A Learning Based Account of Turkish Laryngeal Alternations

Caleb Belth, University of Utah Linguistics Department

A traditional generative conception of the relationship between phonological theory and acquisition is that theory delineates a space of possible grammars and that acquisition is induction from overt linguistic data to the correct grammar in that space [1]. Another conception is that acquisition is abduction of a sufficiently productive linguistic system in response to the sparsity of linguistic input [2, 3]. Under this conception, the structure of particular speakers' psychological grammars is the consequence of the interplay between the learning algorithm and the distributional properties of the input, which makes a theory of acquisition a (psychological) theory of grammar. Prior corpus research [4] found that in a large Turkish corpus [5], laryngeal alternations—in which some (1a) but not all (1b) noun-stem-final stops alternate between voiceless and voiced (or ∅ in the case of [k]) based on whether a vowel-initial suffix (e.g. the possessive in 1) is attached—are statistically predictable by a number of features of the stem.

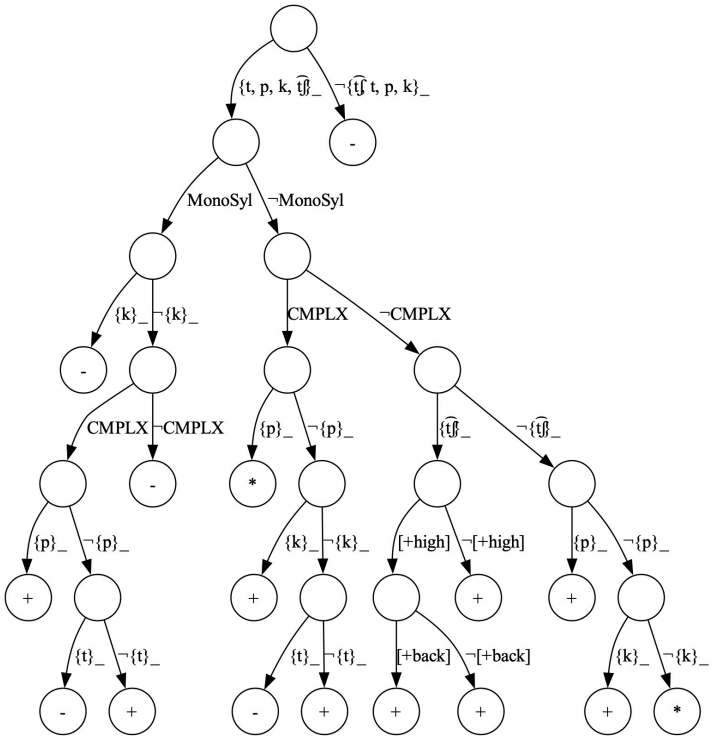
(1) (a) [sɑhi**p**] ~ [sɑhi**b**-im] 'owner'-POSS (b) [to**p**] ~ [to**p**-um] 'ball'-POSS
[ɑde**t**] ~ [ɑde**d**-im] 'amount'-POSS [hizme**t**] ~ [hizme**t**-im] 'employment'-POSS
[gen**t͡ʃ**] ~ [gen**d͡ʒ**-im] 'youth'-POSS [hi**t͡ʃ**] ~ [hi**t͡ʃ**-im] 'worthless one'-POSS
[byjy**k**] ~ [byjym] 'big one'-POSS [il**k**] ~ [il**k**-im] 'first one'-POSS

In addition to well-known predictive features—namely the prosodic size (polysyllabic nouns are more likely to alternate than monosyllabic) and place of articulation (coronal [t] is less likely to alternate than non-coronals [p, k, t͡ʃ])—[4] also found that whether a monosyllabic stem has a complex coda and the vowel quality (backness and height) of the final vowel were statistically predictive. However, in a wug test [4], adult Turkish speakers productively generalized the alternation based only on the size of the word and place of articulation; they showed no evidence of a sensitivity to the vowel quality dependency. Because vowel backness or height influencing following consonant laryngeal features has not been reported typologically, the authors propose that UG specifies possible and impossible interactions, with this interaction being ruled out.

