Bayesian modeling of substantive biases for word order in an artificial language learning paradigm

A central hypothesis of generative linguistics is that typological universals arise because of constraints on the grammars people can learn (e.g. Chomsky 1965; Baker 2001). Recent work suggests that artificial language learning (ALL) experiments with adults can provide direct behavioral evidence for substantive biases parallel to typological tendencies (e.g. Wilson 2006; Finley & Badecker 2008). In this paper, we develop a Bayesian model which formalizes and quantifies biases hypothesized to affect the learning of word order patterns in the nominal domain. Using data from an ALL experiment in which learners exposed to mixtures of grammars tended to shift those mixtures in certain ways rather than others, we show that learners’ inferences are systematically affected by a set of prior biases. These biases are in line with a typological generalization—Greenberg’s Universal 18. This test case illustrates how learners’ internal biases impose properties on the grammars they learn, resulting in the kinds of cross-linguistic regularities known as typological universals.

The experiment

We target Greenberg’s (1963) Universal 18, which concerns ordering of adjectives and numerals with respect to nouns. In particular, of the four patterns in (1), only the first three are well-attested cross-linguistically, and the first two (‘harmonic’ patterns where nouns consistently precede or follow) are much more common than the third (proportions in parentheses are from WALS, Dryer 2008).

<table>
<thead>
<tr>
<th></th>
<th>Adj-N</th>
<th>N-Adj</th>
<th>Num-N</th>
<th>N-Num</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>70%</td>
<td>30%</td>
<td>70%</td>
<td>30%</td>
</tr>
<tr>
<td>2</td>
<td>30%</td>
<td>70%</td>
<td>30%</td>
<td>70%</td>
</tr>
<tr>
<td>3</td>
<td>30%</td>
<td>70%</td>
<td>30%</td>
<td>70%</td>
</tr>
<tr>
<td>4</td>
<td>70%</td>
<td>30%</td>
<td>30%</td>
<td>70%</td>
</tr>
</tbody>
</table>

Table 1. Experiment conditions. ‘Majority’ order is used 70% of the time.

An ALL experiment, whose results we summarize, suggests that learners exposed to a mixture of grammars—resembling a period of change—shift the mixture away from the pattern that violates Universal 18 (4), and toward one of the harmonic patterns. Figure 1a shows the extent to which learners regularize—produce the majority pattern more often than it is used in the input (see Table 1)—and Figure 1b shows individual learner outcomes.

The model

Data from this paradigm is well suited to Bayesian modeling, which assumes that learners combine experience and prior biases—probabilistic constraints on the hypothesis space—in order to make inferences, e.g. about what grammars are likely to have generated the input. We propose a Hierarchical Bayesian model of learning in the experiment described above. The parameters of the model’s prior, estimated from the experiment data, explain the results through the interaction of two prior biases: one preferring regular grammars (a formal bias), and the other favoring harmonic patterns, and disfavoring pattern 4 (the substantive bias). The model not only confirms the impact of these hypothesized biases, it reveals an intriguing pattern of behavior in each condition. Figure 2 illustrates the distribution of learning outcomes predicted by the model (after parameter fitting)—that is, how the grammars learners infer are predicted to shift compared to training as a result of the interaction between the regularization and substantive biases. The model’s parameters enforce a strong regularization bias, preferring grammars closer to the corners of Figure 2. However, the substantive bias, encoded in the model as mixture weights, (whose best-fit values are shown in the corners of Figure 2), leads learners in conditions 3 (with some probability) and 4 (always) to shift toward a harmonic pattern—better satisfying the bias. Although we develop the model in the context of Universal 18, we discuss how it can be generalized to account for biases relevant to a range of typological patterns in both syntax and phonology.
Figure 1. (a) Experiment results. (b) 2D plot of grammar-space (x-axis: probability of producing Adj-N, y-axis: Num-N). Corners labeled L1, L2, L3, L4 correspond to deterministic versions of each pattern in (1), open circles are training-input probabilities for each condition, filled-in points are individual subjects’ productions, arrows are typical shifts (purple: cond. 1, green:2, red:3, blue:4).

Figure 2. Black points labeled T1, T2, T3, T4 correspond to experiment training conditions. Opaque colored points are actual learned probabilities for learners in each condition. Transparent colored points are predicted probabilities according to the model. Arrows indicate likely shifts made by learners.

References