A Learning Based Account of Turkish Laryngeal Alternations

Caleb Belth
University of Utah

1 Introduction

Within generative linguistics, phonological theory has often been viewed as delineating a space of possible grammars. Language acquisition can then be viewed as involving induction from overt linguistic data to the target grammar in that pre-specified space (Tesar & Smolensky, 1998). From this perspective, phonologists have taken interest in cases where learners fail to form a generalization that appears tenable given analysis of language data, because such discrepancies between the regularities present in the input and the generalizations actually formed may indicate something about possible and impossible generalizations (Becker et al., 2011, 2012). A prominent example comes from Becker et al. (2011)'s analysis of Turkish laryngeal alternations. Turkish, like many languages, exhibits a pattern of voicing alternations in which some (1a) but not all (1b) noun-stem-final stops and affricates 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. The form of alternating stem-final consonants is typically taken to be voiced underlyingly (or /k/ in the case of the [k] ∼ ∅ alternation), but to be devoiced in final position (the vowel-initial suffix leading the stop to occur in a syllable onset). Unlike many languages described as having final devoicing, fricatives in Turkish are not devoiced (Kopkallı, 1993).

(1)	a.	[sɑhip]	∼	[sɑhib-im]	'owner'-Poss
		[ɑdet]	∼	[ɑded-im]	'amount'-Poss
		[gent͡ʃ]	∼	[gend͡ʒ-im]	'youth' -Poss
		[byjyk]	∼	[byjym]	'big one'-Poss
	b.	[top]	∼	[top-um]	'ball'-Poss
		[hizmet]	∼	[hizmet-im]	'employment'-Poss
		[hit͡ʃ]	∼	[hit͡ʃ-im]	'worthless one'-Poss
		[ilk]	∼	[ilk-im]	'first one'-Poss

Since not all stop-final stems alternate (1b), the alternation is a lexically-specific property of particular words. Nevertheless, certain features of noun stems are predictive of whether a noun is likely to be an alternating noun or not (as is common in voicing alternations—see for instance Ernestus & Baayen 2003). In particular, Becker et al. (2011) found that in the large Turkish TELL corpus (Inkelas et al., 2000), laryngeal alternations are statistically predictable by a number of features of the stem. In addition to well-known predictive features—namely the prosodic shape (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])—Becker et al. 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. Features predictive in this way, which are usually identified via regression analysis, have been argued to be internalized by learners of other alternations, manifesting in their responses to nonce words (Ernestus & Baayen, 2003; Hayes & Londe, 2006).

* **I thank Charles Yang, Jordan Kodner, and attendees at AMP 2024 for helpful comments.**

Indeed, in a Wug (Berko, 1958) test with adult Turkish speakers, Becker et al. (2011) found evidence of a sensitivity to the prosodic features and the place of articulation of the final stop. However, they found no evidence of a sensitivity to the relationship between vowel quality and the alternation. Becker et al. took this, together with the observation that interactions between vowel quality and consonant laryngeal features are rare cross-linguistically, as evidence that Universal Grammar (UG) might analytically rule out (or bias against) certain kinds of phonological interactions, such as this one.

In this paper, I argue, from a learning-based perspective, for a possible alternative explanation. I argue that the same result follows from two simple aspects of existing morphophonological learning proposals (Yang, 2016; Belth et al., 2021; Belth, 2023a): that learners construct generalizations that are *good enough* (rather than optimal), in a sense to be made precise via the Tolerance/Sufficiency Principle (Yang, 2016), and, during this generalization-constructing process, learners only look for new structures to put into generalizations when the structures they already know do not yield good enough generalizations. In short, I will show that there is independent reason for learners to consider prosodic shape features and features of the stem-final segment before attending to the preceding vowel, and that learners can form an adequate generalization for the voicing alternation with little-to-no reference to the vowel features. Consequently, it is unnecessary to posit that UG analytically filters out this generalization. When implemented as a computational learning model (§ 2), the proposal quantitatively accounts for Becker et al. (2011)'s Turkish speakers' performance as well or better than a model that analytically prohibits consideration of the preceding vowel quality (§ 3). I will discuss some potential implications for how we conceive of the relationship between phonological theory and learning (§ 4).

