

THE SURFEIT OF THE STIMULUS: ANALYTIC BIASES FILTER LEXICAL STATISTICS IN TURKISH LARYNGEAL ALTERNATIONS

MICHAEL BECKER

NIHAN KETREZ

ANDREW NEVINS

Harvard University İstanbul Bilgi University University College London

In an experimental task with novel words, we find that some lexical statistical regularities of Turkish phonotactics are productively extended in nonce words, while others are not. In particular, while laryngeal alternation rates in the lexicon can be predicted by the place of articulation of the stem-final stop, by word-length, and by the preceding vowel quality, this laryngeal alternation is only productively conditioned by place of articulation and word-length. Speakers' responses in a novel word task demonstrate that although they are attuned to the place of articulation and size effects, they ignore preceding vowels, even though the lexicon contains this information in abundance. We interpret this finding as evidence that speakers distinguish between phonologically motivated generalizations and accidental generalizations. We propose that universal grammar (UG), a set of analytic biases, acts as a filter on the generalizations that humans can make: UG contains information about possible and impossible interactions between phonological elements. Omnivorous statistical models that do not have information about possible interactions incorrectly reproduce accidental generalizations, thus failing to model speakers' behavior.*

Keywords: Turkish, laryngeal alternations, vowel-consonant interaction, naturalness, wug test, surfeit of the stimulus

1. INTRODUCTION. Learners and language users can and often do use statistical properties of linguistic input to discover hidden structure and make predictive generalizations about newly encountered items (e.g. Coleman & Pierrehumbert 1997, Bailey & Hahn 2001; see Saffran 2003, Hay & Baayen 2005, Chater & Manning 2006 for recent overviews). While these abilities to track statistical regularities in the input appear to be very powerful, at the same time they also appear to be constrained: some patterns are more readily detected and used than others. For example, Bonatti and colleagues (2005) found that adult learners exposed to artificial grammars were much better at extracting transitional probability regularities over consonants than equally matched transitional probabilities over vowels, suggesting that learners preferentially pay more attention to statistics within consonantal frames. In a study of infant learning of phonotactic patterns, Saffran and Thiessen (2003) showed that infants learned statistical patterns that grouped together /p/, /t/, /k/ (i.e. voiceless stops) as a class of items comprising the first sound in artificial word tokens much better than patterns that grouped /p/, /d/, /k/ as this

* For their valuable comments and suggestions on several versions of this work, we are greatly indebted to Adam Albright, Nicoleta Bateman, Kathryn Flack, Lyn Frazier, Bruce Hayes, William Idsardi, Karen Jesney, Shigeto Kawahara, Michael Key, John Kingston, Joe Pater, Kathryn Pruitt, John McCarthy, Harker Rhodes, Engin Sezer, Jesse Snedeker, Donca Steriade, Anne-Michelle Tessier, Bert Vaux, Andy Wedel, Matt Wolf, and Kie Zuraw. We would also like to thank the audience at the 2004 HUMDRUM, especially Luigi Burzio, Jane Grimshaw, Alan Prince, and Ellen Woolford, the audience at the Yale 2006 Turkish Linguistics Workshop, especially İlhan Çağrı, Atakan İnce, and Jaklin Kornfilt, the audience at the 81st LSA annual meeting in Anaheim, CA, especially Sharon Inkelas and Eduardo Gil, the Harvard phonotactics working group, especially Iris Berent, Peter Graff, and Tim O'Donnell, and the audience at the University of Toronto linguistics department, especially Peter Avery, Elan Dresher, Yoonjung Kang, Alexei Kochetov, and Naomi Nagy. Thanks to the many speakers of Turkish who helped with various aspects of this project, including Mine Doğucu, Anıl and Alper Duman, Beste Kamali, Süleyman Ulutaş, Leyla Zidani-Eroğlu, several students at Boğaziçi University, and the many kind people who filled out our questionnaires. Special thanks go the editor, associate editors, and anonymous referees at *Language*, who helped improve this article in a number of ways. Any errors would be ours.

class, again suggesting that statistical learning may be less efficient when the regularities are inconsistent with natural language structure.

In this article, we examine a number of predictive statistical phonotactic regularities found within the Turkish lexicon, some natural and some unnatural from the point of view of phonological typology, and examine whether they are all kept track of and used to an equal extent in on-line judgment tasks involving novel words. By examining whether adult speakers of a language with robust statistical regularities will detect and extend the use of unnatural patterns in generalization tasks, we can provide potential evidence for the role of analytic biases as active filters on extraction of sublexical statistics.

