

**English compound semantics are predictable by compound prominence
and vice versa [Poster]**

*Dominic Schmitz (HHU Düsseldorf), Melanie J. Bell (Anglia Ruskin University) & Ingo
Plag (HHU Düsseldorf)*

[dominic.schmitz](mailto:dominic.schmitz@uni-duesseldorf.de), [ingo.plag](mailto:ingo.plag@uni-duesseldorf.de)}@uni-duesseldorf.de, melanie.bell@aru.ac.uk

According to reference grammars of English, compound nouns can either be left-stressed, as in *Oxford Street* and *opera singer*, or right-stressed, as in *Oxford Road* and *summer dress* (Bauer et al., 2013). However, extensive empirical research has demonstrated that compound prominence is not categorical but rather varies probabilistically. Prominence has been found to be influenced by several lexical-semantic factors, including the semantic relation between the constituent nouns, their respective semantic categories, and their degree of semantic specificity (e.g., Plag et al., 2008; Bell & Plag, 2012). Although prominence patterns are known to differ across speakers and contexts, no previous study has quantitatively analysed both contextual and lexical effects on the pronunciation of compound nouns or explored how speakers systematically utilise these effects.

This study analyses spoken compound noun tokens from the Boston University Radio Speech Corpus (Ostendorf et al., 1996). Part of this corpus consists of news stories read in laboratory conditions by professional newsreaders. Because every speaker reads the same set of stories containing numerous compound nouns, the corpus is particularly well-suited for disentangling effects related to interspeaker variation, context, and lexical semantics on compound pronunciation. We represent the prosodic form of each token as smoothed pitch and intensity contours extracted using the rPraat package in R (Bořil & Skarnitzl, 2016). These contours are then analysed with generalised additive models (GAMs) to predict their shape based on the compound itself, the speaker, and the surrounding context. Finally, we employ a computational learning approach, specifically the Discriminative Lexicon Model (Baayen et al., 2019), to evaluate the degree to which associations between context-specific prosodic forms and meanings can be learned.

Variability in stress assignment in compounds has been central to debates regarding the nature and role of symbolic rules, associative networks, and analogical mechanisms in linguistic organisation. This study provides initial evidence suggesting a close relationship between acoustic signals and context-dependent semantics. Specifically, the semantics of a compound can be predicted from prosodic information alone at above-chance accuracy. This indicates that human language processing may directly map acoustic signals onto semantic content, with abstract categories such as semantic relations and semantic classes, commonly employed in previous compound research, potentially emerging from this direct mapping.

Literatur

- 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. doi: 10.1155/2019/4895891
- Bauer, L., Lieber, R. & Plag, I. (2013). *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.
- Bell, M. J., & Plag, I. (2012). Informativeness is a determinant of compound stress in English. *Journal of Linguistics*, 48(3), 485–520. doi: 10.1017/S0022226712000199
- Bořil, T., & Skarnitzl, R. (2016). Tools rPraat and mPraat. doi: 10.1007/978-3-319-45510-5_42
- Ostendorf, M., Price, P., & Shattuck-Hufnagel, S. (1996). *Boston University Radio Speech Corpus*. Linguistic Data Consortium.