Learning Non-Local Phonological Alternations via Automatic Creation of Tiers

Consonant and vowel harmony (e.g. Turkish [d**ɑ**l-l**ɑ**r] 'branch-PL' vs. [j**e**r-l**e**r] 'place-PL') frequently involve dependencies between segments that are not adjacent in a word. Nevertheless, linguists often analyze such dependencies as being local on a relevant tier (e.g. the vowel tier <**ɑɑ**>, <**ee**>) [1]. Tiers simplify non-local learning and are consistent with human behavior in artificial language experiments [2,3], but little is known about the mechanism by which the mind may construct such representations. We propose a computational model that uses a sensitivity to adjacent segments to automatically construct a tier that renders the relevant dependencies as local, falling within the model's purview.

Proposal: Motivated by humans' strong proclivity for tracking adjacent dependencies [4], we propose a computational model that tracks only dependencies between alternating segments and the segments adjacent to them. The model is provided with a vocabulary in which some segments alternate, and needs to construct a generalization that can predict the surface form of the alternating segments, which are underlyingly abstract, from the adjacent dependencies that it tracks. If adjacent segments fail to account for the alternation, the model deletes those failed adjacent segments and repeats. These un-predictive segments accumulate into a *deletion set,* the complement of which constitutes a *tier.* Figure 1 provides a toy example of learning sibilant harmony, where underspecified /-S/ takes its [±ant] value from the preceding sibilant. Adjacent vowels, also unspecified for [ant], cannot donate their feature and are thus deleted (iteration 1). The non-sibilant consonants, which give the wrong surface form for /S/, are deleted next (iteration 2), yielding the [+sib] tier.

Results: In a prior artificial language experiment [2], participants were presented with data consisting of <stem, suffixed> pairs fitting the template CV**S**V-**S**V, where a suffixal sibilant harmonized with a stem sibilant across an intervening vowel. At test time, these participants demonstrated that they had learned a harmony pattern, expecting the sibilants in novel CV**S**V-**S**V words to harmonize. However, when presented with words of the form **S**VCV-**S**V, the participants did not systematically expect the sibilants to harmonize. When trained on the same CV**S**V-**S**V words, our proposed model forms a deletion set containing the vowels (equivalently, it learns a [+cons] tier). Consequently, our model mirrors human behavior: harmony generalizes to novel CV**S**V-**S**V words but is blocked by the intervening C in **S**VCV-**S**V words (Tab 1). Asymmetrically, when human participants were exposed to words of the form **S**VCV-**S**V during training, they expected the sibilants in both novel **S**VCV-**S**V and CV**S**V-**S**V words to harmonize. When trained on the same data, our model initially forms a deletion set containing the vowels, then adds the non-sibilant consonants, yielding the [+sib] tier. Consequently, our model once again mirrors human behavior (Tab 2).

Moreover, when trained on natural language data, our model effectively learns harmony generalizations that extend with high accuracy to held-out test words. We trained our model on 958 Turkish words and 975 Finnish words. When trained on Turkish, our model learns a [+vowel] tier and achieves 98.1% accuracy generalizing to 240 test words. Errors are due to harmony exceptions. These results are consistent with standard characterizations of Turkish vowel harmony [5], and acquisition studies, which reveal that Turkish-speaking children as young as 2;0—when their vocabulary likely contains under a thousand words—already know vowel harmony well enough to extend it to nonce words [6]. Similarly, when trained on Finnish, our model learns to exclude the neutral vowels {i, e} from the tier, and achieves 98.19% accuracy on 244 test words. Our model outperforms alternative models [7, 8, 9] (Tab 3).

Conclusion: Our proposed model reflects humans' propensity to track adjacent dependencies [4], mirrors human behavior in artificial language experiments [2], and learns natural-language alternations at a rate comparable to children [6]. Thus, the model gives a possible mechanistic account of the procedure by which humans learn non-local alternations.

