Towards a Learning-Based Account of Underlying Forms: A Case Study in Turkish

Caleb Belth
University of Michigan
cbelth@umich.edu

Abstract

A traditional concept in phonological theory is that of the underlying form. However, the history of phonology has witnessed a debate about how abstract underlying representations ought to be allowed to be, and a number of arguments have been given that phonology should abandon such representations altogether. In this paper, we consider a learning-based approach to the question. We propose a model that, by default, constructs concrete representations of morphemes. When and only when such concrete representations make it challenging to generalize in the face of the sparse statistical profile of language, our proposed model constructs abstract underlying forms that allow for effective generalization. As a case study, we consider the highly agglutinative language, Turkish. We demonstrate that the underlying forms that our model constructs account for the complexities of Turkish phonology resulting from its multifaceted vowel harmony. Moreover, these underlying forms enable the highly-accurate prediction of novel surface forms, demonstrating the importance of some underlying forms to generalization.

1 Introduction

A traditional conception of phonological theory involves abstract underlying representations (URs) together with phonological processes (stated as rules or constraints) mapping between this abstract level of representation and a concrete, surface-level representation. Debates in the 1960's and 1970's questioned how abstract URs should be allowed to be (Hyman, 2018, p. 597), with a particularly famous article by Kiparsky (1968) arguing that the positing of non-concrete representations should only be done when motivated. Any perception of this debate as fading in subsequent years is probably better attributed to the field moving on to other questions than it is to a satisfactory resolution of the debate (Anderson, 2021).

Indeed, some phonologists have taken the position that URs should not be used in phonological theory because doing so is "(i) wrong, (ii) redundant, (iii) indeterminate, (iv) insufficient, or (v) uninteresting," as Hyman (2018, p. 591) summarized the objections. Meanwhile, much of the work on learning phonology has either focused on surface restrictions (e.g., Hayes and Wilson 2008) or continued to assume URs (e.g., Tesar and Smolensky 1998; Boersma 1997), abstracting away from the question of how (and if) such representations are constructed (see Jarosz 2019 for a summary).

One of the main justifications for the use of underlying representations is to capture generalizations. For example, the form of the English plural affix—[z], [s], or [əz]—depends on the stem-final segment, but is predictable from the stem-final segment, as in (1).

(1) [dɑg-z]
[kæt-s]
[hɔrs-əz]

Positing an underlying /-z/ derived by process into [z], [s], or [əz] allows this generalization to be captured. However such an analysis is not necessary. The allomorphs could each be listed along with a set of sounds each occurs after, or the apparent relationship between singulars and plurals could be ignored altogether and both forms could simply be memorized.[1]

How then are we to choose from these analyses? Is the desire to capture a generalization sufficient motivation to choose the /-z/ analysis? In this work we propose a learning-based approach to this question. Specifically, we propose a computational model that assumes, by default, that underlying forms are fully concrete. The model attempts to form morphological generalizations out of sheer

[1] As one reviewer pointed out, evidence of overgeneralization (e.g., MacWhinney 1978) suggests that memorization is not an empirically-tenable hypothesis in all cases.

necessity to deal with the sparse statistical profile of language (Yang 2016, ch. 2; Chan 2008).

The question then becomes learning-based: when does surface-alternation of a morpheme prevent the learner from forming morphological generalizations from concrete representations? In some—but critically not all—cases, surface-alternations are pervasive enough to drive the learner to resort to abstract URs in order to effectively generalize. We present the model in § 2.

We evaluate the model on natural-language corpora of the highly agglutinative language Turkish, demonstrating both when abstract URs are necessary for generalization and when they are not (§ 3). When combined with a recent model for learning local and non-local alternations, the proposed model achieves high accuracy generalizing to held-out test words (§ 3.4).

2 Model

2.1 Model Input

The input to the model is a set of morphologically-analyzed surface forms. An example input of nine forms is shown in Tab. 1. These word forms are processed by the model incrementally, modeling the growth of a learner's lexicon.

While morphological segmentation is an important area of study in its own right, we believe it is a justified assumption given experimental evidence that infants can effectively morphologically segment nonce words. These results have been observed for French-learning 11mo-old (Marquis and Shi, 2012) and English-learning 15mo-old (Mintz, 2013) infants. The finding is corroborated by results for 15mo Hungarian-learning infants, despite the high-level of agglutination in Hungarian (Ladányi et al., 2020).

2.2 Model Output

The output of the model is a lexicon, which contains a representation for each morpheme, and a lexicalized list of any input word forms not decomposable into those morphemes. The representation of a morpheme may be concrete or abstract.[2] As discussed by Ettlinger (2008, sec. 4.3.4), a UR can be called *abstract* because it lacks the phonetic detail of an actual speech sound (e.g., /D/ as an alveolar stop lacking a voicing specification), or because it contains different segments from a surface form. For simplicity, we will refer to the representation constructed in the lexicon as a UR, regardless of its abstractness. This assumes, following prior work (§ 4), that each morpheme has a single UR. Future work will consider scenarios where this may not be the case.