Laryngeal alternations in Turkish are observed at the right edges of nouns, as in 1. Nouns that end in a voiceless aspirated stop in their bare form, such as the prepalatal stop [tʃ^h], can either retain that [tʃ^h] in the possessive (1a,b), or the [tʃ^h] of the bare stem may alternate with the voiced [dʒ] in the possessive (1c,d).¹

- | | | | |
|-----|-----------------------|-----------------------|--------------|
| (1) | bare stem | possessive | |
| | a. atʃ ^h | atʃ ^h -i | ‘hunger’ |
| | b. anatʃ ^h | anatʃ ^h -i | ‘female cub’ |
| | c. gyʃ ^h | gydʒ-y | ‘force’ |
| | d. amatʃ ^h | amadʒ-i | ‘target’ |

Turkish exhibits a contrast between the voiced stops [b, d, dʒ, g] and the voiceless aspirated stops [p^h, t^h, tʃ^h, k^h] in onset position, for example, *t^her* ‘sweat’ vs. *der* ‘give (aorist)’. In coda position, however, the contrast is lost, with stops appearing voiceless and aspirated through complete phonetic neutralization (Kopkallı 1993, Wilson 2003). This restriction on the distribution of voiced stops applies productively to loanwords, such as *rop^h* ‘dress’ < French *robe*. Voiced coda stops are allowed in the initial syllable of the word, for example, *ad* ‘name’ or *ab.la* ‘older sister’, and in a limited number of exceptional words.

When nouns that end in a voiceless stop are followed by a vowel-initial suffix, the final stop may surface with its voiced counterpart, for example, *dʒop^h* ‘club’ vs. the possessed form *dʒo.b-u* ‘club.3SG’; when followed by a consonant-initial suffix, however, the final stop remains in coda position and thus stays voiceless: *dʒop^h.-lar* ‘club.PL’. This alternation occurs in 54% of the nouns of the language (Inkelas et al. 2000) and applies productively to loanwords, for example, *gu.rup^h* vs. *gu.ru.b-u* ‘group.3SG’. For the remaining 46% of stop-final nouns, the stop is voiceless in all suffixed forms of the word, for example, *sop^h* ~ *so.p^h-u* ‘clan.3SG’ ~ *sop^h.-lar* ‘clan.PL’.

The velar stops [k^h, g] contrast in onset position, for example, *so.k^hak^h* ‘street’ vs. *ga.ga* ‘beak’. In word-final position, they neutralize to the voiceless stop [k^h]. While postconsonantal dorsals, as in *renk^h* ~ *reng-i* ‘color’, display the general process of laryngeal alternation, intervocalic velar stops undergo deletion rather than voicing; that is, when nouns ending in a POSTVOCALIC velar stop are followed by a vowel-initial suffix, the velar stop deletes, for example, *etek^h* ~ *ete-i* ‘skirt’ (Zimmer & Abbott 1978, Sezer 1981).² Since laryngeal alternation and deletion are in complementary distribution, depending on the segment that precedes the final dorsal, we treat the two processes as one. Additionally, as is shown below, whether a noun stem shows the k/∅ alternation is corre-

¹ Turkish orthography does not represent aspiration, since it is predictable from a combination of voicing and morphological structure. For a discussion of laryngeal features in Turkish, see §4.1.

² We focus on the laryngeal alternations and deletions that occur in derived environments, leaving aside morpheme-internal intervocalic stops. Thus, the root-medial dorsal stops in *sok^hak^h* and *gaga* are protected in their nonderived environment.

lated with the same type of lexical statistics as other stop consonant alternations, thereby justifying a unified treatment for the purpose of the current experimental inquiry.

The distinction between alternating and nonalternating stops is traditionally captured within generative phonology as the difference between an underlying voiced stem-final stop in the case of $\text{ɕop}^h \sim \text{ɕob-u}$ and an underlying voiceless stem-final stop in the case of $\text{sop}^h \sim \text{sop}^h\text{-u}$, with the underlying contrast being neutralized in word-final coda position (Lees 1961). While the difference between alternating and nonalternating nouns may be captured in a variety of alternate theoretical frameworks that do not incorporate underlying representations (e.g. via reference to identity relations among surface forms alone; Burzio 2002, Albright 2008a), it is clear that under any way of representing morphophonemic alternation, Turkish nouns fall into two distinct classes of words, one of which alternates and one of which does not.