References [1] Goldsmith 1976. *Autosegmental phonology*. MIT. **[2]** Finley 2011. *The privileged status of locality in consonant harmony*. J Mem Lang. **[3]** McMullin & Hansson 2019. *Inductive learning of locality relations in segmental phonology.* Lab Phon. **[4]** Gómez & Maye 2005. *The developmental trajectory of nonadjacent dependency learning*. Infancy. **[5]** Kabak 2011. *Turkish vowel harmony*. The blackwell companion to phonology. **[6]** Altan 2009. *Acquisition of vowel harmony in Turkish*. 35. yıl Yazıları. **[7]** Goldsmith & Riggle 2012. *Information theoretic approaches to phonological structure: the case of finnish vowel harmony.* Nat Lang Ling Theory. **[8]** Gouskova & Gallagher 2020. *Inducing nonlocal constraints from baseline phonotactics*. Nat Lang Ling Theory. **[9]** Jardine & McMullin 2017. *Efficient learning of tier-based strictly k-local languages*. LATA.

Figure 1: A toy example of the model learning sibilant harmony.

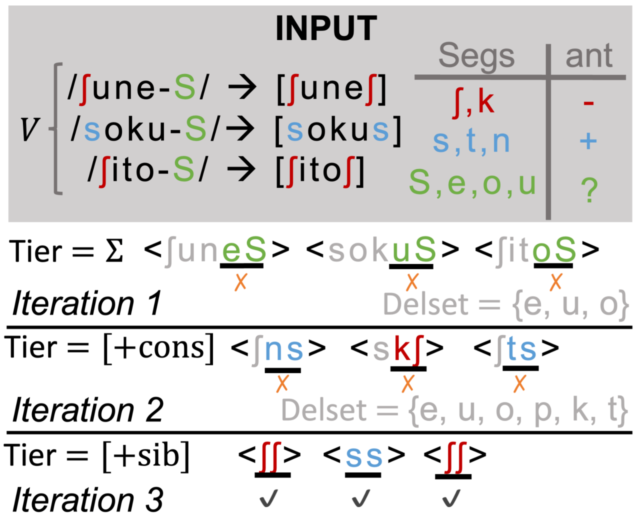

Table 1: Our model mirrors human behavior in [3]'s first artificial language experiment, generalizing (≈1.0) when humans do (✔) and not generalizing (≈0.5) when humans do not (✗).

	Train CVSV-SV		
	CVSV-SV (Old)	CVSV-SV (New Train-Like)	SVCV-SV (Novel)
Humans	✓	✓	✗
Our Model	1.0000 ± 0.00	1.0000 ± 0.00	0.5000 ± 0.14
GR [7]	0.5778 ± 0.32	0.7121 ± 0.22	0.5722 ± 0.11
GG [8]	0.4722 ± 0.12	0.5182 ± 0.17	0.5028 ± 0.12
FS [9]	0.5083 ± 0.13	0.4909 ± 0.13	0.5556 ± 0.13
Trigram	1.0000 ± 0.00	1.0000 ± 0.00	0.7028 ± 0.04

Table 2: Our model mirrors human behavior in [3]'s second experiment, generalizing (≈1.0) when humans do (✔).

	Train SVCV-SV		
	SVCV-SV (Old)	SVCV-SV (New Train-Like)	CVSV-SV (Novel)
Human	✓	✓	✓
Our Model	1.0000 ± 0.00	1.0000 ± 0.00	1.0000 ± 0.00
GR [7]	1.0000 ± 0.00	0.4167 ± 0.00	0.2500 ± 0.00
GG [8]	0.4500 ± 0.14	0.5111 ± 0.15	0.5250 ± 0.11
FS [9]	0.5111 ± 0.16	0.5250 ± 0.15	0.5556 ± 0.13
Trigram	1.0000 ± 0.00	0.5833 ± 0.00	0.2500 ± 0.00

Table 3: Our model learns accurate harmony generalizations.

Model	Test Accuracy (Turkish)	Test Accuracy (Finnish)
Our Model	**0.9810** ± **0.00**	**0.9819** ± **0.00**
GR [7]	0.6145 ± 0.27	0.8001 ± 0.03
GG [8]	0.4224 ± 0.01	0.5644 ± 0.03
FS [9]	0.4112 ± 0.01	0.5766 ± 0.01
Trigram	0.4890 ± 0.01	0.8347 ± 0.02