Surface Form	Morphological Analysis
1. [buz-lɑr]	'ice-PL'
2. [kɯz-lɑr]	'girl'-PL
3. [el-ler]	'hand-PL'
4. [jer-ler-in]	'place'-PL-GEN
5. [søz-ler]	'word-PL'
6. [dɑl-lɑr-ɯn]	'branch'-PL-GEN
7. [sɑp-lɑr]	'stalk-PL'
8. [jyz-yn]	'face'-GEN
9. [ip-ler-in]	'rope'-PL-GEN

Table 1: An example Turkish input consisting of morphologically-analyzed surface forms.

2.3 Model Description

By default, the model creates a concrete UR for each morpheme. Prior work (§ 4) often resorts to phonological processes to produce the various surface forms of a morpheme at the first instance of surface alternation. Our model differs from this approach by treating underlying forms as concrete even after the first instance of surface alternation. Instead of immediately collapsing surface forms into a single, abstract UR, our model simply lexicalizes all word forms in which a morpheme occurs as something other than its most frequent form. It is only when the resulting lexicalization becomes unsustainable (see § 2.4) that the model then constructs abstract underlying forms from which the surface realizations are derived by morphophonological process.

The pseudocode for the algorithm is shown in (2).[3] As discussed in § 2.1, the input to the model is an incremental stream of morphologically-analyzed surface forms. Whenever the model receives a new surface form (2; step 1), it initially creates a concrete underlying form for each morpheme, storing the most frequent form of the morpheme concretely (2; step 3), and lexicalizes any wordforms that contain a different form of the morpheme (2; step 8). However, if too many word-

[2]We treat surface and underlying representations, whether concrete or abstract, as sequences of segments, where each segment is a set of distinctive features.

[3]Code is available at https://github.com/cbelth/underlying-forms-SCiL

Meaning	UR	Plural Form
PL	/lɑr/	N/A
‘ice’ ‘girl’	/buz/ /kɯz/	**Stem-PL**
‘hand’	/el/	/el-ler/

Table 2: When the first three words from Tab. 1 enter the lexicon, the stems and plural affix are all stored concretely (left two columns). The plural form of the ‘ice’ and ‘girl’ stems are predictably decomposable into their concrete stems and the PL affix (denoted with the boldface concatenation), so those forms need not be stored in the lexicon. However, with /-lɑr/ as the UR of the plural, the plural form of ‘hand’ cannot be so decomposed, so it is instead lexicalized.

forms in the lexicon are exceptions—where the measurement of “too many” occurs as described in § 2.4—the model instead constructs an abstract UR (2; step 5) and then learns a phonological process, via a separate model (see § 2.6), to account for the resulting alternation.

(2) **Input:** Incremental stream of morphologically analyzed SRs

1. **While** surface form in input **do**
2. – **For** morpheme in segmentation **do**
3. — Morpheme UR ← most freq form
4. — **If** too many alternative forms **do**
5. —– Construct abstract UR
6. —– Learn phonological process
7. — **Else do**
8. —– Lexicalize exceptions

For example, consider the PL suffix after the first 2 (of 9) inputs listed in Tab. 1 have entered the learner’s lexicon. At this point, the model will be storing the only attested surface form [-lɑr] as the concrete UR /-lɑr/.

When the third word enters the lexicon, our model will lexicalize the form ‘hand-PL’ as /el-ler/, rather than immediately constructing an abstract PL morpheme. This is shown in Tab. 2, where each stem and the plural affix have concrete underlying forms, and the plural form of ‘ice’ and ‘girl’ are formed by suffixing the plural to the stem, but the plural form of ‘hand’ is lexicalized.

By the time all 9 words enter the lexicon, however, there will be 4 instances of [-lɑr] and 4 of [-ler], making it no longer sustainable to keep a concrete underlying form. The difference between these two scenarios and, more generally, the decision of when to create an abstract underlying form, is made by the Tolerance Principle (Yang, 2016), as described next.

2.4 When is Abstraction Needed?

In order to detect when the amount of surface alternation that prohibits generalization from concrete representations, the model uses the Tolerance Principle (TP), proposed by Yang (2016). The TP is a cognitively-grounded tipping point, which hypothesizes that children form productive generalizations when the number of exceptions to a proposed generalization results in a real-time processing cost lower than that without the generalization. The exact derivation of the TP is provided by Yang (2016, ch. 3), but rests critically upon the empirical observation of linguistic sparsity. The TP has had much prior success in computational modeling, lexical, and experimental studies (Schuler et al., 2016; Yang, 2016; Richter, 2018; Koulaguina and Shi, 2019; Emond and Shi, 2021; Richter, 2021; Belth et al., 2021; Payne, 2022; Belth, 2023).

Our model’s default treatment of underlying forms as concrete can be stated as a morpheme-specific rule. In the example above, where only the first 2 words of Tab. 1 have entered the lexicon, the rule for the PL form would be (3), which predicts that the PL morpheme is realized as [-lɑr].

(3) If PL then [-lɑr]

The TP threshold, which evaluates a linguistic rule (generalization), is stated in (4), where n is the number of items the rule applies to and e is the number of exceptions to the rule.

(4)
$$e \leq \frac{n}{\ln n}$$

Thus, our model tracks—for each morpheme—the number of observed words in which the morpheme appears (n) and the number of those where surface alternation leads the morpheme to be realized as something other than its hypothesized concrete form (e).

If the (4) threshold is met, then the UR remains concrete and the word forms where the suffix is realized as something else are lexicalized[4] as exceptions. For example, when the 3rd item in Tab. 1

[4]By lexicalization, we mean that the word form is stored in the lexicon verbatim instead of being decomposed into the underlying morphemes. See Tab. 2 for an example.