2 Proposal

Because of the sparsity of structures in a learner's input, generalization is necessary for successful language acquisition. While sparsity is a fundamental characteristic of the learning input in all languages, the problem is perhaps particularly pernicious in a language like Turkish, which uses much agglutination. Thus, when learning Turkish morphophonology, learners must construct morphological generalizations that allow them to generalize from the set of words they know to the long, Zipfian tail of the language (Chan, 2008; Yang, 2016; Belth, 2023c). However, if morphological generalizations do not take into account alternations, then alternating forms look to the learner like exceptions. For instance, the Poss suffix in (1) alternates, as do the alternating stems (1a), potentially foiling any morphological generalization that simply attempts to concatenate a suffix to a stem. The suffix alternation is due to vowel harmony, which is acquired early—often by age 2—as evidenced by overgeneralizations (Ekmekci, 1979; Altan, 2009). I have given a learning-based account of this aspect of the acquisition of Turkish vowel harmony in Belth (2023b, 2024b). In comparison, Turkish-learning children appear to take longer to acquire voicing alternations, with errors persisting for all stops through at least age 8. The errors largely include producing a voiceless stop in a vowel-initial-suffix context where the voiced counterpart would be expected (Nakipoglu et al., 2016), such as *[sɑhipim] for [sɑhibim] 'owner'-Poss. Consequently, the following proposal will proceed under the assumption that learners have acquired the vowel harmony alternation by the time they are figuring out the voicing alternation.

This means that, for instance, when learning how the Poss is formed, nouns exhibiting the voicing alternation (1a) will looks like exceptions to the generalization of attaching /-Hm/ (where /H/ is a high vowel underspecified for backness and rounding) and applying vowel harmony. If the exceptions introduced by alternation rise to the level that it is untenable to lexicalize them, then forming a generalization requires separating the alternating (+) from non-alternation (−) nouns. To formalize this proposal, I will now present a model, based closely on ATP, as proposed by me and my colleagues (Belth et al., 2021) in the context of morphological inflection and extended to handle alternations (Belth, 2024a). ATP takes as input (*source, feature, target*) triples. In the present case, the *source* of a triple is the phonological form of a Turkish noun stem, the *target* is '+' if the final segment alternates (in this case in the Poss form) and '−' if it does not. The features are properties of the stem, specifically aspects of prosodic shape and segmental content of the stem that may be useful in partitioning the nouns into those that alternate and those that do not; I will describe these features in more detail in § 2.2. ATP recursively subdivides the input triples one feature at a time (I will describe the feature-selection process in § 2.3), motivated by the experimental observation that humans sort items into categories by attending to one dimension of the items at a time (Smith, 1979; Nelson, 1984; Medin et al., 1987). This recursive, binary subdivision yields what can be interpreted as a *decision tree*, which partitions the input triples into mutually-exclusive sets, represented by the tree's leaves. Recursion stops (the

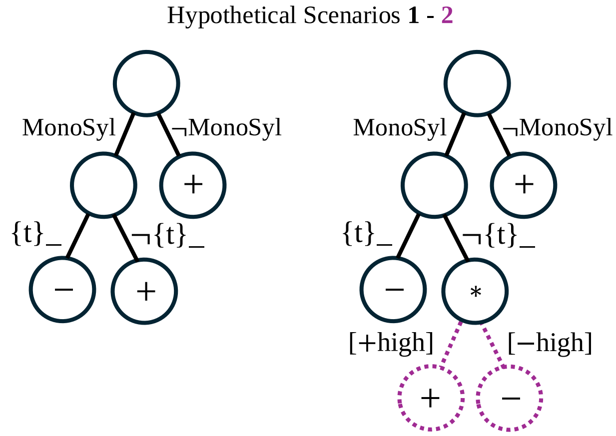

Figure 1: Two hypothetical states of the proposed learner when learning over different inputs. The pink, dotted scenario demonstrates that the model is capable of conditioning its predictions about whether nouns are alternating (+) or non-alternating (−) based on the quality (e.g., height) of the preceding vowel. However, the black, solid scenario demonstrates that if the features already available to it (e.g., monosyllabicity and the identity of the final stop) are sufficiently predictive, the learner, following Belth (2023a, 2024b), will not move deeper into the segmental representation of the noun to find the vowel dependency.

base case) at a particular level if, in that partition of the triples, enough either do or do not alternate to satisfy the Tolerance/Sufficiency Principle (TSP) (Yang, 2016).[1] I will introduce the TSP in § 2.1. Thus, any path from the tree's root to a leaf constitutes the structural condition of a rule; the rule predicts that the nouns meeting that structural condition either do (+) or do not (−) alternate.

I have visualized the basic workings of the proposal in Figure 1. The left tree shows a hypothetical state of the learner where the input has been recursively subdivided into three sets (tree leaves) based on the monosyllabicity and final segment of the noun stem. This represents a learning scenario where the learner's generalizations make no reference to vowel quality. In contrast, the right tree requires further subdivision to find productive rules, eventually (in pink) subdividing on features of the preceding vowel. As we will see below, the order in which segmental features enter into the tree follows from independent considerations of how humans attend to items in phonological representations (Belth, 2023a, 2024b).

2.1 *When A Rule Does or Does Not Work: The Tolerance/Sufficiency Principle* A central aspect of the present proposal is that learners incrementally construct and refine generalizations only as needed to sufficiently account for the input. To make this notion concrete, following Belth et al. (2021), I will use the Tolerance/Sufficiency Principle (TSP) (Yang, 2016). The TSP evaluates a generalization/rule, R, based on the number of items that match its structural description (that is, its scope n) and either the number of those that follow the rule (m) or the number that violate it ($e = n - m$). The TSP is a proposed psychological tipping point for how many exceptions a rule can *tolerate*, or how many rule-followers is *sufficient*, for a learner to internalize the rule and apply it productively. The threshold it provides is shown in (2).

