Stochastic Phonological Knowledge: General Constraints, Gradient Ranking

Bruce Hayes
UCLA

Recent research has documented an ability of language learners to project statistical knowledge from the lexicon. Where the language contains competing morphological patterns (for example ɪŋ ~ ʌŋ vs. X ~ Xd past tenses in English, or competing verbal conjugation classes in Romance) language learners become tacitly aware of the relative lexical frequencies of the rival patterns. Supporting evidence comes from experiment: asked to inflect novel stems compatible with more than one pattern, speakers behave stochastically, adopting one or another pattern at random in frequencies matching the lexical frequencies.

This ability has been shown for opaque alternations (Zuraw 2001), conjugation class choice (Albright 2002), and projection of underlying representations (Ernestus and Baayen 2003). The present work extends this research line to a classical case of natural, transparent phonology, the vowel harmony system of Hungarian. A formal analysis using stochastic Optimality Theory is proposed, and it is demonstrated that the rankings are learnable using a combination of algorithms from the literature.

I conclude with data suggesting, tentatively but intriguingly, that language learners favor general—that is, formally simple—constraints. When the Hungarian system is modeled using constraints that are too detailed, the result is a better match to the training set, but a worse match to native speaker intuitions. The suggested conclusion is that the speakers learn a grammar that diverges somewhat from the input data pattern, in the direction of greater generality.