Whether the final stop of a given noun will or will not alternate is unpredictable. The noun's size strongly correlates with its status, however: most monosyllabic nouns do not alternate, while most polysyllabic nouns do. Section §2 discusses several other factors that correlate with laryngeal alternations, and §3 shows that Turkish speakers use only a subset of the available factors: they use the noun's size and the place of articulation of the final stop, but they do not use the quality of the vowel that precedes the word-final stop. A back vowel before a word-final $[\text{t}^h]$, for instance, correlates with more alternations, but Turkish speakers seem to ignore this correlation. These language-specific patterns can be understood given a crosslinguistic perspective: typological observations commonly correlate the distribution of laryngeal features with a word's size and a consonant's place of articulation, but rarely or never with the quality of a neighboring vowel. Indeed, speakers are reluctant to learn patterns that correlate vowel height with the laryngeal features of a neighboring consonant (Moreton 2008).

From a crosslinguistic perspective, it is unsurprising that monosyllabic nouns would behave differently from polysyllabic nouns with respect to the laryngeal alternation. Initial syllables are often protected from markedness pressures, showing a wider range of contrasts and an immunity to alternations (Beckman 1998).³ Specifically in Turkish, the privileged status of the laryngeal features [voice] and [s.g.] (spread glottis) in initial syllables is not just seen in laryngeal alternations. Generally in the language, a coda obstruent followed by an onset obstruent will surface with the laryngeal features of the onset obstruent (e.g. *t̃.rab.zon* 'Trapezus' (place name), **t̃.rab.son*), but a coda obstruent in the initial syllable may surface with its independent laryngeal specification (e.g. *ab.sor.be* 'absorption' vs. *t̃ep̃.si* 'tray').

The backness of a neighboring vowel, however, is never seen to interact with a consonant's laryngeal features across languages. While such a connection is mildly phonetically plausible (vowel backness correlates with tongue-root position, which in turn correlates with voicing), there is no known report of any language where consonant laryngeal features change depending on the backness of a neighboring vowel, or vice versa. Given this gap in the universal inventory of possible phonological interactions, it is not surprising that in Turkish, speakers showed no sign of using vowel backness as a predictor of laryngeal alternations in the experiment we describe in §3.

³ Our account contrasts with that in Wedel 2002, Ussishkin & Wedel 2009, where it is suggested that the source of the size effect is in neighborhood density. This argument is effectively refuted by Pycha and colleagues (2007) and by Becker and Nevins (2009), who show that similarity-based measurements such as neighborhood density do not correlate with alternation rates.

In OPTIMALITY THEORY (OT; Prince & Smolensky 2004 [1993]), typological observations are encoded in the structure of the universal inventory of constraints (CON). The constraints and their interactions produce all and only the observed sound patterns of the world's languages. The preferred status of initial syllables is encoded with a set of faithfulness constraints specific to initial syllables. The lack of interaction between vowel backness and laryngeal features is encoded by the exclusion of constraints from CON that refer to some value of [\pm back] next to some value of [\pm voice] or [\pm s.g.], for example, *[+back][+voice]. In the absence of such constraints, there is never a reason to change one of these features in the presence of the other, and the lack of interaction is predicted. The account of the Turkish facts offered here capitalizes on these aspects of CON, while remaining agnostic about the mechanism that excludes these constraints, be it by assuming an innate set of constraints (as has been assumed since Prince & Smolensky 2004 [1993], and in the context of learning in Tesar & Smolensky 1998, 2000, Tesar 1998, Prince 2002, Hayes 2004, Jarosz 2006, Tesar & Prince 2006, among others), or by a mechanism of constraint induction (as in Flack 2007, Hayes & Wilson 2008, Moreton 2010).

Our analysis, as presented in §4, crucially depends on a grammatical encoding of the predictors of lexical trends, rather than direct access to the lexicon. A grammar that encodes phonological interactions will be sufficient for excluding interactions of vowel quality and laryngeal features, as this is a second-order interaction involving disjoint features.

Many possible grammatical architectures could be employed for this purpose, and our experimental results do not arbitrate among them. For the sake of concreteness, in this article, we implement such a grammar with a version of optimality theory where the pattern of individual lexical items is recorded in terms of lexically specific constraint rankings (cf. Itô & Mester 1995, Inkelas et al. 1997, Anttila 2002, Pater 2006, 2009, Coetzee 2008). A noun with a nonalternating final stop, like *anatf^h ~ anatf^h-i*, is associated with the ranking IDENT(lar) >> *VtjV, meaning that faithfulness to laryngeal features outweighs the markedness pressure against voiceless intervocalic palatal stops. A noun with a final alternating stop, like *amatf^h ~ amaɟ-i*, is associated with the opposite ranking, that is, *VtjV >> IDENT(lar). This assumes that the final stop in *amatf^h* is underlyingly voiceless and aspirated, and that it surfaces unfaithfully in *amaɟ-i*, contrary to the traditional generative analysis of Turkish (Lees 1961, Inkelas & Orgun 1995, Inkelas et al. 1997). This aspect of the analysis is discussed and motivated in §4.

