Simulating the accumulation of lexical complexity

Adam King and Andrew Wedel
University of Arizona

The lexicon is a particularly interesting locus for the effects of language evolution, and has been the subject of several recent (and not so recent) investigations into its status as an optimized code for communication in a noisy channel[7;5;4]. Accordingly, word forms in the lexicon should balance between ease of production, i.e. short forms, and maximal likelihood of identification, i.e. high form-internal information, with word frequency playing a key role in a word's shape. Here, we present a tool to simulate diachronic change in word forms, showing that lexical properties, such as Zipf's law of abbreviation, emerge naturally under constraints of small generation-to-generation changes.

Furthermore, we demonstrate that simple lexicons with uniform distributions of word lengths and phoneme frequencies can evolve towards the complexity evident in natural language, without appealing to a system-wide metric of efficiency, doing so when the following observed properties of language change are included in the simulation: a) frequent words are more likely to undergo change[1;3] and phonological segments that are less informative during incremental word processing are more likely to b) be lost[2] or c) merge with higher information segment classes[6]. These results support the prediction that linguistic evolution is shaped by small locally-beneficial changes, which compound over time to ultimately form the shape of the lexicon and linguistic structure broadly.

References

[1] J. L. Bybee and P. J. Hopper. *Frequency and the emergence of linguistic structure*, volume 45. John Benjamins Publishing, 2001.

[2] U. Cohen Priva. Informativity affects consonant duration and deletion rates. *Laboratory Phonology*, 6(2):243–278, 2015.

[3] J. Kanwal, K. Smith, J. Culbertson, and S. Kirby. Zipf's law of abbreviation and the principle of least effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, 165:45–52, 2017.

[4] K. Mahowald, I. Dautriche, E. Gibson, and S. T. Piantadosi. Word forms are structured for efficient use. *Cognitive science*, 42(8):3116–3134, 2018.

[5] S. T. Piantadosi, H. Tily, and E. Gibson. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9):3526–3529, 2011.

[6] A. Wedel, A. Kaplan, and S. Jackson. High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, 128(2):179–186, 2013.

[7] G. K. Zipf. *The psycho-biology of language*. Mifflin Houghton Publishing, 1935.