(2) **Tolerance/Sufficiency Principle**: a rule R is productive *iff*

$$e = n - m \leq \frac{n}{\ln n}$$

The threshold was derived (see Yang 2016:ch.3 for a description of the derivation) by considering the Zipfian character of linguistic data and its psychological impacts on processing. Experimental evaluation of

[1] Recursion also stops if no more features are available to split on.

the TSP provides support for it as a categorical tipping point of generalization for human learners (Schuler et al., 2016; Koulaguina & Shi, 2019; Emond & Shi, 2021; Shi & Emond, 2023). In the last several years, the TSP has been used in numerous corpus and computational studies of acquisition (e.g., Yang 2016; Kodner 2020; Belth et al. 2021; Björnsdóttir 2021; Pearl & Sprouse 2021; Richter 2021; Van Tuijl & Coopmans 2021; Trips & Rainsford 2022; Belth 2023b; Henke 2023; Payne 2023). The TSP threshold evaluates a rule once one is proposed, but it naturally lends itself to a recursive procedure of proposing, evaluating, and refining (proposing anew) rules (Yang, 2016:ch.3). Belth et al. (2021) developed a computational implementation of such a recursive approach, which I described above in the context of Turkish voicing alternations: the procedure recursively subdivides the input, yielding a decision tree where the path to a leaf is the structural description of a rule and the base case of the recursion is when the rule represented by the path passes the TSP (2). I now turn to describing the features available to the recursive procedure in subdividing (§ 2.2) and then how these are chosen (that is, how rules are proposed and refined), also via the TSP (§ 2.3).

2.2 *Features* In Becker et al. (2011)'s corpus analysis, whether a noun stem was monosyllabic or polysyllabic influences whether it is likely to be an alternating or non-alternating noun. I will call this feature MonoSyl. Whether the stem ends in a complex coda (i.e. a consonant cluster) was also an identified factor. I will call this feature CMPLX. By 3 years of age, Turkish-learning children are producing both monosyllabic and bisyllabic words without reducing bisyllabic words to monosyllabic ones (Kopkalli-Yavuz & Topbas, 2000). Likewise, by 3 years of age children are producing many final consonant clusters without reduction, with a few taking until age 5 (Topbas, 2006). While Topbas & Kopkalli-Yavuz (1998) report that, shortly after age 2, they found no errors for alternating words in a naturalistic corpus, in an elicitation study, Nakipoglu et al. (2016) found extensive errors—mostly producing the voiceless counterpart of a stop where a voiced one would be expected—through age 8;11. Based on these developmental studies, it appears that children have decent command of the relevant aspects of the prosodic inventory (monosyllabicity and complex codas) prior to fully acquiring the voicing alternation.

In Becker et al. (2011)'s study, the identity or place of articulation of the final stop/affricate was also substantially predictive, along with, to a lesser degree, the quality (height and backness) of the preceding vowel. These latter two features are segmental features of the stem. As I proposed in Belth (2023a, 2024b), when learning a phonological alternation, learners' attention is drawn first to the alternating segment. In this case, that would be the final stop. Only when a sufficient generalization cannot be formed do learners look further (Belth, 2023a) or change representations (Belth, 2024b). Consequently, the preceding vowels will only enter into the learner's attention if the already-available prosodic shape features and the segmental features of the final consonant have already been considered. This is based on the hypothesis that the well-attested and non-language-specific ability to track adjacent dependencies (Saffran et al., 1996, 1997; Aslin et al., 1998; Saffran et al., 1999; Fiser & Aslin, 2002) underlies alternation learning. The hypothesis also has some direct experimental support through an artificial language learning study (Belth, 2023c:ch. 6). In short, the basic idea is that attention starts at the alternating segment and moves outward over a representation from there. Consequently, the learner's attention runs into segments that are closer before it runs into those that are farther away.

2.3 *Choosing the Split Feature* In choosing, out of the currently-available features, one to split on, the proposed model considers all features that occur with at least one of the nouns at the current level in the recursion. A feature is chosen with the goal of separating alternating from non-alternating nouns, focusing on the type (alternating vs. non-alternating) that is more frequent. If the alternating nouns are more frequent, then the model searches for a feature that occurs with the alternating nouns but not with the non-alternating nouns. On the flip side, if the number of non-alternating nouns is greater, the model instead searches for a feature that occurs with the non-alternating nouns but not with the alternating nouns. Let us call the kind of nouns that are more frequent (alternating or non-alternating) the *target* nouns.

The model searchers for any feature where the number of target nouns that has the feature (m) out of the total number of target nouns (n) passes the TSP—that is, that being a target item implies having the feature, at least enough to pass the TSP (3).

