

Quantifying similarities in the interaction between prefix, base class, and individual base in scalar *out*-prefixation

Sven Kotowski (Heinrich-Heine-Universität Düsseldorf), Martin Schäfer (Universität Tübingen)

It is well-known that English *out*-prefixation productively gives rise to scalar-comparative interpretations (Kotowski 2020; Talmy 2000). As shown by the underlined contextual information in (1)-(2), scalar dimensions are not necessarily fixed on the lexical level. Thus, the same derivative, *outsing* here, allows for different interpretations regarding the bases for comparisons, the LOUDNESS- and QUALITY-dimensions, respectively:

- (1) Good vocal control is essential, a backing singer must not try to **‘out-sing’ the lead vocalist** [...] Whilst most sound engineers and producers will adjust vocal volumes in the mix, it is important for singers [...] to back off the microphone a bit. (iWeb)
- (2) LBT is my guilty pleasure, you can't deny their vocal abilities [...] nobody out there can **outsing them** from a technical standpoint. (iWeb)

At the same time, it appears unlikely that a given base of an *out*-derivative will allow for arbitrary dimensions of comparison. This paper investigates whether scalar *out*- has a uniform effect within as well as across different verb classes by combining qualitative and quantitative analyses. We hypothesize that (I) the verb-class of the base form is a major determinant of the target dimensions for scalar *out*-, while at the same time (II) the common scalar semantics of the derivatives and the change in syntactic argument structure will also make all *out*-derivatives more similar across verb-classes.

Hypothesis (I) is investigated in two iWeb-corpus studies (Davies 2018): (Ia) with all possible bases from seven different VerbNet classes (Kipper et al. 2008), and (Ib) with 100 tokens each of 12 *out*-lemmas from three of these classes (RUN, EXIST, and PERFORMANCE). We show that all verb classes exhibit clear-cut dimension preferences, that preference-strength correlates with lower degrees of polysemy, and that individual lemmas within a class largely follow the same pattern. Divergent behavior of individual lemmas can be explained by polysemy as derived from cross-listing in different VerbNet-classes. For example, *out*-verbs with bases from the PERFORMANCE-class show significant bias for the QUALITY-dimension, while its member OUTSING shows a preference for LOUDNESS. This discrepancy is reflected in the cross-listing of SING as a SOUND-EMISSION-verb.

Hypothesis (II) is investigated using distributional semantics measures (Sahlgren 2006) extracted from ukWaC (Baroni et al. 2009) and iWeb. We used a two-pronged approach: First, we extracted similarity measures for four pairs of derivatives and their bases from each verb class used in Study (Ib). Second, to allow for contextualization, we extracted similarity measures for a control group of such word pairs with other prefixes from overlapping verb classes. Base-base and derivative-derivative similarity measures are indicative of a uniform effect of semantic narrowing for scalar *out*- only. Across the three *out*-sets, the PERFORMANCE class shows the least uniformity, in line with its high cross-listing score identified in Study (Ia).

Our findings underline the importance of semantics in understanding derivational patterns, in particular semantic narrowing of base interpretations, and show how distributional semantics can support and refine qualitative analyses.

REFERENCES

- Baroni, M., Bernardini, S., Ferraresi, A. and Zanchetta, E. 2009. The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3). 209–226. 10.1007/s10579-009-9081-4
- Davies, M. 2018. The 14 Billion Word iWeb Corpus. Available online at <https://corpus.byu.edu/iWeb/>.
- Kipper, K., A. Korhonen, N. Ryant & M. Palmer. 2008. A large-scale classification of English verbs. *Language Resources and Evaluation* 42(1). 21–40.
- Kotowski, S. 2020. The semantics of English *out*-prefixation: A corpus-based investigation. *English Language and Linguistics*. 1-29. doi:10.1017/S1360674319000443
- Sahlgren, M. 2006. *The Word-Space Model. Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*. PhD thesis, Univeristy of Stockholm.
- Talmy, L. 2000. *Toward a Cognitive Semantics. Vol. 2: Typology and Process in Concept Structuring*. Cambridge, MA: MIT Press.