Proposal. We combine the learning algorithms ATP [1] and PLP [2] to propose a learning model for this laryngeal alternation. ATP constructs morphophonological rules by recursively subdividing training words based on their features. The recursive splits yield a decision tree, with a path down the tree being a rule; recursion stops when such a rule becomes productive, as measured by the Tolerance Principle (TP) [5]. Thus, each rule is a leaf in the tree. PLP learns phonological alternations by beginning rule construction around an alternating segment (here the stem-final stop), and expanding its window of attention linearly outward in response to the alternation not being sufficiently predictable, also measured by the TP. We combine these models by using PLP's attention window to extract features for ATP. Thus, ATP has access to features indicating the prosodic size of the word (MonoSyl and CMPLX), which we assume learners have independently learned (the subjects in [4] were adults), and to phonological features from PLP, but in the order of PLP's window increase: PLP starts tracking the alternating [p, t, t͡ʃ, k]. Thus, if the alternating nouns can be sufficiently predicted with only the prosodic features and the final [p, t, t͡ʃ, k] (equivalent to the place of articulation), then PLP never has a reason to look further. In this case, the model would not generalize based on the final vowels because PLP never had reason to begin tracking those dependencies. Critically though, the model is capable of representing such dependencies should they be necessary; UG need not rule them impossible.

Results. We augmented TELL [5] nouns with token frequencies from a web crawl [7] and simulated Turkish learners by sampling nouns weighted by frequency. We performed 24 samples (the number of participants in [4])—choosing each sample size from a normal distribution

(μ=10K σ=1K)—and ran the model on each. We then had each model predict, for each of [4]'s nonce words, whether or not it alternates. We compared the fraction of models that predicted each nonce to alternate to the fraction of [4]'s participants who did so. The correlation (Pearson's *r*) was 0.83 ($p < 0.001$). We also fit two logistic regression models to the TELL corpus, corresponding to [4]'s models. The first includes vowel features as predictors: *alter ~ place*size + place*back + place*high*. The second includes only place and size as predictors: *alter ~ place*size*. We used these regression models to estimate the probability of each nonce alternating and compared this to the human alternation rates. They achieved 0.80 and 0.82 (both $p < 0.001$) correlation respectively. Since the model without vowel predictors correlated better, this is consistent with [4]'s finding that Turkish speakers did not show a statistically robust sensitivity to the vowel dependencies. However, our model, which allows the possibility of vowels being tracked, achieves comparable correlation with the regression model that, by specification, rules out the possibility of vowel quality influencing the alternation. An example tree is shown below (+/− = +/−alternate; * = not productive). Interestingly, the model does sometimes condition rules on vowel quality, but the resulting rules are of very narrow scope (covering <1% of training nouns) and thus do not contribute substantially to prediction. Indeed, the model's correlation was marginally higher than the regression model that excluded vowels, and is thus consistent with [4]'s wug test. Thus, we conclude that our model provides a possible learning based account for why Turkish speakers may not be sensitive to the vowel-consonant interaction: it could be that the lack of sensitivity is a consequence of the algorithms by which generalizations are abductively constructed, rather than a reflection of an analytical constraint in UG.



References

[1] Tesar, B., & Smolensky, P. (1998). Learnability in optimality theory. *Linguistic Inquiry*.

[2] Belth, C., Payne, S., Beser, D., Kodner, J., & Yang, C. (2021). The greedy and recursive search for morphological productivity. *CogSci*.

[3] Belth, C. (In Press). A Learning Based Account of Local Phonological Processes. *Phonology.*

[4] Becker, M., Ketrez, N., & Nevins, A. (2011). The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language*.

[5] Inkelas, S., Küntay, A., Orgun, O., & Sprouse, R. (2000). Turkish electronic living lexicon (TELL). *Turkic Languages*.

[6] Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. MIT press.

[7] Wenzek, G., Lachaux, M. A., Conneau, A., Chaudhary, V., Guzmán, F., Joulin, A., & Grave, E. (2019). CCNet: Extracting high quality monolingual datasets from web crawl data. *LREC.*