(3)
$$n - m \leq \frac{n}{\ln n}$$

If more than one such feature is found, the model selects the one that occurs least often with the non-target nouns. If no such feature exists, the model instead seeks a feature where *m* out of the total number of nouns with that feature (*n*) are target nouns—in other words that being a target item is implied by having the feature, at least enough to pass the TSP.

3 Evaluation

3.1 *Data and Setup* The data for the evaluation comes from the Turkish Electronic Living Lexicon (TELL; Inkelas et al. 2000). TELL contains about 15K nouns that have transcribed pronunciations for both stem and possessive (Poss) forms. This was the same corpus used by Becker et al. (2011). I augmented the TELL Poss nouns with token frequencies from a web crawl (Wenzek et al., 2020). This enabled taking frequency-weighted samples to simulate multiple Turkish learners with variation in their vocabulary.

I simulated 24 learners, which is the number of participants in Becker et al. 2011's Wug test. For each simulated learner, I sampled a vocabulary size from a normal distribution $n \sim \text{Normal}(\mu = 10,000, \sigma = 1,000)$ and then sampled *n* unique nouns from TELL, weighted by frequency. After training each model I then had it predict, for each of Becker et al. (2011)'s nonce words, whether or not it alternates. I then compared the fraction of models that predicted each nonce to alternate to the fraction of Becker et al. (2011)'s participants who did so.

For comparison, I fit two logistic regression models to the TELL Poss nouns. These correspond to the regression models fit by Becker et al. (2011). The first regression model included vowel features as predictors: $alter \sim place * size + place * back + place * high$. The second included only place and size as predictors: $alter \sim place * size$. The predictor $place$ captures the place of articulation of the final consonant. Following Becker et al. (2011), the predictor $size$ combined the monosyllabicity (MonoSyl) and the complex coda features (CMPLX). Thus, the first model represents the hypothesis that learners have access to vowel features just as all the other features, whereas the second model represents the hypothesis that learners have no access to the vowel features. I used each of these regression models to estimate the probability of each nonce alternating, and compared this probability to the human alternation rates.

3.2 *Results* The results are shown in Table. 1. The proposed model achieves 0.83 correlation with Becker et al. (2011)'s participant responses. This is comparable to the 0.82 correlation achieved by the logistic regression model that, by specification, rules out the possibility of vowel quality influencing the alternation. The regression model that lacks vowel predictors correlated better than the model with vowel predictors (0.80), which is consistent with Becker et al. (2011)'s finding that Turkish speakers did not show a statistically robust sensitivity to the vowel dependencies. Thus, the proposed model accounts for the Turkish speakers' behavior at least as well as a model that analytically rules out the possibility of vowel quality influencing the laryngeal features of a following consonant.

An example tree constructed by the proposed model is shown in Figure 2. In the tree, +/− means +alternate/−alternate and "*" means that neither alternation nor non-alternation were productive for that partition. The first split in the tree partitions words based on whether the stem ends in [t, p, k, t͡ʃ], which makes sense since only these sounds alternate in Turkish (fricatives and sonorants do not). The second split is based on whether the noun stem is monosyllabic or not. Many of the leaves down the right branch of this second split are "+," which is consistent with Becker et al. (2011) corpus analysis, where polysyllabic nouns are more likely to alternate. Also consistent with Becker et al.'s corpus analysis, a number of splits are based on the identity (equivalently place of articulation) of the final stop and whether the stem ends in a complex coda. 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 (0.83 vs. 0.82), and is thus as consistent with Becker et al. (2011)'s wug test. It is possible that Becker et al.'s participants were in fact marginally sensitive to the vowel quality, just not enough for it to show up statistically.

In short, the proposed 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 constructed, rather than a reflection of an analytical filter in UG.

Table 1: The proposed model's results, compared to two regression models that constitute the hypotheses that the vowel features cannot be used (No Vowels) in predicting the voicing alternation or that they have equal potential to be used (Vowels). Despite not ruling out the the vowel features, the proposed model correlates with Turkish speakers' Wug test behavior marginally better than the No Vowels model, which analytically rules out the possibility of vowel dependencies being used.

Model	Pearson's r	p-value
Proposed Model	0.83	< 0.001
No Vowels ($alter \sim place * size$)	0.82	< 0.001
Vowels ($alter \sim place * size + place * back + place * highi$)	0.80	< 0.001

4 Discussion

In this paper, I have made a brief argument about why Turkish speakers show no robust evidence of generalizing a dependency between the laryngeal features of stem-final stops/affricates and the quality of the preceding vowel even though Becker et al. (2011) provide evidence that it is tenable to form such a generalization from a corpus of the language. I have argued that it may be a reflection of a particular learner-input scenario rather than a UG analytic bias against this sort of dependency. In particular, the result arises from a learner that attempts to form a grammar that works well enough with the sparse input, together with considerations about the representations over which the learner generalizes. Despite the smallness of the argument, I believe that it both carries implications for and weaves into a bigger story about the relationship between phonological theory and learning. In the learning proposal I have given, whether a particular grammar contains reference to the vowel-consonant dependencies of interest is not a property of the learner. Figure 1 makes this clear: the learner is capable of forming grammars that do (right) or do not (left) make use of the vowel quality in the rules contained in the grammar. Nor is it a property of the input, divorced from the learner: we cannot look at a corpus and determine, without reference to a learner, what will go into a grammar. Rather, the content and form of the grammar is a property of the learner interacting with the input.