Morphemes		Word Forms		
Meaning	UR	PL Form	GEN From	PL, GEN Form
PL	/lɑr/	N/A	N/A	N/A
GEN	/in/	N/A	N/A	N/A
'ice'	/buz/		??	??
'girl'	/kɯz/	**Stem-PL**	??	??
'stalk'	/sɑp/		??	??
'hand'	/el/	/el-ler/	??	??
'word'	/søz/	/søz-ler/	??	??
'face'	/jyz/	??	/jyz-yn/	??
'place'	/jer/	??	??	/jer-ler-in/
'branch'	/dɑl/	??	??	/dɑl-lɑr-ɯn/
'rope'	/ip/	??	??	/ip-ler-in/

Table 3: The left two columns contain morphemes—meaning and form (UR); the right three columns contain word forms. Boldface denotes word forms that can be predictably decomposed into concrete underlying forms, while '/-/' notation denotes word forms that must be lexicalized. The '??' denotes word forms that are unknown. Once all nine words from Tab. 1 enter the lexicon, most forms (6 of 9) cannot be predictably decomposed into concrete underlying forms, so the model constructs abstract URs, as described in § 2.5.

enters the lexicon, the realization of PL as [-ler] violates (3). However, with only three word forms containing PL this one exception can be lexicalized, since $1 \leq 3/\ln 3$.

On the other hand, if the (4) threshold is violated—i.e., $n > \frac{n}{\ln n}$—then the model constructs an abstract underlying form. For example, when the 9th item of Tab. 1 enters the lexicon, the realization of PL as [-ler] becomes the 4th of 8 forms in which PL is realized as [-ler] instead of the [-lɑr] predicted by (3). Because $4 > 8/\ln 8$, the model will construct an abstract UR for the PL morpheme.

This is shown in Tab. 3, where the plural is realized as [-lɑr] in 3 plural forms and 1 plural, genitive form, but there are 4 forms that must be lexicalized because they instead have the [-ler] form.[5]

Constructing abstract URs introduces discrepancies between URs and SRs for any word forms containing the morpheme, so our model then passes the (UR, SR) pairs implicit in its lexicon[6] to a model that learns phonological alternations to account for the newly-introduced discrepancies. The process of constructing abstract URs is described in § 2.5 and the process of learning what conditions the alternations is described in § 2.6.

[5]Note that the PL, GEN of 'branch' is lexicalized because the GEN affix is realized in a form other than [in], not because of the PL affix, which is why that form does not get counted as an exception in the TP calculation for the PL affix.

[6]See § 2.6 for a description of how the set of (UR, SR) pairs is computed.

2.5 Constructing Abstract URs

The model's first step in constructing an abstract UR for a morpheme is to create the set of forms that the morpheme is realized as. For example, the forms of the GEN affix attested in Tab. 1 are [-in] / [-ɯn] / [-yn], and of the PL affix are [-lɑr] / [-ler].

Next, the model aligns each of the forms. This is trivial for fixed-length affixes (e.g., the case of the PL affix). If the length of the forms are not all the same, then the model counts the lengths of the morpheme's realizations. For example, the dative affix can be realized as [-ɑ] or [-e], but may contain an affix-initial [j] when attaching to a morpheme that ends in a vowel. The model thus counts the number of words in which [-ɑ] or [-e] (length 1) is the realization, and the number in which [-jɑ] or [-je] is the realization (length 2), and chooses the most frequent length as the length of the UR. If a shorter length is chosen, the extra segment(s) are treated as epenthesized; if the longer is chosen, they are treated as deleted. For simplicity, we assume that these segments epenthesize or delete on the left, which is a simplification. This process is not guaranteed to generalize to other languages, so future work will develop a more robust alignment process by more tightly combining the problems of abstract UR construction and rule construction.

Once the forms are aligned, the UR is constructed one segment at a time. Each segment is set to match in features where all realizations of the affix match; features that alternate across forms are unspecified underlyingly. For example, [-lɑr] / [-ler] will lead to /-lAr/, where A is the low, unround vowel with backness unspecified, because both forms agree in the initial and final segments, but the vowel alternates on backness. Similarly, [-in] / [-ɯn] / [-yn] will result in /-Hn/, where H is the high vowel with backness and roundness unspecified, since [i] and [y] differ in backness from [ɯ] while [i] and [ɯ] differ from [y] in roundness.

2.6 Learning Alternations

When the number of words where the morpheme's surface alternation requires the word be lexicalized becomes too great, the model constructs an abstract UR for the morpheme. This abstract UR introduces a discrepancy between the abstract UR and its surface realization. The model thus constructs a set of (UR, SR) pairs from the lexicon, which it passes to a model that learns a phonological process to derive the various surface forms.

For example, when the 9th item from Tab. 1 causes /lɑr/ to no-longer be sustainable as the PL affix UR, the lexicon is as described in Tab. 3. The surface form for the PL forms of the roots 'ice', 'girl', and 'stalk' are computed by concatenating /lɑr/ to the stem (i.e., Stem-PL), and the remaining six known surface forms, which were lexicalized, are extracted directly from the lexicon. Since the PL is being collapsed into /lAr/, each word's UR is computed by replacing the surface realization of the PL affix with this new UR. Thus, the (UR, SR) pairs at this point would be {(/buzlAr/, [buzlɑr]), (/kɯzlAr/, [kɯzlɑr]), (/sɑplAr/, [sɑplɑr]), (/ellAr/, [eller]), (/søzlAr/, [søzler]), (/jyzyn/, [jyzyn]), (/jerlArin/, [jerlerin]), (/dɑllArɯn/, [dɑllɑrɯn]), (/iplArin/, [iplerin])}.

