Complexity Bias and Substantive Bias in Phonotactic Learning

Eleanor Glewwe (University of California, Los Angeles)

A number of studies have used artificial grammar learning (AGL) to investigate the extent to which the phonological typology is shaped by synchronic learning biases (Wilson 2006, Moreton 2008, Finley 2012, White 2013, a.o.). Moreton & Pater's (2012) review concluded that there is robust evidence for complexity bias (bias toward featurally simpler patterns) but scant evidence for substantive bias (bias toward phonetically natural patterns). I present an AGL experiment that could test for multiple biases by comparing how well subjects learned different distributions of a stop voicing contrast. The results do not support substantive bias but do support complexity bias. A second experiment confirms the complexity bias effect and demonstrates the impact the broader phonological structure of an artificial language can have on performance.

In Experiment 1, there were four training languages defined on two dimensions: Trained Contrast Position (whether the language exhibited a stop voicing contrast word-initially or word-finally) and Trained Neutralization Value (whether stops in the other word-edge position "neutralized" to voiced (D) or voiceless (T)). Table 1 shows what types of stops occurred in which positions in each condition. For instance, in the Initial-T condition, both voiceless and voiced stops occurred word-initially, but only voiceless stops occurred word-finally. Items were of the form CVCVC (e.g. *lanit*), and the other two Cs in each item were sonorants, so subjects in all conditions heard word-initial and word-final sonorants.

Table 1: Exp. 1 Conditions

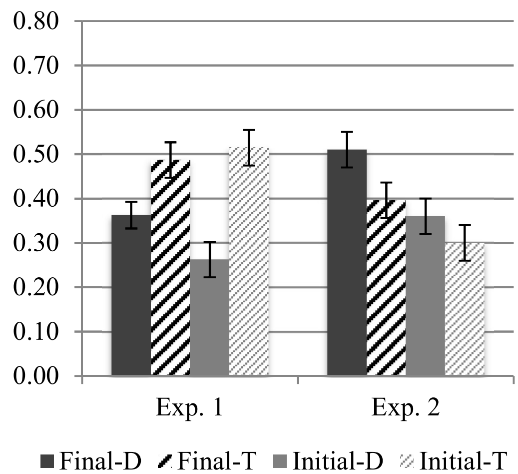
Condition	#T	#D	T#	D#
Final-D	✗	✓	✓	✓
Final-T	✓	✗	✓	✓
Initial-D	✓	✓	✗	✓
Initial-T	✓	✓	✓	✗

In the training phase, subjects were told they would be learning words of a new language. Each word was paired with an image. In the test phase, subjects heard additional words, without images, and had to say whether each word could be a word of their language or not. The test items were the same in all four conditions and consisted of #T, #D, T#, and D# items. Test items fell into three categories: familiar conforming (items heard in training), novel conforming (new items consistent with the pattern heard in training), and novel nonconforming (new items of the type not heard in training; e.g. D# items in the Initial-T condition).

Subjects' acceptance rates of novel nonconforming items indicate how well they have learned to reject words not in their language, or, conversely, how much they have extended the stop voicing contrast to the other word-edge position. In the typology, if a language contrasts voicing in obstruents word-finally, it contrasts voicing in obstruents word-initially, but not necessarily vice versa (Steriade 1997). This implicational is phonetically motivated: the voicing contrast is harder to perceive word-finally than word-initially, so if it exists word-finally, it should also exist word-initially. Thus one prediction from substantive bias is an effect of Trained Contrast Position such that subjects in Final conditions accept nonconforming items more (i.e. extend the contrast to the opposite position more) than subjects in corresponding Initial conditions. A second prediction from substantive bias concerns the effect of Trained Neutralization Value. Because voiced stops are more marked than voiceless stops, subjects in neutralizing-to-voiced (D) conditions should accept nonconforming (i.e. T) items more than subjects in neutralizing-to-voiceless (T) conditions accept nonconforming (i.e. D) items. Finally, there is a prediction from complexity bias that follows from the presence of sonorant Cs in the training items. In neutralizing-to-voiced (D) conditions, subjects can posit a *[-voice] constraint to exclude the type of item they were not exposed to (e.g. *lani**t***), but in neutralizing-to-voiceless conditions, subjects must posit the more complex *[-son, +voice] to exclude nonconforming D items (e.g. *lani**d***) while not also excluding items with (voiced) sonorants in the same position (e.g. *tiri**l***). This predicts that subjects in T conditions will (erroneously) accept nonconforming items more than subjects in D conditions, the opposite of the second substantive bias prediction.

The left side of Figure 1 shows the mean acceptance rates of novel nonconforming items by condition in Experiment 1. (Acceptance rates of novel conforming items were above chance in all four conditions and did not differ significantly between conditions.) I fit a mixed-effects logistic regression to the novel nonconforming items with response (accept or reject) as the dependent variable and Trained Contrast Position, Trained Neutralization Value, and their interaction as fixed effects. The only significant effect was a main effect of Trained Neutralization Value: subjects in neutralizing-to-voiceless (T) conditions were more likely to accept nonconforming items than subjects in neutralizing-to-voiced (D) conditions. This result supports complexity bias and not a voicing-based substantive bias. Although the difference in acceptance rate of novel nonconforming items between the Final-D and Initial-D conditions was in the direction that would support a positional substantive bias, the difference in acceptance rate between Final-T and Initial-T was in the opposite direction. Thus subjects did not consistently extend the voicing contrast more from final to initial position than from initial to final position.

Fig. 1: Acceptance Rates of Novel Nonconforming Items



To test the effect of complexity bias further, I conducted a second experiment identical to Experiment 1 in all respects but one: the non-critical (i.e. non-stop) Cs were all voiceless fricatives instead of sonorants. This change flips the complexity bias prediction. Now subjects in T conditions can posit a *[+voice] constraint to exclude the type of item they were not exposed to (e.g. *fisi**b***) while subjects in D conditions must posit the more complex *[-cont, -voice] to exclude nonconforming T items (e.g. *fisi**p***) while not also excluding items with licit voiceless fricatives (e.g. *tusi**f***). Thus subjects in D conditions should accept nonconforming items more than subjects in T conditions. The right side of Figure 1 shows the mean acceptance rates of novel nonconforming items in Experiment 2. Acceptance rates of nonconforming items were in fact higher in D conditions than in corresponding T conditions. A mixed-effects logistic regression fit to the novel nonconforming items yielded a marginally significant main effect of Trained Neutralization Value in the direction supporting complexity bias. There was also a significant main effect of Trained Contrast Position: unlike in Experiment 1, subjects in Experiment 2 did consistently extend the voicing contrast more from final position to initial position than vice versa, supporting the positional substantive bias.

The two experiments yield mixed support for substantive bias but stronger support for complexity bias. They also show that an artificial language's non-critical sounds crucially affect performance. Subjects seem to infer phonotactic constraints according to the experiment-internal distribution of sounds, opting for the simplest constraint with which they can master the pattern.