**Variability in the relation between prominence and semantics in English compounds**

Dominic Schmitz[1], Ingo Plag[1] & Melanie J. Bell[2]
*[1]Heinrich-Heine-Universität Düsseldorf, [2]Anglia Ruskin University*

Previous research highlighted many factors influencing the variation we find in compound prominence (as in *chíld care* vs. *home phóne*), mostly related to semantics (e.g. Bell & Plag, 2012, 2013). However, the nature of this variation remains unclear and raises crucial concerns about the reality and theoretical treatment of the commonly assumed categorical binary stress patterns (e.g. Kunter, 2010, 2011).

Following ideas in recent research on pitch contours in Mandarin (Chuang et al., 2024), the current study implements a linear discriminative learning model (LDL; Baayen et al., 2019) in which the meaning of compounds is predicted from their pitch and intensity contours (as a proxy for a compound's prosody), and vice versa.

Using 760 English compound tokens from the Boston University Radio Speech Corpus (Ostendorf et al., 1996; Plag et al., 2008), LDL networks were trained on the combined pitch and intensity contours, and on context-dependent semantic vectors computed with BERT (Devlin et al., 2018). Testing the trained networks on held-out compound tokens, we found that in 39 % of cases the model identified the correct compound semantics based on the combined pitch and intensity contours. The baseline for correctly identifying a token by chance was at 9 %.

The present study provides first evidence suggesting that the acoustic signal on the one hand, and context-dependent semantics on the other are closely connected. Using prosodic information from a form's signal alone, one can predict its semantics.

Regarding the puzzle of prominence in noun-noun compounds, our findings point away from discrete factors (semantic relation, semantic classes of the constituents) used in previous research and instead advocate an approach that leverages the features of the phonetic signal and the corresponding semantics in a discriminative learning process.

## References


Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. P. (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*, *2019*, 4895891. https://doi.org/10.1155/2019/4895891

Bell, M. J., & Plag, I. (2012). Informativeness is a determinant of compound stress in English. *Journal of Linguistics*, *48*(3), 485–520. https://doi.org/10.1017/S0022226712000199

Bell, M. J., & Plag, I. (2013). Informativity and analogy in English compound stress. *Word Structure*, *6*(2), 129–155. https://doi.org/10.3366/word.2013.0042

Chuang, Y.-Y., Bell, M. J., Tseng, Y.-H., & Baayen, R. H. (2024). Word-specific tonal realizations in Mandarin. *Preprint*.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*, 4171–4186. https://arxiv.org/abs/1810.04805v2

Kunter, G. (2010). Perception of Prominence Patterns in English Nominal Compounds. *Proceedings of Speech Prosody 2010*, 1–4.

Kunter, G. (2011). *Compound stress in English: The phonetics and phonology of prosodic prominence*. De Gruyter Mouton.

Ostendorf, M., Price, P., & Shattuck-Hufnagel, S. (1996). *Boston University Radio Speech Corpus*. Linguistic Data Consortium.

Plag, I., Kunter, G., Lappe, S., & Braun, M. (2008). The Role of Semantics, Argument Structure, and Lexicalization in Compound Stress Assignment in English. *Language*, *84*(4), 760–794. https://doi.org/10.1353/LAN.0.0072