Dominik Rus
Department of Linguistics and
Brain & Language Lab, Department of Neuroscience,
Georgetown University,
Washington DC

Biology Meets Statistics: Towards a Variational Model of Morphosyntactic Learnability

Central to modern linguistics is the view that human language is a biological entity whose evolution, design, and acquisition resemble the processes and characteristics found in natural objects ("biolinguistics"; Chomsky 2000; Jenkins 2004). In this tradition, (most of) language is (believed to be) selectional rather than instructional; hard–wired rather than purely emergent; and—similarly to the predominant view in basic natural sciences and neurosciences (Gazzaniga 1993)—subjected to natural selection (Duchaine, Cosmides, and Tooby 2001; Pinker 1999). Language as an empirical inquiry must thus be accounted for not only descriptively (e.g., via linguistic theories), but particularly explanatorily (i.e., via learnability theories). However, predominant deterministic theories of learnability (i.e., U(niversal) G(rammar)–based) in the Chomskyan tradition have had a hard time explaining variability within and across learners, as well as maintaining formal sufficiency (i.e., causality and concreteness) and developmental compatibility (i.e., gradualness and quantitativeness) (Yang 2002).

In this talk, I show how deterministic approaches have tackled one of the most common cross–linguistic phenomena found in early morphosyntax—Root Infinitives, where in "finite" contexts in matrix clauses, young children opt for a bare verb stem or infinitive. Central to all UG–based theories is some kind of structural deficit in the syntax, due to the initial underspecification of T(ense) and/or Agr(eement) in the phrase marker, arising from a different (immature) feature checking option (Wexler 1998) or complete omission of functional material (Radford 1995). Based on early English, Romance languages, and Slovenian, I show that kids are indeed extremely sensitive to target (T/Agr) morphosyntactic properties (Guasti 1994; Hyams 2005; Salustri and Hyams 2003; Rus 2006; Rus and Chandra 2005), though careful quantification of the data reveals important information that otherwise gets skewed in more undifferentialist accounts (Aguado-Orea and Pine to appear). UG *does* constrain acquisition, but the process is gradual and far from "perfect", contrary to the common arguments in the mainstream generative acquisition accounts (e.g., Crain and Pietroski 2006; Radford 2000; Wexler 1998), and more in line with constructionist, frequency–based accounts (e.g., Aguado-Orea and Pine to appear). However, probabilistic approaches alone, *cannot* predict which systems are impaired and why, and—to an even lesser degree—*how* exactly the systems get "perfect" (i.e., they're rather descriptive than explanatory, and generally pushing for a single-level linguistic representation).

I argue that one needs to adopt some "golden middle" approach in which biology directly interacts with experience (i.e., the input). Following Yang (2002, to appear), I formulate language acquisition as a process where UG "merely" delimits the space of possible internalized grammars which are in competition (and which in Yang's model vary along parametric specifications), interacting with stochastic learning. Such variational model of learnability conceptually satisfies our earlier theses of biolinguistic research, and at the same time empirically accounts for gradualness, errors, and inconsistencies found across child grammars. I show quantitatively how Legate and Yang's (to appear) simple, universal constraint, related to the morphological marking of time anchoring *can* predict the distribution of RIs cross-linguistically. At the same time, however, I discuss some of the shortcomings of the model, based mainly on the data from child Italian, Slovenian, and Spanish, arguing that a variational model is on the right track, but has to be "sharpened" and tuned to language–specific properties, which in turn can be captured by more universal *biases*/*constraints* that need to be first and foremost theoretically well motivated.