Learning phonological processes from UR-SR pairs is an active area of study, and many models have been proposed (see Jarosz 2019 for an overview). In this work we chose Belth (2023)'s model, which is a cognitively-grounded model that provides a unified ability to learn local and non-local alternations, which is important, given Turkish's non-local vowel harmony combined with local processes like voicing assimilation (see § 3.1).

Belth (2023)'s model is grounded in humans' strong tendency to track adjacent dependencies. For example, artificial language experiments have repeatedly demonstrated that learners more easily learn local phonological processes than non-local ones (Baer-Henney and van de Vijver, 2012) and, when multiple possible phonological generalizations are consistent with exposure data, learners systematically construct the most local generalization (Finley, 2011; White et al., 2018; McMullin and Hansson, 2019).

The Belth (2023) model learns rules to predict the surface form of alternating segments—in this case those that are underlyingly abstract. To do so, the model tracks only dependencies between alternating segments and the segments adjacent to them. If these adjacent segments fail to allow the surface form to be accurately predicted, the model deletes any adjacent segments that prevent the surface form from being predicted, and repeats. The iteratively deleted segments accumulate into a deletion set, the complement of which is interpreted as a tier. The learned rules are applied locally over the tier projection. Because segments are deleted only when adjacent dependencies fail to make the surface form predictable, local processes are a special case, and thus local and non-local processes are learnable by a unified model.

3 Evaluation

This section provides a case study of our proposed model on the highly agglutinative language, Turkish. In § 3.1 we describe some relevant details of Turkish. We then describe the setup of our evaluation in § 3.2. Finally, we present qualitative results in § 3.3 and quantitative results in § 3.4.

3.1 Turkish

Turkish phonology receives attention often because of its apparently complex vowel harmony system. It exhibits both primary front/back harmony and secondary rounding harmony, which is parasitic on height: only [+high] vowels harmonize for roundness. Moreover, Turkish has a number of exceptional suffixes whose vowels do not participate in harmony, and even half-harmonizing suffixes, which have multiple vowels, some of which harmonize and some of which do not. These harmony processes occur alongside other processes, such as local voicing assimilation. The Turkish vowel inventory is shown in (5).

		front		back	
		unround	round	unround	round
(5)	high	i	y	ɯ	u
	low	e	ø	ɑ	o

The primary harmony process is observed in affix vowels that alternate between [+back] when the preceding vowel is [+back] and [−back] when it is [−back], as in (6) (examples from Nevins 2010, p. 28; Kabak 2011, p. 3).

(6)	[dɑ̲l-lɑ̲r-ɯ̲n]	branch-PL-GEN
	[je̲r-le̲r-i̲n]	place-PL-GEN
	[i̲p-le̲r-i̲n]	rope-PL-GEN

The secondary rounding harmony involves suffixal [+high] vowels matching in roundness to the vowel to the left, as in (7) (examples from Nevins 2010, p. 29; Kabak 2011, p. 3). This harmony occurs regardless of whether the vowel to the left is [+high] (7a) or [−high] (7b).

(7)	a.	[i̲p-i̲n]	rope-GEN
		[jy̲z-y̲n]	face-GEN
		[kɯ̲z-ɯ̲n]	girl-GEN
		[bu̲z-u̲n]	ice-GEN
	b.	[e̲l-i̲n]	hand-GEN
		[sø̲z-y̲n]	word-GEN
		[sɑ̲p-ɯ̲n]	stalk-GEN
		[jo̲l-u̲n]	road-GEN

Some suffixes have vowels that do not participate in harmony. For example, the suffix [-ki] can attach to a temporal or spatial nominal root to yield adjectival forms as in (8), where the suffix surfaces with the vowel [i] regardless of the final vowel of the stem (examples from Oflazer 1994, p. 144). The PL suffix, which alternated in (6), here harmonizes with the [i] vowel (8b), thus surfacing as [e].

(8)	a.	[ønʤe-ki]	'(the one) before'
		[jɑrɯn-ki]	'(the one) tomorrow'
	b.	[ønʤe-ki̲-le̲r]	'(the ones) before'
		[jɑrɯn-ki̲-le̲r]	'(the ones) tomorrow'

The situation gets more complex, as some suffixes are *half harmonizing*, meaning they have two vowels with one harmonizing and one not.[7] An example is shown in (9a), where the first vowel of the abilitative (ABIL) suffix harmonizes with the vowel to the left, but the second vowel is always [−back] [i] even when the first vowel is [+back] (Kornfilt, 2013). The aorist (AOR) suffix vowel then harmonizes with the abilitative's non-harmonizing second vowel [i] in (9a). The example (9b) demonstrates that the AOR suffixal vowel surfacing as [i] in (9a) is indeed due to harmony, as it harmonizes in (9b) with [o].

