound words such as /thunderstorm/ or /tailgate/, this ambiguity is amplified: Not only can both constituents be polysemous, also the relation between them is underspecified and open to different interpretations [1]. And within the class of compound words, this under-specification is especially relevant for novel compounds such as /gold student/ or /star model/, where these interpretations are not yet conventionalized through lexicalisation [1–3]. Notably, novel compounds differ in their interpretational diversity: While some (such as /gold student/) are interpreted very consistently by different speakers, others (such as /face room/) show considerable interpretational diversity [1, 3]. It remains an open question which factors explain those differences in interpretational diversity. Although previous studies have shown that word embeddings (meaning representations derived from text-based language models), can encode the interpretational diversity of such expressions, these previous studies have been limited to a small, rigid and high-level closed set of relational interpretations (e.g., ‘student MADE OF gold’, ‘student ABOUT gold’) [1].

Method & Results: In contrast to previous work, the present study uses open-format interpretations provided by human participants, who were presented with novel compound words and asked to freely type what these mean. These open-format interpretations were classified by annotators in a bottom-up manner in order to compute quantitative measures of interpretational diversity. For our first exploratory study, we used a dataset from a previous study that contained these measures for 45 English novel compound words [3]. As potential predictors of interpretational diversity, we considered a range of semantic measures derived from word embedding models. The vector norm of the embeddings –a measure of meaning specificity of a word [4] that for existing words is correlated to their number of senses –emerged as the best predictor: Compounds with lower vector norms (and thus, less specific meanings according to the model) also showed higher interpretational diversity.

Based on these results, we performed a simulation-based power analysis to estimate the sample size required to replicate a vector norm effect of the observed effect size. We subsequently performed a high-powered confirmatory study with 193 English novel compounds, which were systematically sampled for maximal variation of vector norm. We presented sets of 22 to 25 of these novel compounds to a total of 160 native English speakers, collected open-format interpretations, had these annotated and classified, and estimated the interpretational diversity of these compounds (following the procedure by [3]). In this confirmatory study, the effect of vector norm identified in the exploratory study was replicated.

Discussion: Our study shows that text-based word embedding models encode the unconstrained interpretational diversity of linguistic expressions, even for novel expressions that have never been observed in their training data. This shows that a model that learns meaning representations only from experience with text data is able to pick up these nuances of linguistic information, and that the required information to do so can be learned and extrapolated from the patterns of natural language use.

References

- [1] Schmidtke, D., Gagné, C., Kuperman, V., Spalding, T., & Tucker, B. (2018). Conceptual relations compete during auditory and visual compound word recognition. *Language, Cognition and Neuroscience*, 33, 923–942. <https://doi.org/10.1080/23273798.2018.1437192>

-
- [2] Günther, F., & Marelli, M. (2022). Patterns in CAOSS: Distributed representations predict variation in relational interpretations for familiar and novel compound words. *Cognitive Psychology*, 134, 101471. <https://doi.org/10.1016/j.cogpsych.2022.101471>
 - [3] Ryder, M. (1994). *Ordered chaos: The interpretation of English noun-noun compounds* (Vol. 123). University of California Press.
 - [4] Schäfer, M., & Bell, M. (2020). Constituent polysemy and interpretational diversity in attested English novel compounds. *The Mental Lexicon*, 15(1), 42–61. <https://doi.org/10.1075/ml.00013.sch>
 - [5] Oyama, M., Yokoi, S., & Shimodaira, H. (2023). *Norm of word embedding encodes information gain*. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.emnlp-main.131>