

Usage-based implicit grammar and discrimination learning

Harald Baayen

Implicit grammar is a usage-based theory of language in which the consequences of discrimination learning are taken seriously and investigated quantitatively using corpus-based computational models. According to this approach, a substantial part of knowledge of grammar builds up over the lifetime through implicit learning, with continuous fine-tuning of the association strengths between cues (features) and outcomes (classes to be discriminated). Furthermore, it is regarded as an empirical question whether classic linguistic constructs such as phonemes, morphemes, or multi-word units are required.

In this workshop, I will first discuss two examples that illustrate the new kinds of explanations of linguistic phenomena that become possible within the framework of discrimination learning. First, I will discuss a model for auditory comprehension that is trained on real conversational speech, and recognizes words spliced out of such speech with human-like accuracy, without ever seeking to identify phonemes, morphemes, or word forms (such as "yesterday" and "yeshay"). Instead, the model learns to discriminate between meanings straight from low-level acoustic features. Interestingly, both the model and native speakers only correctly identify some 30% of the words spliced out of real conversational speech that they were presented with. This low recognition rate indicates that the initial stage of word recognition does not yield a crisp and clear verdict on what is in the acoustic signal, and that context-driven expectations are crucial for making sense of spontaneous speech.

The many different variants that recent corpus studies have documented for idioms raise the question of whether idioms could be recognized along similar lines as the reduced variants of a word such as "yesterday". I will present a corpus-inspired simulation study that distinguishes between meanings, including those of idioms, on the basis of sublexical features, and that provides a tentative explanation for why idioms can be semantically idiosyncratic and yet formally flexible.

Having presented examples of what can be achieved with discrimination learning, we continue with a hands-on lab session that introduces the `ndl` package for R and illustrates how this package can be used to build predictive computational models grounded in error-driven discrimination learning.

References

- Arnold, D., Tomaschek, F., Lopez, F., Sering, T. and Baayen, R. H. (2016). From speech to comprehension without phonemes. Submitted.
- Baayen, R. H. and Ramscar, M. (2015). Abstraction, storage and naïve discriminative learning. In Dabrowska, E. and Divjak, D. (eds), *Handbook of Cognitive Linguistics*, 99-120. Berlin: De Gruyter Mouton.
- Baayen, R. H., Shaoul, C., Willits, J. and Ramscar, M. (2015). Comprehension without segmentation: A proof of concept with naive discrimination learning. *Language, Cognition, and Neuroscience*.
- Geeraert, K., Newman, J. and Baayen, R. H. (2016). Idiom flexibility: experimental data and a computational model. Submitted.