(9)	a.	[jɑ̲z-ɑ̲bil-ir]	'write'-ABIL-AOR
		[jy̲z-e̲bil-ir]	'swim'-ABIL-AOR
	b.	[o̲l-u̲r]	'become'-AOR

Vowel harmony often goes in hand with other phonological processes, such as voicing assimilation. This can be seen, for example, in the locative (LOC) suffix, which exhibits vowel harmony, but begins with an alveolar stop, which assimilates in voicing to the segment to its left, as in (10) (examples from Dobrovolsky 1982; Çöltekin 2010; Kornfilt 2013).

(10)	[byro̲-d̲ɑ]	'office'-LOC
	[ev̲-d̲e]	'house'-LOC
	[ʤep̲-t̲e]	'pocket'-LOC

In the remaining subsections, we demonstrate how our proposed model elegantly accounts for these complexities in Turkish (§ 3.3), and how this allows for novel surface forms to be accurately predicted (§ 3.4). First, though, we introduce the setup and data we used for our experiments (§ 3.2).

3.2 Setup and Data

To simulate learning in Turkish, we constructed two Turkish datasets consisting of frequency-annotated and morphologically-analyzed surface forms (see below). To simulate one learning trajectory, we sampled words with replacement from the corpus, weighted by frequency. Each time a new word form is sampled, the learner adds it to its lexicon. We then investigate the underlying forms of each morpheme, seeing which are concrete and which are abstract (§ 3.3). We then evaluate how accurately the model, combined with a model for learning alternation rules, allows novel surface forms to be predicted (§ 3.4).

We constructed two datasets, called MorphoChallenge and CHILDES. The first used data from MorphoChallenge (Kurimo et al., 2010), which contains a large Turkish corpus annotated

[7] The term *half harmonizing* is from Nevins (2010), but one reviewer pointed out that, in principle, other fractions of vowels (1 of 3) could harmonize.

with word frequencies. To generate morphological analyses of words, we used Çöltekin (2010, 2014)'s finite state morphological analyzer, which is designed for Turkish. This is similar to the process used in the MorphoChallenge, but is publicly available.[8] We dropped any word in MorphoChallenge that had fewer than 25 occurrences or for which the morphological analyzer failed to provide an analysis. We also removed forms with affixes that are analyzed by Çöltekin (2010, 2014) as having multiple underlying forms. For example, the highly irregular aorist suffix is sometimes described as having four underlying forms: /-Ar/, /-Hr, /-z/, /-null/. Future work will consider scenarios where multiple URs are necessary. This resulted in 22,315 frequency-annotated and morphologically-analyzed surface forms, which we transcribed into IPA.

The second dataset is derived from the child-directed speech in the Aksu (Slobin, 1982) and Altinkamis corpuses of the CHILDES database (MacWhinney, 2000). We computed the frequency of each word in the corpuses and used the same process as above to morphologically analyze each word. This dataset is much smaller, so we did not exclude words with low corpus counts from this dataset. The resulted in 1,727 frequency-annotated and morphologically-analyzed surface forms, transcribed into IPA.

Note that some Turkish suffixes exhibit deletion/epenthesis to avoid CC or VV clusters. These additional processes are at present ignored, because the implementation provided by Belth (2023) was designed for harmony and disharmony. Future work will extend the implementation to epenthesis and deletion by incorporating Belth (In Press)'s model, which handles such processes.

3.3 Suffixes: Abstract and Concrete

Remarkably, the apparent complexity of Turkish vowel harmony, discussed in § 3.1, vanishes when we investigate the output of our model.[9] As before, we will let A denote the Turkish low, unround vowel with backness unspecified (extensionally, {e, ɑ}) and H be the Turkish high vowel with both backness and height unspecified (extensionally, {i, y, ɯ, u}). Moreover, we will use D to denote the alveolar stop with voicing unspecified (extensionally, {d, t}).

[8]https://github.com/coltekin/TRmorph

[9]This analysis is performed on a random, frequency-weighted 80% sample of the MorphoChallenge dataset.

We will walk through the complexities exemplified by (6)-(10) one-by-one. First, the PL suffix in (6), which has a low unrounded vowel, participates in front/back harmony, but not rounding harmony because it is not a [+high] vowel. Our model constructed the underlying form /-lAr/ for this suffix, capturing the fact that it only harmonizes for backness.

The GEN suffix in (6)-(7) has a [+high] vowel and participates in both primary and secondary harmony. Our model constructed the underlying form /-Hn/ for this suffix, which captures the surface alternation of this morpheme.

Next, the [-ki] suffix in (8) does not participate in harmony, and our model consistently represents it with a concrete form /-ki/.

For the abilitative suffix in (9), our model abstracts the first, harmonizing vowel, but keeps the second, non-harmonizing vowel concrete -/Abil/.

Lastly, the UR for the locative suffix in (10) is constructed with both segments abstract /-DA/, capturing both the voicing assimilation of the initial alveolar stop and the vowel harmony of the second segment.

These underlying forms allow Belth (2023)'s model to learn two rules, which allow for the accurate prediction of novel surface forms. On the resulting (UR, SR) pairs, Belth (2023) learns a vowel harmony rule, which targets both /A/ and /H/ vowels, and enforces harmony with respect to their unspecified values: [back] for /A/ and both [back] and [round] for /H/. The model automatically constructs a vowel tier and enforces harmony locally over that tier (see Belth 2023 for details). Belth (2023)'s model also learns a local voice assimilation rule, which causes /D/ to take its [voice] value from the segment to its left.

