

TITLE: Simulating language, family, and feature evolution: a review of the state of the art

CATEGORY: Oral

THEME SESSION: 5. Quantitative Linguistic Typology: State-of-the-Art and Beyond

In recent years simulation models have become current as a complement to empirical inquiries into language dynamics (understood as a broad field encompassing not only language change—the domain of traditional historical linguistics—but also "external factors" such as interaction among languages and their speakers). Simulation models have been designed to test hypotheses concerning the development of linguistic diversity (Nettle 1999, Holman et al. 2008), language competition (e.g., Abrams and Strogatz 2003, Patriarca and Leppänen 2004, Mira and Paredes, Stauffer et al. 2006, Schulze et al. 2008, Murilo et al. 2008), the measurement of differential stabilities of abstract typological features (Wichmann and Holman 2009), the relationship between language change and population structures (Nettle 1999b, Wichmann et al. 2008), and the performance of different phylogenetic methods (Barbañçon et al. 2013). (In addition, numerous studies have applied computer simulations as an approach to language evolution, but this field is not considered here).

Among agent-based models, which are suitable for studying the interaction among languages and their speakers, some operate with a simulated space (a lattice) and no internal language structure (de Oliveira et al. 2006a,b, Patriarca & Leppänen 2004, Pinasco and Romanelli 2006), whereas others similarly study language dynamics in a simulated space but also has some simple way of representing language structure—typically as a string of binary features (bitstrings) (Schulze and Stauffer 2005, Kosmidis et al. 2005, Stauffer et al. 2006, Murilo et al. 2008, Teşileanu and Meyer-Ortmanns 2006). The latter class of models is obviously of most immediate interest from a linguistic point of view.

In the present paper we review existing literature on simulation models suitable for the investigation of topics relevant for historical-typological linguistics in a broad sense, propose evaluation criteria, and set forward a model representing a consensus of the group of co-authors. With the proliferation of work in this area it is important to distinguish between less and more adequate models. Since the simulation of the dynamics of linguistic diversity is a field of great potential we feel that it is important to assess previous results as well as to outline a generic simulation model that strikes a balance between the two opposed desiderata of maximal realism and minimal complexity. Our currently evolving general-purpose simulation model is intended primarily as a tool for evaluating specific methods for estimating diachronic typological developments, not as a tool for modeling everything we know about language history. It is simple, containing in its basic structure mainly parameters for birth, death, and change, but will allow for extensions via plug-ins (e.g. different kinds of networks for modeling causes or effects of change). Languages are modeled as vectors of properties (values of variables) as its basic units (the units whose behavior is simulated). A distinction is made between cognation variables (extremely low probability of recurrence, but defining genealogies) vs. typological variables. Finally, it is tuned to fit reality to the extent that it replicates known global patterns of languages such as family sizes, language sizes, and quantitative features of tree topologies.