Productivity and semantic transparency:
An exploration of compounding in Mandarin

Shen Tian and Harald Baayen

Shanghai Jiao Tong University and Eberhard-Karls University, Tübingen

In structuralist linguistics, compounds are argued not to constitute morphological categories, due to the absence of systematic form-meaning correspondences. This study investigates subsets of compounds for which systematic form-meaning correspondences are present: adjective-noun compounds in Mandarin. We show that there are substantial differences in morphological productivity for polar adjectives in nominal adjective-noun compounds. One set of productivity measures (the count of types, the count of hapax legomena, and the estimated count of unseen types) (Baayen, 1992) reflect compounds' profitability or 'extent of use'. By contrast, the category-conditioned degree of productivity $\mathcal{P}$ characterizes the internal semantic systematicity of the words belonging to a morphological category, and its potential for extension to new words. We provided evidence that greater semantic transparency, as gauged with semantic vectors (also known as word embeddings), predicts greater category-conditioned productivity (Shen and Baayen, 2021). Consistent with this finding, a linear discriminative learning model (Baayen et al., 2019) trained to predict words' semantic vectors from their triphone vectors performs with greater accuracy for more productive adjectives.

We are currently extending this approach to other sets of Mandarin compounds. One such set comprises verb-noun compounds with as verbal constituents antonymous verbs (e.g. go up/go down, remember/forget, buy/sell) in 上学 (*shang4xue2*, go up school, 'go to school'), 记仇 (*ji4chou2*, remember hatred, 'hold a grudge'), and 卖国 (*mai4guo2*, sell country, 'treason'), and a second such set comprises X-person compounds, where X can be any major word category (e.g 高徒, *gao1tu2*, tall apprentice, 'brilliant student'; 影迷, *ying3mi2*, 'movie fan'; 动人, *dong4ren2*, touch man, 'touching'). For both sets of compounds, just as for AN compounds, we again observe a positive correlation between semantic transparency and degree of productivity, as gauged by the $\mathcal{P}$ measure. Whereas for the AN and VN compounds, two measures of transparency show significant correlations with $\mathcal{P}$, only one such measure is predictive for X-person compounds, possibly because the latter formations are semantically more heterogeneous. Our current results dovetail well with the hypothesis that semantic transparency is a prerequisite for a word formation process to be productive.

References

Baayen, R. H. (1992). A quantitative approach to morphological productivity. In Booij, G. E. and Marle, J. v., editors, *Yearbook of Morphology 1991*, pages 109–149. Kluwer, Dordrecht.

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., and Blevins, J. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*.

Shen, T. and Baayen, R. H. (2021). Adjective–noun compounds in Mandarin: a study on productivity. *Corpus Linguistics and Linguistic Theory*.