

Signal evolution within the word

Andrew Wedel

Languages have been shown to optimize their lexicons over time with respect to the amount of signal allocated to words relative to their informativity: words that are on average less predictable in context tend to be longer, while those that are on average more predictable tend to be shorter (Piantadosi et al 2011, cf. Zipf 1935). Further, psycholinguistic research has shown that listeners are able to incrementally process words as they are heard, progressively updating inferences about what word is intended as the phonetic signal unfolds in time. As a consequence, phonetic cues early in the signal for a word are more informative about word-identity because they are less constrained by previous segmental context. This suggests that languages should not only optimize the total amount of signal allocated to different words, but optimize the distribution of that information across the word relative to existing competitors in the lexicon. Specifically, words that are on average less predictable in context should evolve highly informative phonetic cues early in the word, while tending to preserve a 'long tail' of redundant cues later in the word. In this talk I will review recent work in our group showing that these predictions are borne out in several languages.

I will also present recent statistical work in our group supporting the hypothesis that languages tend to develop phonological rules which enhance phonetic cue informativity at the beginnings of words, but reduce cue informativity later in words. I will argue that this typological tendency plausibly arises from the word-level tendency to preserve higher informativity cues at word beginnings.

References

- Zipf, George Kingsley .1935. "The psycho-biology of language."
Piantadosi, Steven T., Harry Tily, and Edward Gibson. 2011. "Word lengths are optimized for efficient communication." *Proceedings of the National Academy of Sciences* 108.9: 3526-3529.