It is worth noting that others—in particular Nevins (2010)—have similarly argued that Turkish vowel harmony can be elegantly accounted for with an underspecification approach. Our model builds on Nevins (2010)'s observations by providing an explicit computational model that constructs underlying forms, which turn out to be consistent with this analysis.

As a further analysis, we show the 10 most frequent affixes in a 1K word sample of the CHILDES corpus in Tab. 4, along with the UR that our model constructed for each. Of the 10 affixes, 7 have been collapsed into abstract forms. However, there are 3 forms (P1S, lH, P2S) that were quite frequent,

Affix	UR	Abstract
PL	/-lAr/	Y
P3S	/-H/	Y
P1S	/-m/	N
GEN	/-Hn/	Y
DAT	/-A/	Y
ACC	/-H/	Y
LOC	/-DA/	Y
VN:INF	/-mA/	Y
lH	/-lɯ/	N
P2S	/-n/	N

Table 4: Top 10 most frequent affixes in a random, frequency-weighted sample of 1K words from the CHILDES dataset, and the URs that our model learned. See http://coltekin.net/cagri/trmorph/trmorph-manual.pdf for a description of affix names.

but are still able to be stored concretely. The P1S and P2S affixes do not have alternating segments in Turkish, so it is expected that these would be concrete. The "lH" affix, as captured by its name, can surface with any high vowel. However, in the training data, the [-lɯ] form occurs 25 out of 32 times, so the 7 words where it surfaces as something else are lexicalized ($7 <= 32/\ln 32$).

3.4 Quantitative Evaluation

We also evaluated how the model enables generalization, when paired with a model for learning phonological alternations. We used our model in tandem with Belth (2023)'s model to learn to map a stem and morphological analysis of a surface form to an actual surface form. For example, given the stem [dɑl] and morphological analysis Stem-PL-GEN, our model's underlying forms for -PL and -GEN are concatenated to the stem to form a UR, to which the generalizations learned by Belth (2023) can then be applied to predict a surface form, such as [dɑllɑrɯn].

We ran the model on both datasets, simulating incremental learning by sampling words with replacement and weighted by frequency, and adding them to the lexicon when sampled. As this process incrementally adds words to the lexicon, our model operates as described in (2). In 250-word increments (i.e., every time the lexicon grows by 250 unique words), we evaluated the model by using the rules learned by Belth (2022)'s model—on our learned underlying forms—to predict the surface form of all the words not in the lexicon. We carried out 5 simulations on each dataset, using different random seeds for sampling on each.

The results are shown in Fig. 1, where the x-axis shows the incremental growth of the learner's lexicon (i.e., the training size), and the y-axis shows the accuracy at predicting novel surface forms at that point during training. The accuracy is computed over all surface forms not currently in the training data. Each subfigure is for one of the two datasets. The MorphoChallenge results (Fig. 1a) are reported up to a size of 3K words, so the test results are on 10s of thousands of novel words.

The model's performance appears to be consistent with acquisition studies. Altan (2009) found that Turkish-speaking children as young as 2;0 extend vowel harmony to nonce words. Studies across languages reveal that a child's vocabulary is quite modest at this age, with an upper bound around 1K words (Fenson et al., 1994; Hart and Risley, 1995; Szagun et al., 2006; Bornstein et al., 2004). The model's performance on both datasets is above 90% accuracy when its vocabulary contains 1K words.

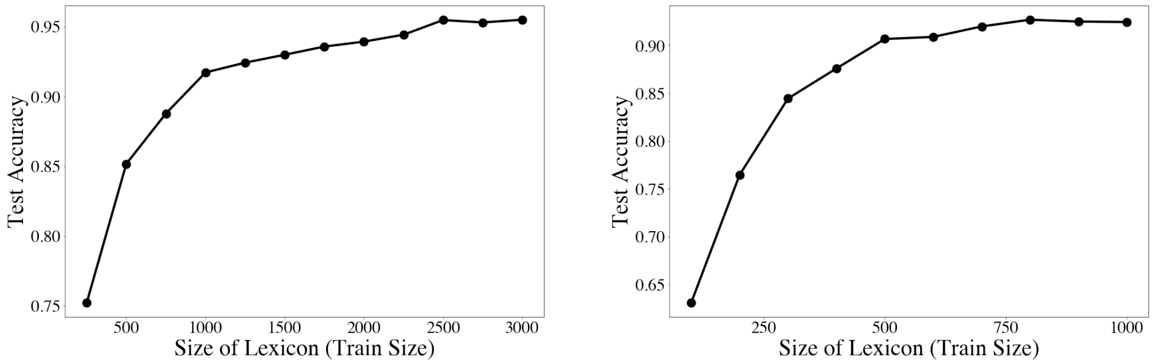
3.4.1 Error Analysis

Of the errors, around 52% result from the model having a concrete form of an affix, which it then errantly predicts for a novel word that exhibits alternation in that affix. For example, there are insufficient forms in the training data to make /ɯp/ as the concrete CV:IP affix prohibitive ($e = 5 <= n = 13/\ln 13$), even though vowel harmony leads it to sometimes surfaces with other high vowels. As a result, novel words like [gel-ip], which take the [ip] form of the affix are mispredicted.

