Two Eras in Learning Theory: Implications for Cognitively Faithful Models of Language Acquisition and Change

Partha Niyogi (niyogi@cs.chicago.edu)
Department of Computer Science & Department of Statistics, 1100 E. 58th Street
Hyde Park, Chicago, IL 60637 USA

Robert C. Berwick (berwick@csail.mit.edu)
Department of EECS and Brain and Cognitive Sciences, MIT, 32D-728, 77 Mass Ave.
Cambridge, MA 02139 USA

Abstract
We review recent advances towards more cognitively-faithful models of language acquisition and change that parallel conceptual shifts in computational learning theory, and how these new models can yield improved empirical accounts. We have now obtained statistics from the Penn-Helsinki corpus of Old and Middle English that permits estimation of historical parameter values. Specifically, we have analyzed the competition between two grammatical systems in Middle English, (one primarily verb-final (OV-type) and the other verb-initial (VO-type)).

In this setting we assume two grammars with corresponding languages L₁ and L₂. g₁ speakers produce expressions with probability P₁ over L₁ and g₂ speakers, probability P₂ over L₂. Parameter a = P₁(L₁ ∩ L₂) and b = P₂(L₁ ∩ L₂), a and b are the probabilities with which speakers of pure g₁ and g₂ produce “ambiguous” expressions. If xᵣ is the proportion of g₁-type grammars in the rth generation, then:

\[ x_{r+1} = \frac{(1-a)x_r)}{(1-a)x_r + b(1-x_r)} \]

This has bifurcations as \( a - b \) continuously changes. We estimate \( a \) and \( b \) at a single time point, using \( a - b \) to predict which grammatical type dominates in successive generations. However, given data from a mixture distribution \( P = xP₁ + (1-x)P₂ \), can we even estimate \( a \) and \( b \)? Yes: we collect data from the Penn-Helsinki corpus by sampling a few individuals at the same time point. This is nontrivial, because only surface forms of writers’ expressions are available; one cannot always uniquely decode underlying grammars. We overcome this by “tying” parameters in a novel way. Importantly, this new estimation procedure permits empirical tests of this class of models for language change using data from historical corpora for the first time, and again validates the need for a fully population view of language acquisition, evolution, and change.

References


