

Structure and change in the space of possible grammars

Edward Stabler, UCLA, stabler@ucla.edu

0. This paper shows how many linguistic proposals can be formally modeled by a theory of language change that can then be refined using standard quantitative methods. In particular, the recognition of empirical and logical structure in the space of possible grammars allows a simple formal model that predicts ‘S-shaped’ transitions and convergence in a range of conditions, and allows the estimation of distinct probabilities of parameter changes in language phylogeny.

A formal model of language change is given by a space of grammars (e.g. specified by parameter settings or rules) together with transition probabilities among these grammars and an initial population. As usual the model should account for as much data as possible (making actual historical changes probable, and allowing plausible explanations of exceptional cases), with the fewest and best supported assumptions.

1. Occurrence frequencies and non-instantaneous changes. Although a language learner will sometimes abandon one construction in favor of another rather quickly, the transition is rarely instantaneous; for a brief period of time both constructions may be roughly equally likely. Almost step-like but really S-shaped transitions are also found in historical studies of language change, as is expected in certain conditions that can be defined population-theoretically. In some accounts it arises in a population from indeterminacy in the competition between alternative grammars, and in others from indeterminacy in the choice of parametric settings (or rules or constraints) in a ‘meta-grammar’. But these alternatives are notational variants when the probability distribution over grammars can be regarded as an equivalent probability distribution over the parametric settings of a single grammar. So for the present study we will assume an approach of the latter kind, as formalized by [9] in which ‘competing’ parameters are defined by features of functional categories. This approach is especially simple and concise, with natural performance models [3, 1], but can also be represented as grammar competition [7, 11].

2. Structure in the parameter space and language phylogeny. A simple approach to comparative, parametric analysis clusters grammars based on the assumption that all parameters are equally likely to change (and parsimony-based phylogenetic methods are based on similar uniformity assumptions: minimize changes in the tree). Recent work has shown that even with these simple models we obtain an impressive correspondence with results of the more common lexically-based and genetic methods [6]. And there is reason to suspect that structural methods like this may ultimately provide reasonable conclusions at greater time depths than lexical methods [2]. But there are logical and empirical relations among the parameter settings: when these can be determined – even approximately – they can replace the arbitrary uniformity assumptions, providing an important step towards a sophisticated structural phylogeny.

As an example of logical structure in the theory, consider the proposal about head movement in [5, 6]. In grammars respecting the ‘head movement constraint’, languages that differ in how high the noun moves in DP will be logically related: a grammar cannot get a noun to a high position without also allowing it to get to the intermediate positions. In human grammars, there are many ‘feeding’ relations of this kind: one parameter setting can enable others. These relations determine an entailment partial order on parameter settings, such that changing from one language may require several steps.

In some natural cases, the relation ‘accessible by one grammatical change’ yields a

lattice structure, useful for intuitions about the structure of the space, with the ‘least marked’ grammar at the ‘bottom’. For example, in the setting of §1, the ‘union’ of two distinct grammars $G_1 \sqcup G_2$ is a perfect bilingual, and the common components of two grammars are given by a kind of intersection $G_1 \sqcap G_2$. In a structured space of the sort defined here, the perfect bilingualism $G_1 \sqcup G_2$ will always be harder to reach than either G_1 or G_2 , suggestive of the relative instability of perfect bilingualism in many social settings. Furthermore, each grammar will be harder to achieve than the grammar that adheres to just their common components $G_1 \sqcap G_2$, in a way suggestive of the emergence of unmarked constructions in creoles. This kind of structure comes for free with the identification of entailment relations in the parametric space of variation.

In addition to logical structure, there is ‘empirical structure’ manifest in tendencies for one parameter value to co-occur with others. (Some, e.g. those underlying some Greenbergian word order parameters, could originate in biases coming from performance or interface factors.) In the present framework, these tendencies are modeled by setting the transition probabilities so that certain parameter values increase the likelihood of setting another (logically independent) parameter to a specific value.

As a consequence of this structure, certain sets of distinct parameter values will tend to co-occur. It is tempting to speculate that such clusters may define some of the ‘macro’ characters of languages that linguists have noted, predicting their relative resistance to change. The structure of the parameter space also immediately enables easier convergence among speakers in a range of conditions, demonstrated in [4]. But the most exciting prospect is that this structured language space provides an independently motivated initial hypothesis about language change that can be further refined with quantitative methods [8, 10]. The states of the population of competing grammars define a Markov model and, with reasonable idealizations, the transition probabilities among grammars determine the probabilities of population state transitions. This model determines the probability of ending up in each particular state from any initial condition, and it can be trained to better fit observed trajectories. A simple example and phylogenetic consequences are presented in the full paper.

References

- [1] BURDEN, H., AND LJUNGLÖF, P. Parsing LCFRSSs. *Proceedings IWPT 2005*.
- [2] DUNN, M., ET AL. Structural phylogenetics. *Science* 309 (2005), 2072–2075.
- [3] HARKEMA, H. A recognizer for minimalist grammars. *Proceedings IWPT 2000*.
- [4] LEE, Y., COLLIER, T. C., KOBELE, G. M., STABLER, E. P., AND TAYLOR, C. E. Grammar structure and the dynamics of language evolution. In *ECAL05* (2005), pp. 624–633.
- [5] LONGOBARDI, G. The structure of DPs. In *The Handbook of Contemporary Syntactic Theory*, M. Baltin and C. Collins, Eds. Blackwell, Oxford, 2001.
- [6] LONGOBARDI, G. Methods in parametric linguistics and cognitive history. In *Linguistic Variation Yearbook*, P. Pica, Ed. John Benjamins, Philadelphia, 2003.
- [7] NIYOGI, P. *The Computational Nature of Language Learning and Evolution*. MIT Press, Cambridge, Massachusetts, 2003. Forthcoming.
- [8] SIEPEL, A., AND HAUSSLER, D. Phylogenetic hidden Markov models. In *Statistical Methods in Molecular Evolution*, R. Nielson, Ed. Springer, NY, 2005.
- [9] STABLER, E. P. Derivational minimalism. In *Logical Aspects of Computational Linguistics*, C. Retoré, Ed. Springer-Verlag, NY, 1997, pp. 68–95.
- [10] WARNOW, T. ET AL. A stochastic model of language evolution that incorporates homoplasy and borrowing. In *Phylogenetic Methods and the Prehistory of Languages*. CUP, 2005.
- [11] YANG, C. *Knowledge and learning in natural language*. PhD thesis, MIT, 2000.