Lexical dynamics, iterated sampling, and Zipf's Law

Robert Daland
UCLA

Piantadosi (2014) points out that Zipf's Law can be derived by many different approaches, none of them are adequate to the full range of facts (e.g. 'mother' is frequent in every language – not predicted by theories which exclude semantics). I simulate the 'acquisition-production' loop with iterated, agent-based models in which children learn words they are exposed to, and then repeat those words to their own children. *Contra* Piantadosi, I show that an everyday assumption (the probability of a word's occurrence is proportional to its experienced frequency) leads to decidedly non-Zipfian outcomes, because it fails to model productivity. I introduce and develop the mathematical properties of **utility-based sampling**, in which words are produced frequently not because they are already frequent, but because they are useful. One interesting result: interacting agents can create a language with a Zipfian distribution.