Coral Hughto, Joe Pater and Robert Staubs, University of Massachusetts Amherst

Grammatical agent-based modeling of typology

Motivation, overview. What effect does learning have on the typological predictions of a theory of grammar? One way to answer this question is to examine the output of agent-based models (ABMs), in which learning can shape the distribution over languages that result from agent interaction. Prior research on ABMs and language has tended to assume relatively simple agent-internal representations of language, with the goal of showing how linguistic structure can emerge without being postulated *a priori* (e.g. Kirby and Hurford 2002, Wedel 2007). In this paper we show that when agents operate with more articulated grammatical representations, typological skews emerge in the output of the models that are not directly encoded in the grammatical system itself. This of course has deep consequences for grammatical theory construction, which often makes fairly direct inferences from typology to properties of UG. *We argue that abstracting from learning may lead to* ***missed opportunities*** *in typological explanation, as well as to* ***faulty inferences*** *about the nature of UG.*

Model structure. We generate languages using a very simple agent network, which we take as an initial idealization; we will also present some comparisons with alternative types of network. Two agents repeatedly learn from one another, with one being randomly selected as the teacher on each learning trial. The result of multiple runs of this model yields a distribution over the languages that the grammatical model can represent. This approach to language generation could be instantiated with a range of grammar and learning theories; our work to date has focused on Maximum Entropy grammars (MaxEnt; Goldwater and Johnson 2003) learned with a sampling version of stochastic gradient ascent (Jäger 2007). As in Optimality Theory, candidates compete as the output for a given input. On each trial, an input is randomly chosen, and the teacher and learner each sample from the distribution over outputs defined by its MaxEnt grammar. If these outputs differ, the learner updates its constraint weights using the delta rule, thus moving probability onto the teacher's output. All simulations reported here started with weights at zero, and had a learning rate of 0.1.

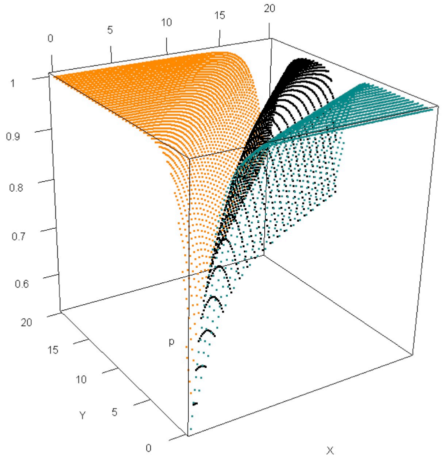
Missed opportunities? Because the output of our ABMs yields a gradient frequency distribution over languages, typological predictions are generated that go beyond the categorical possible *vs*. impossible distinction of grammatical models operating on their own. As an illustration, we take a case of typological gradience that has been claimed to follow from grammatical models having a particular structure. We show that it does, once learning is incorporated into typological explanation.

The postulation of feature geometric nodes (Clements 1985, McCarthy 1988, Sagey 1990) or of feature classes (Padgett 2002) has been claimed to account for the greater prevalence of processes that target multiple features within a class than those that target unrelated features. For example, the existence of a consonantal [place] node or class is meant to explain the relative prevalence of processes that target labials, coronals, dorsals and pharyngeals, as opposed to ones that would target unrelated features, for example labials along with consonants that are either [+voice], [–continuant] or [+sonorant]. This explanation does not go through without some auxiliary assumptions – in classic feature geometry a preference for simple linking or delinking rules was often cited, though without any specific proposal about the form or location of that bias.

Taking the case of feature classes (Padgett 2002), adding a constraint that refers to an entire class does not change the set of languages that the grammatical theory can generate. For example, adding general NoCoda[Place] and Ident[Place] to a constraint set that already has specific NoCoda and Ident constraints for each of the individual features (e.g. NoCoda[Labial] and Ident[Labial]) does not affect the ability of the theory to generate coda neutralization or non-neutralization for any subset of the features, for both place features and any others. It does, however, affect the outcome in our ABM incorporating these constraints.

We ran a simulation with 4 tableaux, representing four places of articulation in final position, with either preservation of place or delinking as candidates. Preservation violated both the specific and general NoCoda constraints and delinking violated the specific and general Ident constraints. In 40/50 runs of 10,000 trials each, neutralization occurred as the highest probability candidate in all tableaux, or none. When we repeated the simulation leaving out the general NoCoda[Place] and Ident[Place] constraints, the count for uniform application of neutralization across places of articulation was lowered to 10/50, close to the 0.125 expected by chance. This second simulation can be taken to represent the outcome for a set of unrelated features, thus showing that this grammatical ABM does capture the desired general typological skew. A notable feature of this explanation is that it does not require a stipulated learning bias – generalization comes from the fact that weight updates affect both the specific and the general constraints.

Faulty inferences? The grammatical framework we adopt has two properties that set it apart from those that are typically used in the modeling of typology in generative linguistics: it is probabilistic, and it can represent cumulative constraint interactions, or gang effects. The output of the full ABMs that we have studied to date, however, tend toward categoricity and non-cumulative interactions. Both of these can be seen in the output of a simulation that used the two tableaux below. The indicated candidates show a cumulative interaction – these candidates have highest probability in their tableaux if the weight of Con-X is greater than that of Con-Y, but still less than twice the weight of Con-Y. The zero initial weights give both candidates in each tableaux equal probability, and we stopped the simulations when the probability of one of the candidates in each tableau, averaged across the agents, was at least 0.95 (mean *n* trials = 2709). When we let simulations run longer, we find that once they reach this level of categoricity, it tends not to decrease. A tendency towards categoricity emerges in these simulations because as the probability of one of the candidates approaches 1, updates are less frequent (since the agents more often choose the same candidate), and weight changes have less of an effect (due to the use of exponentiation in the calculation of MaxEnt probability). The output of this ABM also displays a strong tendency away from the cumulative interaction shown in the tableaux. There are three pairs of candidates that can be jointly given greater probability than their competitors: (A, C), (B, D), and the one shown, (A, D). Over 1000 runs, 312 produced (A, C), 688 produced (B, D), and none produced (A, D). Randomly sampling weights from a uniform positive bounded distribution yields 0.5 (B, D), 0.25 (A, C), and 0.25 (A, D). The absence of cumulative patterns in the output of the ABM is likely in part due to the overlap in the weight space between non-cumulative and relatively categorical patterns. This can be seen in a graph that plots weights of X and Y from 0 to 20 against average candidate probability, for the three patterns. Relatively little of the high probability space is occupied by (A, D), in black.



This result, along with the others that we will present, raises the possibility that it may not be safe to infer that UG must be categorical and non-cumulative to account for observed typological tendencies in those directions.

Tableau 1	Con-X	Con-Y
→ Can-A		*
Can-B	*	

Tableau 2	Con-X	Con-Y
Can-C		**
→ Can-D	*	