About 47% of the errors are the result of vowel harmony or consonant assimilation being predicted for a novel form that exceptionally does not involve harmony. For example, the word [sɑɑt-ler] 'watch-PL' is predicted by our model to be [sɑɑt-lɑr] because the UR for the plural suffix is /lAr/, as it systematically harmonizes. According to a Wiktionary search,[10] the root [sɑɑt] is of Arabic origins. Because Arabic has a different vowel system, vowels in Arabic loan words may conform to the Turkish vowel system when entering Turkish, and thus sometimes behave oddly. Indeed, Altan (2009) observed that children may overextend vowel harmony to such words.

The remaining 1% of errors result from very low

[10] https://en.wiktionary.org/wiki/saat#Turkish



(a) MorphChallenge

(b) CHILDES

Figure 1: Our proposed model's accuracy (averaged over 5 simulations) at predicting novel surface forms. The x-axis shows the growth of the learner's lexicon (i.e., the training size).

frequency affixes which are simply unattested in the training data.

4 Prior Work

Tesar (2014) and Hua et al. (2020) focus on theoretical analyses of the nature of the problem of learning URs. O'Hara (2017); Rasin et al. (2018); Ellis et al. (2022) proposed computational models, but evaluate on small, phonology-textbook-like data, not large, natural-language corpora.

Cotterell et al. (2015) also predominately models textbook-like problems, but presents some limited analysis on more realistic corpora. However, these corpora only involve very simple morphological paradigms involving a single suffix, and present to the model a fairly curated subset of the corpus that isolates the relevant morphophonological process.

Richter (2021) studies the question of when allophonic surface segments are collapsed into an abstract underlying segment, focusing on the English flap [ɾ] allophone of /T/. While Richter (2021) focuses on allophones, our proposed model is inspired by it and can be viewed as extending the same principles to morphophonological alternations.

Of these prior models, we were only able to get access to code for Cotterell et al. (2015) and Rasin et al. (2018), which we were unable to get to run on our large datasets. In future versions of this work, we intend to implement some of these existing models in order to compare their performance and behavior to that of our proposed model.

5 Conclusion

This work proposed a learning-based account of underlying forms, taking the highly agglutinating language of Turkish as a case study. The proposed model starts with concrete underlying representations and constructs abstract URs only in cases where doing so helps to form generalizations that deal with the sparsity of morphological forms in the learner's input.

The model constructs abstract underlying forms when they are critical for generalization, but allows for concrete forms when abstraction is unecessary. This flexibility is at the core of the model's success, as evidenced by the fact that the representations of Turkish suffixes in § 3.3 are minimally abstract. For example, the half-harmonizing suffixes consist of concrete segments except for the single, harmonizing vowel. Similarly, exceptional, non-harmonizing suffixes remain fully concrete.

When combined with a model for learning local and non-local alternations, the proposed model achieves >95% accuracy predicting the surface form of held-out test words.

This work presents a preliminary case study in Turkish. Future work will evaluate the model on other languages. Moreover, the algorithm takes as input morphologically-segmented surface forms. As discussed in § 2.1, there is experimental evidence that children are able to perform morphological segmentation. In future work, we will attempt to bring the problems together, jointly segmenting surface forms, learning underlying forms, and morphophonological grammars.

References

Asli Altan. 2009. Acquisition of vowel harmony in Turkish. *Dilbilim 35. Yıl Yazıları*, pages 9–26.

Stephen R. Anderson. 2021. *Phonology in the Twentieth Century*. Number 5 in History and Philosophy of the Language Sciences. Language Science Press, Berlin.

Dinah Baer-Henney and 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.

Caleb Belth. 2022. A learning-based account of phonological tiers. *Linguistic Inquiry*. Under Review.

Caleb Belth. 2023. Learning non-local phonological alternations via automatic creation of tiers. In *Oral Presentation at the 97th Annual Meeting of the LSA*.

Caleb Belth. In Press. A learning-based account of local phonological processes. *Phonology*.

Caleb Belth, Sarah Payne, Deniz Beser, Jordan Kodner, and Charles Yang. 2021. The greedy and recursive search for morphological productivity. In *CogSci*.

Paul Boersma. 1997. How we learn variation, optionality, and probability. In *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, volume 21, pages 43–58. Citeseer.

Marc H Bornstein, Linda R Cote, Sharone Maital, Kathleen Painter, Sung-Yun Park, Liliana Pascual, Marie-Germaine Pêcheux, Josette Ruel, Paola Venuti, and Andre Vyt. 2004. Cross-linguistic analysis of vocabulary in young children: Spanish, dutch, french, hebrew, italian, korean, and american english. *Child development*, 75(4):1115–1139.

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

Çagri Çöltekin. 2010. A freely available morphological analyzer for turkish. In *LREC*, volume 2, pages 19–28.

Çagri Çöltekin. 2014. A set of open source tools for turkish natural language processing. In *LREC*, pages 1079–1086.

Ryan Cotterell, Nanyun Peng, and Jason Eisner. 2015. Modeling word forms using latent underlying morphs and phonology. *Transactions of the Association for Computational Linguistics*, 3(0):433–447.

Michael Dobrovolsky. 1982. Some thoughts on turkish voicing assimilation. In *Calgary Working Papers in Linguistics*, volume 7, pages 1–6.