The present proposal contrasts with Becker et al. (2011)'s suggestion that UG might provide an analytic bias against vowel quality influencing the laryngeal features of a following consonant. In comparing these proposals, I followed Becker et al.'s use of a regression model that excludes vowel feature predictors as representing this UG proposal. Both my proposal and theirs do more or less the same at accounting for the Turkish speakers' Wug test results (0.83 vs. 0.82 correlation). Thus, these results alone do not serve to decide between the proposals. At least, however, these results alone do indicate that it is not *necessary* to appeal to a bias against this particular kind of interaction to achieve this level of correlation with human behavior.

What about further considerations? The same sort of approach as I have applied to this case study has been applied with success to the acquisition of morphological inflection (Belth et al., 2021), morphological marking (Payne, 2023), Dutch voicing alternations (Belth, 2024a), and the formation of phonological tiers (Belth, 2024b), to name a few examples. These prior results might lend support for the approach's application to Turkish voicing alternations. On the other hand, Becker et al. (2011) marshal as further support for their proposal the typological observation that it is rare cross-linguistically for vowel quality to influence the laryngeal features of a following consonant. This sort of typological appeal is common in phonological theory. It carries with it the hope that the same sorts of causes can simultaneously explain certain facts of particular grammars (here those of the Turkish participants) and typological patterns. If a single UG bias can account for this quirk of Turkish grammars and for the typological pattern, then, in a sense, it appears to have greater explanatory breadth and thereby to be preferable to an account of the grammars, such as mine, that leaves the typological pattern explanation to other causes.

In my opinion, the fact that my proposal does not simultaneously explain this quirk of Turkish grammars and the typological generalization is no flaw. First, the preference for theories that explain properties of particular grammars and, simultaneously, typology, is, I think, rooted in the idea that theory development should strictly maximize coverage. Yet successful theoretical development often violates this prescriptive

