Multi-word units in a discriminative framework

Saskia E. Lensink
Leiden University
s.e.lensink@hum.leidenuniv.nl

R. Harald Baayen
Universität Tübingen
harald.baayen@uni-tuebingen.de

Phrasal frequency effects (Bannard & Matthews, 2008) have often been taken as evidence for the existence of representations for multi-word units. However, Baayen et al. (2013) were able to simulate phrasal frequency effects using a naive discriminative learning (NDL) network that does not rely on representations for multi-word units, by connecting letter digraphs to the constituent words of the multi-word units.

Even though it may sometimes be possible to model phrasal frequency effects without relying on representations of multi-word units, it is not clear whether this is the most optimal architecture to use. Therefore, we built an NDL network where a layer of input cues, consisting of the constituent words of a set of 300 semantically transparent trigrams, and a layer of outcomes, the trigrams themselves, are connected.

The validity of this NDL architecture was tested against two experimental data sets where stimuli consisted of these 300 trigrams. Statistical analyzes showed that the NDL measures performed to comparable levels as traditional frequency measures in explaining the empirical data. This testifies to the usefulness of an NDL architecture with full-form representations for semantically transparent multi-word units, and with their constituent words used as learning cues.

One problem noted by Baayen et al. (2013) is that it is unclear how multi-word units can be distinguished from each other. Our discriminative network provides a first tentative solution to this question. We will also argue that a discriminatory perspective clarifies why multi-word units have to be short (with a most five words).

References: • Baayen, R. H., Hendrix, P., & Ramscar, M. (2013). Sidestepping the combinatorial explosion: An explanation of n-gram frequency effects based on naive discriminative learning. *Language and Speech, 56*(3), 329-347. • Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning the effect of familiarity on children's repetition of four-word combinations. *Psychological science*, *19*(3), 241-248.