On the Analytic Bias toward Phonetically Driven Tonal Phonotactics

Tsung-Ying Chen Dept. Foreign Languages and Literature Nat'l Tsing Hua Univ.

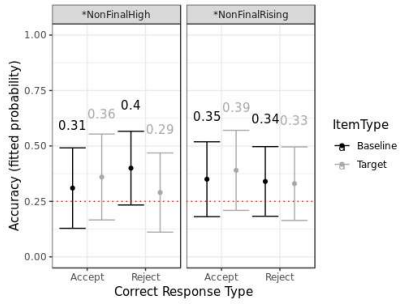
Background & Summary The current artificial grammar (AG) learning research presents evidence in favor of a phonological analytic bias (e.g., Moreton, 2008; Pater & Moreton, 2012a,b) toward learning the phonetically driven and widely attested tonal phonotactics constraint *NonFinalR (i.e., no non-word-final rising tones; Zhang, 2007, et seq.) and against learning an arbitrary and perhaps unattested tonal phonotactics constraint *NonFinalH (i.e., no non-word-final high tones). Typological surveys (e.g., Zhang, 2002) reveal a cross-linguistic trend to avoid rising tones either as a toneme or in a non-final position. The latter tonal phonotactics, referred to as *NonFinalR, in particular, may have arisen from phonetically shortened non-final syllables that make the full realization of a rising f0 contour more effortful. Tone sandhi rules reducing a non-final rising tone to a level tone were found to be more productive than arbitrary ones (Zhang & Lai, 2010; Zhang et al., 2011), thus suggesting a potential analytic bias toward *NonFinalR. In our AG learning experiment, we tested this analytic bias hypothesis by training Mandarin Chinese (MC) native speakers to learn either *NonFinalR or *NonFinalH. Since neither constraint needs to be generalized (or highly ranked in OT's term) by L1 MC speakers, the null hypothesis is that L1 MC speakers could learn both constraints equally well. However, our AG experimental results indicate a significantly better learning performance by the *NonFinalR group, therefore supporting the analytic bias hypothesis.

Method Our AG learning experiment was composed of a training session and two test sessions. The training session included 320 disyllabic words randomly created with segments, tones, and syllable templates in MC. The two training conditions were created by alternating the word-initial tone of the training stimuli: In the *NonFinalR condition, word-initial rising tones were replaced with high level tones, whereas in the *NonFinalH conditions, word-initial high level tones were replaced with rising tones. These disyllabic stimuli were recorded by the author and to avoid inter-stimuli variation, the original f0 contours were replaced with the f0 contours averaged by tone type and context. MC retroflex consonants were also excluded from the training stimuli in order to test the learning of *Retroflex as a baseline. This constraint is considered natural as many languages lack retroflex consonant phonemes (e.g., Hawaiian), which is thus expected to be learnable in an AG learning context. If a tonal constraint is natural, it should be as learnable as *Retroflex. Test sessions included 128 novel disyllabic stimuli created following the same procedure used to create the training stimuli. A subset of 64 items served as the target stimuli with half violating *NonFinalR and the other half violating *NonFinalH. The other 64 baseline items conformed to both tonal constraints, but half of them had an MC retroflex consonant as the onset of their second syllable, therefore violating *Retroflex. All participants first took part in the training session, In the training phase, in which participants were asked to learn a 'minor Chinese dialect' and exposed to the training stimuli in random order via headphones. Every 4 to 7 trials, participants were asked to recall a previously presented training stimulus; a correct recall rate of 90% was set as the threshold for continuing to the test sessions. Qualified participants then proceeded to complete an immediate test session as well as a delayed test session. The two test sessions were essential to examine the consistency in response patterns; if a constraint is more learnable, participants would respond to relevant test items more accurately and more consistently. The test sessions required participants to listen to the 128 novel test stimuli presented in random order and judge whether they sounded like the 'minor Chinese dialect' with a binary response ('like' vs. 'unlike'). The

first test session was administered immediately after the training session, and the other was administered on a separate day to test the consistency in accurate judgments across the test sessions. Both training and test sessions were administered using PsychoPy v1.85.4 (Pierce, 2009) in a quiet room. Fifty-three MC L1 speakers enrolled as college students were recruited and randomly assigned to either training condition.

Result Analysis Due to limited space, here we only discuss the consistency in correct responses across the two test sessions, which is most informative for understanding our learners' performance. This consistency data set were analyzed using mixed-effects logistic regression, and in which binary-coded consistency (Correct-Correct vs. Others) served as the dependent variable. Main predictors directly derived from our experimental design were Group (*NonfinalR vs. *NonfinalH), ItemType (Target vs. Filler), and CorrectType (Accept vs. Reject) with their three-way interaction included. The random intercepts of Participant and Test Item and the by-participant random slopes for ItemType and CorrectType were included in the model. The model is summarized in Figure 1 based on the three-way main interaction in this model, with the Group × ItemType interaction being significant. This is obvious by specifically looking at how learners in the two training condition rejected test items violating the target tonal constraint or *Retroflex. The *NonFinalR group was consistent in correctly rejecting test items violating *NonFinalR or *Retroflex with a performance significantly above the chance level (p = .25; the dotted line). The *NonFinalH group, while correctly rejecting the same test items violating *Retroflex, failed to reject those violating *NonFinalH. In addition, the *NonFinalR group did equally well in accepting or rejecting both types of test stimuli, but the *NonFinalH group showed a substantial variation in their judgment performance across item and response types.

Figure 1. Accuracy Consistency across Session



Discussion The above analysis is the evidence for the analytic bias toward *NonFinalR, consistent with previous experimental findings. One intriguing pattern found in our results is that the *NonFinalH group seemed to converge on the *Retroflex grammar faster, considering their better performance in consistently rejecting test items violating *Retroflex. It could be the case that since *NonFinalH is absent from the innate language faculty and cannot be generalized via implicit learning, the *NonFinalH AG is in fact a less complex grammar allowing for a single focus on the learning of *Retroflex. We will also report our analysis of accuracy rates in the two test sessions that also indicates an advantage for the *NonFinalR group. Futhermore, we are collecting more evidence with an experiment for participants to learn directly from feedback on their judgments in a test session and will report the variation in accuracy rate over time to discuss the learnability of *NonFinalR and *NonFinalH. If *NonFinalR is more learnable, its learners should converge on *NonFinalR *faster*.

References Moreton (2008) *Phonology* 25, 83-127. Myers (2015) *Language & Linguistics* 16, 791-818. Optiz & Freiderici (2013) *NeuroImage* 19, 1730-1737. Pater & Moreton (2012a) *Language & Linguistics Compass* 6, 686-701. Pater & Moreton (2012b) *Language & Linguistics Compass* 6, 702-818. Pierce (2009) *Frontiers in Neuroinformatics* 2, 10. Zhang (2002) *Doctoral Dissertation, UCLA*. Zhang (2007) *J East Asian Linguistics* 16, 259-302. Zhang & Lai (2010) *Phonology* 27, 153-201. Zhang et al. (2011) *Lingua* 121, 181-206.