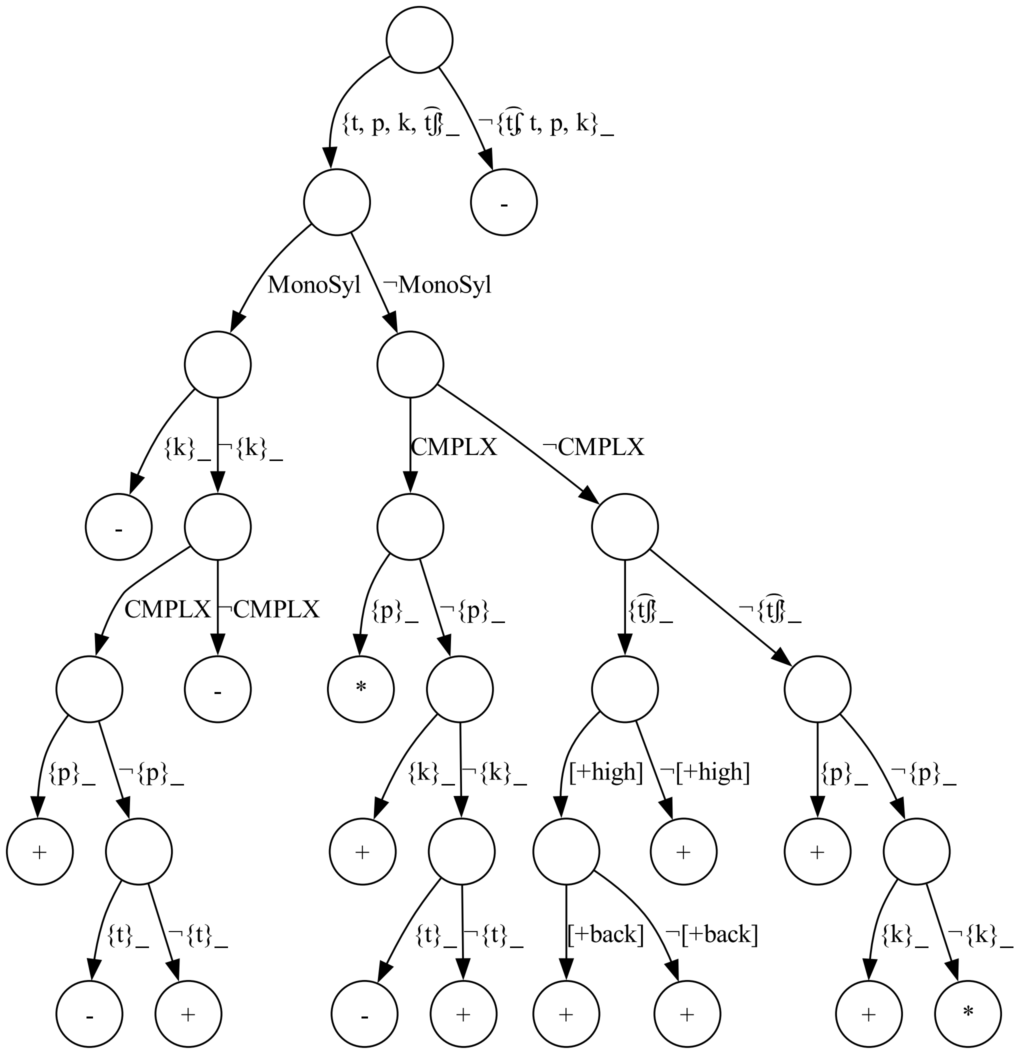

Figure 2: An example of a grammar that the proposed model constructed. The first split separates the nouns with stems that end in stops or affricates from those that do not, as only the former alternate in Turkish. The model regularly subdivides based on the prosodic shape of the word (MonoSyl and CMPLX) and on the identity (or, equivalently, the place of articulation) of the final stop/affricate. The model comparatively rarely subdivides based on the height or backness of the preceding vowel, with the few cases where it does covering less than 1% of training nouns.

demand (Feyerabend, 1975). Moreover, different kinds of questions typically demand different kinds of explanations (Mayr, 1961): there is no reason to believe that an explanation for what generalizations are (or are not) in a grammar can or should always simultaneously serve as an explanation for why languages pattern the way they do. The latter sort of *why* question, in domains such as linguistics and evolutionary biology, where the objects of study have inherited properties, compel historical explanations (Mayr, 1961; Gould, 1989). In fact there is good reason to think that the causes underlying what generalizations learners form are not, in general, the same causes underlying why languages pattern the way that they do. In experimental settings, humans time and again learn typologically rare or unattested patterns with little trouble (Seidl & Buckley, 2005; Baer-Henney & van de Vijver, 2012; Beguš, 2018; Belth, 2023c). And, while patterns that are typologically rare are of course rare on their own (tautologically), linguists are nevertheless constantly turning up such patterns, making *typologically rare* patterns, as a category, quite common (Bach & Harms, 1972; Buckley, 2000; Coetzee & Pretorius, 2010; Johnsen, 2012). Each of these, when productive for speakers, constitute evidence that humans must be able to learn them in natural language learning scenarios.

Ultimately, it is preferable for the proposal, and competing alternatives, to be evaluated on further empirical grounds in addition to these conceptual considerations. In future work, I plan to apply the proposal incrementally to model the acquisition trajectory of voicing alternations. If the model accounts for the developmental patterns of Turkish voicing alternations (Topbas & Kopkalli-Yavuz, 1998; Nakipoglu et al., 2016), this would lend further support to the proposal as an explanation for adult Turkish speakers' behavior in Becker et al. (2011)'s study. Furthermore, since my proposal predicts that nothing inhibits learners from forming generalizations that capture a relation between vowel quality and the laryngeal features of a following consonant, experimental settings could be created to test this prediction.

References

Altan, Asli (2009). Acquisition of vowel harmony in Turkish. *35. yıl Yazıları*, Multilingual Yabancı Dil Yayınları, 9–26.

Aslin, Richard N., Jenny R. Saffran & Elissa L. Newport (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science* 9:4, 321–324, URL https://doi.org/10.1111/1467-9280.00063.

Bach, Emmon & Robert T Harms (1972). How do languages get crazy rules. *Linguistic change and generative theory* 1, p. 21.

Baer-Henney, Dinah & Ruben van de Vijver (2012). On the role of substance, locality, and amount of exposure in the acquisition of morphophonemic alternations. *Laboratory Phonology* 3:2, 221–249, URL https://doi.org/10.1515/lp-2012-0013.

Becker, Michael, Nihan Ketrez & Andrew Nevins (2011). The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 84–125, URL https://muse.jhu.edu/article/422024.

Becker, Michael, Andrew Nevins & Jonathan Levine (2012). Asymmetries in generalizing alternations to and from initial syllables. *Language* 231–268, URL https://muse.jhu.edu/article/479325.

Beguš, Gašper (2018). *Unnatural phonology: A synchrony-diachrony interface approach*. Ph.D. thesis, Harvard University.

Belth, Caleb (2023a). A learning-based account of local phonological processes. *Phonology* 40, 1–33, URL https://doi.org/10.1017/S0952675724000083.