Kevin Ellis, Adam Albright, Armando Solar-Lezama, Joshua B Tenenbaum, and Timothy J O'Donnell. 2022. Synthesizing theories of human language with bayesian program induction. *Nature communications*, 13(1):5024.

Emeryse Emond and Rushen Shi. 2021. Infants' rule generalization is governed by the Tolerance Principle. In *Proceedings of the 45nd annual Boston University Conference on Language Development*, pages 191–204.

Marc Ettlinger. 2008. *Input-driven opacity*. University of California, Berkeley.

Larry Fenson, Philip S Dale, J Steven Reznick, Elizabeth Bates, Donna J Thal, Stephen J Pethick, Michael Tomasello, Carolyn B Mervis, and Joan Stiles. 1994. Variability in early communicative development. *Monographs of the Society for Research in Child Development*, pages i–185.

Sara Finley. 2011. The privileged status of locality in consonant harmony. *Journal of memory and language*, 65(1):74–83.

Betty Hart and Todd R Risley. 1995. *Meaningful differences in the everyday experience of young American children.* Paul H Brookes Publishing, Baltimore, MD.

Bruce Hayes and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic inquiry*, 39(3):379–440.

Wenyue Hua, Adam Jardine, and Huteng Dai. 2020. Learning underlying representations and input-strictly-local functions. In *Proceedings of the 37th West Coast Conference on Formal Linguistics*.

Larry M Hyman. 2018. Why underlying representations? *Journal of Linguistics*, 54(3):591–610.

Gaja Jarosz. 2019. Computational modeling of phonological learning. *Annual Review of Linguistics*.

Bariş Kabak. 2011. Turkish vowel harmony. *The Blackwell companion to phonology*, pages 1–24.

Paul Kiparsky. 1968. *How abstract is phonology?* Indiana University Linguistics Club.

Jaklin Kornfilt. 2013. *Turkish*. Routledge.

Elena Koulaguina and Rushen Shi. 2019. Rule generalization from inconsistent input in early infancy. *Language Acquisition*, 26(4):416–435.

Mikko Kurimo, Sami Virpioja, Ville Turunen, and Krista Lagus. 2010. Morpho challenge 2005-2010: Evaluations and results. In *Proceedings of the 11th Meeting of the ACL Special Interest Group on Computational Morphology and Phonology*, pages 87–95. ACL.

Enikő Ladányi, Ágnes M Kovács, and Judit Gervain. 2020. How 15-month-old infants process morphologically complex forms in an agglutinative language? *Infancy*, 25(2):190–204.

Brian MacWhinney. 1978. *The acquisition of morphophonology*. Monographs of the Society for Research in Child Development. University of Chicago Press.

Brian MacWhinney. 2000. *The CHILDES Project: Tools for analyzing talk. transcription format and programs*, volume 1. Psychology Press.

Alexandra Marquis and Rushen Shi. 2012. Initial morphological learning in preverbal infants. *Cognition*, 122(1):61–66.

Kevin McMullin and Gunnar Ólafur Hansson. 2019. Inductive learning of locality relations in segmental phonology. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 10(1).

Toben H Mintz. 2013. The segmentation of sub-lexical morphemes in english-learning 15-month-olds. *Frontiers in psychology*, 4:24.

Andrew Nevins. 2010. *Locality in vowel harmony*, volume 55. Mit Press.

Kemal Oflazer. 1994. Two-level description of turkish morphology. *Literary and linguistic computing*, 9(2):137–148.

Charlie O'Hara. 2017. How abstract is more abstract? learning abstract underlying representations. *Phonology*, 34(2):325–345.

S Payne. 2022. *When collisions are a good thing: the acquisition of morphological marking*. Bachelor's thesis, University of Pennsylvania.

Ezer Rasin, Iddo Berger, Nur Lan, and Roni Katzir. 2018. Learning phonological optionality and opacity from distributional evidence. In *Proceedings of NELS*, volume 48, pages 269–282.

Caitlin Richter. 2018. Learning allophones: What input is necessary. In *Proceedings of the 42nd annual Boston University Conference on Language Development.* Cascadilla Press.

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

Kathryn D Schuler, Charles Yang, and Elissa L Newport. 2016. Testing the tolerance principle: Children form productive rules when it is more computationally efficient to do so. In *CogSci*, volume 38, pages 2321–2326.

Dan I Slobin. 1982. Universal and particular in the acquisition of language. *Language acquisition: The state of the art*, 57.

Gisela Szagun, Claudia Steinbrink, Melanie Franik, and Barbara Stumper. 2006. Development of vocabulary and grammar in young german-speaking children assessed with a german language development inventory. *First Language*, 26(3):259–280.

Bruce Tesar. 2014. *Output-driven phonology: Theory and learning*. 139. Cambridge University Press.

Bruce Tesar and Paul Smolensky. 1998. Learnability in optimality theory. *Linguistic Inquiry*, 29(2):229–268.

James White, René Kager, Tal Linzen, Giorgos Markopoulos, Alexander Martin, Andrew Nevins, Sharon Peperkamp, Krisztina Polgárdi, Nina Topintzi, and Ruben van De Vijver. 2018. Preference for locality is affected by the prefix/suffix asymmetry: Evidence from artificial language learning. In *NELS*, pages 207–220.

Charles Yang. 2016. *The price of linguistic productivity: How children learn to break the rules of language*. MIT press.