Belth, Caleb (2023b). Towards a learning-based account of underlying forms: A case study in Turkish. Hunter, Tim & Brandon Prickett (eds.), *Proceedings of the Society for Computation in Linguistics, 2023, Amherst, MA*, Association for Computational Linguistics, 332–342, URL https://aclanthology.org/2023.scil-1.29.

Belth, Caleb (2023c). *Towards an Algorithmic Account of Phonological Rules and Representations*. Ph.D. thesis, University of Michigan, URL https://dx.doi.org/10.7302/8399.

Belth, Caleb (2024a). A learning-based account of non-productivity in Dutch voicing alternations. *Proceedings of the 48nd annual Boston University Conference on Language Development, 2023, Boston*, URL https://www.lingref.com/bucld/48/BUCLD48-05.pdf.

Belth, Caleb (2024b). A learning-based account of phonological tiers. *Linguistic Inquiry* p. 1–37, URL https://doi.org/10.1162/ling_a_00530.

Belth, Caleb, Sarah Payne, Deniz Beser, Jordan Kodner & Charles Yang (2021). The greedy and recursive search for morphological productivity. *Proceedings of the Annual Meeting of the Cognitive Science Society, 2021, Online*, vol. 43, URL https://escholarship.org/uc/item/1md2p6j5.

Berko, Jean (1958). The child's learning of english morphology. *Word* 14:2-3, 150–177, URL https://doi.org/10.1080/00437956.1958.11659661.

Björnsdóttir, Sigríður Mjöll (2021). Productivity and the acquisition of gender. *Journal of Child Language* 48:6, 1209–1234, URL https://doi.org/10.1017/S0305000920000732.

Buckley, Eugene (2000). What should phonology explain. *Handout from SUNY Buffalo linguistics colloquium*.

Chan, Erwin (2008). *Structures and distributions in morphology learning*. Ph.D. thesis, University of Pennsylvania.

Coetzee, Andries W & Rigardt Pretorius (2010). Phonetically grounded phonology and sound change: The case of Tswana labial plosives. *Journal of Phonetics* 38:3, 404–421, URL https://doi.org/10.1016/j.wocn.2010.03.004.

Ekmekci, Ozden Fatma (1979). *Acquisition of Turkish: A Longitudinal Study on the Early Language Development of a Turkish child.* Ph.D. thesis, The University of Texas at Austin.

Emond, Emeryse & Rushen Shi (2021). Infants' rule generalization is governed by the Tolerance Principle. Dionne, Danielle & Lee-Ann Vidal Covas (eds.), *Proceedings of the 45nd annual Boston University Conference on Language Development, 2020, Boston*, Somerville, MA, 191–204.

Ernestus, Mirjam Theresia Constantia & R Harald Baayen (2003). Predicting the unpredictable: Interpreting neutralized segments in dutch. *Language* 79:1, 5–38, URL https://doi.org/10.1353/lan.2003.0076.

Feyerabend, Paul (1975). *Against method*. Verso, London.

Fiser, József & Richard N Aslin (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28:3, 458–467, URL https://doi.org/10.1037//0278-7393.28.3.458.

Gould, Stephen J (1989). *Wonderful Life: The Burgess Shale and the Nature of History*. WW Norton & Company.

Hayes, Bruce & Zsuzsa Cziráky Londe (2006). Stochastic phonological knowledge: The case of hungarian vowel harmony. *Phonology* 23:1, 59–104, URL https://doi.org/10.1017/S0952675706000765.

Henke, Ryan E (2023). Rules and exceptions: A Tolerance Principle account of the possessive suffix in Northern East Cree. *Journal of Child Language* 50:5, 1119–1154, URL https://doi.org/10.1017/S0305000922000277.

Inkelas, Sharon, Aylin Küntay, Orhan Orgun & Ronald Sprouse (2000). Turkish electronic living lexicon (TELL). *Turkic Languages* 4, 253–275.

Johnsen, Sverre Stausland (2012). A diachronic account of phonological unnaturalness. *Phonology* 29:3, 505–531, URL https://www.jstor.org/stable/23325590.

Kodner, Jordan (2020). *Language acquisition in the past*. Ph.D. thesis, University of Pennsylvania.

Kopkallı, Handan (1993). *A phonetic and phonological analysis of final devoicing in Turkish*. Ph.D. thesis, University of Michigan, URL https://deepblue.lib.umich.edu/handle/2027.42/129074.

Kopkalli-Yavuz, Handan & S Topbas (2000). Children's preferences in early phonological acquisition: How does it reflect sensitivity to the ambient language. *Studies on Turkish and Turkic languages: Proceedings of the Ninth International Conference on Turkish Linguistics*, 289–295.

Koulaguina, Elena & Rushen Shi (2019). Rule generalization from inconsistent input in early infancy. *Language Acquisition* 26:4, 416–435, URL https://doi.org/10.1080/10489223.2019.1572148.

Mayr, Ernst (1961). Cause and effect in biology: kinds of causes, predictability, and teleology are viewed by a practicing biologist. *Science* 134:3489, 1501–1506, URL https://www.jstor.org/stable/1707986.

Medin, Douglas L, William D Wattenmaker & Sarah E Hampson (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive psychology* 19:2, 242–279, URL https://doi.org/10.1016/0010-0285(87)90012-0.

Nakipoglu, Mine, Reyhan Furman & Asli Üntak (2016). Acquisition of morphophonemic alternations and the role of frequency. *Exploring the Turkish linguistic landscape: Essays in honor of Eser Erguvanlı-Taylan*, John Benjamins Publishing Company, 157–182, URL https://doi.org/10.1075/slcs.175.10nak.

Nelson, Deborah G Kemler (1984). The effect of intention on what concepts are acquired. *Journal of Verbal Learning and Verbal Behavior* 23:6, 734–759, URL https://doi.org/10.1016/S0022-5371(84)90442-0.

Payne, Sarah (2023). Contrast, sufficiency, and the acquisition of morphological marking. *Proceedings of BUCLD*, vol. 47, 604–617, URL https://www.lingref.com/bucld/47/BUCLD47-49.pdf.

Pearl, Lisa & Jon Sprouse (2021). The acquisition of linking theories: A tolerance and sufficiency principle approach to deriving utah and rutah. *Language Acquisition* 28:3, 294–325, URL https://doi.org/10.1080/10489223.2021.1888295.

Richter, Caitlin (2021). *Alternation-Sensitive Phoneme Learning: Implications For Children's Development And Language Change*. Ph.D. thesis, University of Pennsylvania.

Saffran, Jenny R, Richard N Aslin & Elissa L Newport (1996). Statistical learning by 8-month-old infants. *Science* 274:5294, 1926–1928, URL https://doi.org/10.1126/science.274.5294.1926.

Saffran, Jenny R., Elissa L. Newport, Richard N. Aslin, Rachel A. Tunick & Sandra Barrueco (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science* 8:2, 101–105, URL https://doi.org/10.1111/j.1467-9280.1997.tb00690.x.

Saffran, Jenny R, Elizabeth K Johnson, Richard N Aslin & Elissa L Newport (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70:1, 27–52, URL https://doi.org/10.1016/s0010-0277(98)00075-4.

Schuler, Kathryn, Charles Yang & Elissa Newport (2016). Testing the Tolerance Principle: Children form productive rules when it is more computationally efficient to do so. *Proceedings of the 38th annual meeting of the Cognitive Science Society, 2016, Philadelphia*, URL https://escholarship.org/uc/item/1cn5n52p.

Seidl, Amanda & Eugene Buckley (2005). On the learning of arbitrary phonological rules. *Language Learning and Development* 1:3–4, 289–316, URL https://doi.org/10.1080/15475441.2005.9671950.

Shi, Rushen & Emeryse Emond (2023). The threshold of rule productivity in infants. *Frontiers in Psychology* 14, p. 1251124, URL https://doi.org/10.3389/fpsyg.2023.1251124.

Smith, Linda B (1979). Perceptual development and category generalization. *Child Development* 705–715.

Tesar, Bruce & Paul Smolensky (1998). Learnability in optimality theory. *Linguistic inquiry* 29:2, 229–268, URL https://doi.org/10.1162/002438998553734.

Topbas, Seyhun (2006). Türkçe sesletim-sesbilgisi testi: Geçerlik-güvenirlik ve standardizasyon çalismasi. *Türk Psikoloji Dergisi* 21:58, p. 39.

Topbas, S & H Kopkalli-Yavuz (1998). The onset of a linguistic system: Is there evidence from the acquisition of final devoicing in Turkish. *Proceedings of GALA*, 284–293.

Trips, Carola & Thomas Rainsford (2022). Tolerating subject-experiencers? Yang's Tolerance Principle applied to psych verbs under contact in Middle English. *Journal of Historical Syntax* 6:12, 1–43, URL https://doi.org/10.18148/hs/2022.v6i12.118.

Van Tuijl, Rosita & Peter Coopmans (2021). The productivity of Dutch diminutives. *Linguistics in the Netherlands* 38:1, 128–143, URL https://doi.org/10.1075/avt.00053.tui.

Wenzek, Guillaume, Marie-Anne Lachaux, Alexis Conneau, Vishrav Chaudhary, Francisco Guzmán, Armand Joulin & Edouard Grave (2020). CCNet: Extracting high quality monolingual datasets from web crawl data. Calzolari, Nicoletta, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Twelfth Language Resources and Evaluation Conference*, European Language Resources Association, Marseille, France, 4003–4012, URL https://aclanthology.org/2020.lrec-1.494.

Yang, Charles (2016). *The Price of Linguistic Productivity: How Children Learn to Break the Rules of Language*. The MIT Press, Cambridge, MA, URL https://doi.org/10.7551/mitpress/9780262035323.001.0001.