Given this approach, the pattern of monosyllabic nouns, like *atf^h ~ aɟ-i*, can be recorded separately from the pattern of polysyllabic nouns, by using a faithfulness constraint that protects the laryngeal features of stops in the base's initial syllable, IDENT(lar)₀₁. The existence of constraints in CON that are specific to initial syllables allows Turkish speakers to learn separate lexical trends for monosyllabic and polysyllabic nouns. By contrast, in the absence of universal constraints that relate laryngeal features and vowel backness, the backness of the stem-final vowel cannot be used in recording the pattern of any lexical items, and this aspect of the lexicon goes ignored by speakers.

Speakers' ability to project trends from their lexicon onto novel items is a well-established observation (see Zuraw 2000, Albright et al. 2001, Ernestus & Baayen 2003, Hayes & Londe 2006, among many others). The novel observation offered here, that only universal trends are projected, also finds support in previous work, which we discuss in §5.

This article offers empirical data that bears on the relation between the projection of language-specific lexical trends and crosslinguistic patterns of phonological interac-

tions, by deriving both from the inventory of universal constraints, and uses a grammar to filter lexical trends from those items onto novel nouns. We discuss the more general consequences of our findings for phonological typology and phonological learning in §5.

2. A QUANTITATIVE STUDY OF PATTERNS IN THE TURKISH LEXICON. The distribution of laryngeal alternations in the lexicon of Turkish depends heavily on the phonological shape of nouns. For instance, while the final stop in most monosyllabic nouns does not alternate (2a), the final stop in most polysyllabic words does alternate with its voiced counterpart (2b). This section offers a detailed quantitative survey of the Turkish lexicon, using the Turkish Electronic Living Lexicon (TELL, Inkelas et al. 2000, <http://linguistics.berkeley.edu/TELL/>). Some nouns in TELL are listed as both alternators and nonalternators (2c), which we call ‘vacillators’. Our statistical analysis treats them as intermediate between alternators and nonalternators, although in reality it is possible that their actual rate of alternation is different from 50%.

- | | | | |
|-----|---------------------|----------------------------------|---------------|
| (2) | bare stem | possessive | |
| a. | atʃ ^h | atʃ ^h -i | ‘hunger’ |
| b. | amatʃ ^h | amaɖʒ-i | ‘target’ |
| c. | gyvetʃ ^h | gyvetʃ ^h -i, gyvedʒ-i | ‘cooking pot’ |

Several phonological properties of Turkish nouns are discussed below, and we show that four of them correlate with stem-final alternations: (i) the noun’s size (monosyllabic vs. polysyllabic), (ii) the place of articulation of the stem-final stop, (iii) the height of the vowel that precedes the stem-final stop, and (iv) the backness of that vowel.

TELL lists a total of about 30,000 nouns, verbs, and adjectives that were collected from a variety of extant dictionaries. Of these, 18,000 were produced and transcribed by a native speaker in various inflected forms. Nouns are listed with their bare citation forms and with four suffixed forms (1SG possessive, accusative, professional, and 1SG predicative).

2.1. EXPLORATION OF LARYNGEAL ALTERNATION PREDICTORS IN TELL. In Figure 1, we offer a closer look at the 3,002 nouns in TELL whose bare stems end in a voiceless stop. Figure 1a shows the distribution of alternation by the PLACE of articulation of the final stop. Most word-final labials, palatals, and dorsals alternate,⁴ but only a small proportion of the final coronals do. This is represented in the mosaic plots, where alternation status is plotted against place of articulation, word length, preceding vowel height, and preceding vowel backness. Each of these plots shows the 3,002 stop-final nouns of TELL.

Figure 1b shows the effect of phonological SIZE: while ~60% of polysyllables (which we mark as CVCVC) alternate, most monosyllables do not. Monosyllables with a simplex coda (CVC) are even less likely to alternate than monosyllables with a complex coda (CVCC). Figures 1c,d show that a stem-final stop is more likely to alternate when the stem’s final syllable contains a HIGH vowel or a BACK vowel. These last two correlations are rather surprising, since crosslinguistically, vowel quality is not known to influence the laryngeal features of a neighboring obstruent.

The interaction of size and place is plotted in Figure 2a. In all places of articulation, CVC nouns alternate less than CVCVC nouns, but the pattern of CVCC words is not uniform. For labials and palatals, a majority of CVCC words alternate, patterning with the CVCVC words. For the dorsals, the CVCC words pattern together with the shorter CVC words, showing a modest proportion of alternators. Finally, the coronals show a

⁴ Dorsals delete postvocally, and voice postconsonantly; see §1 above for discussion.

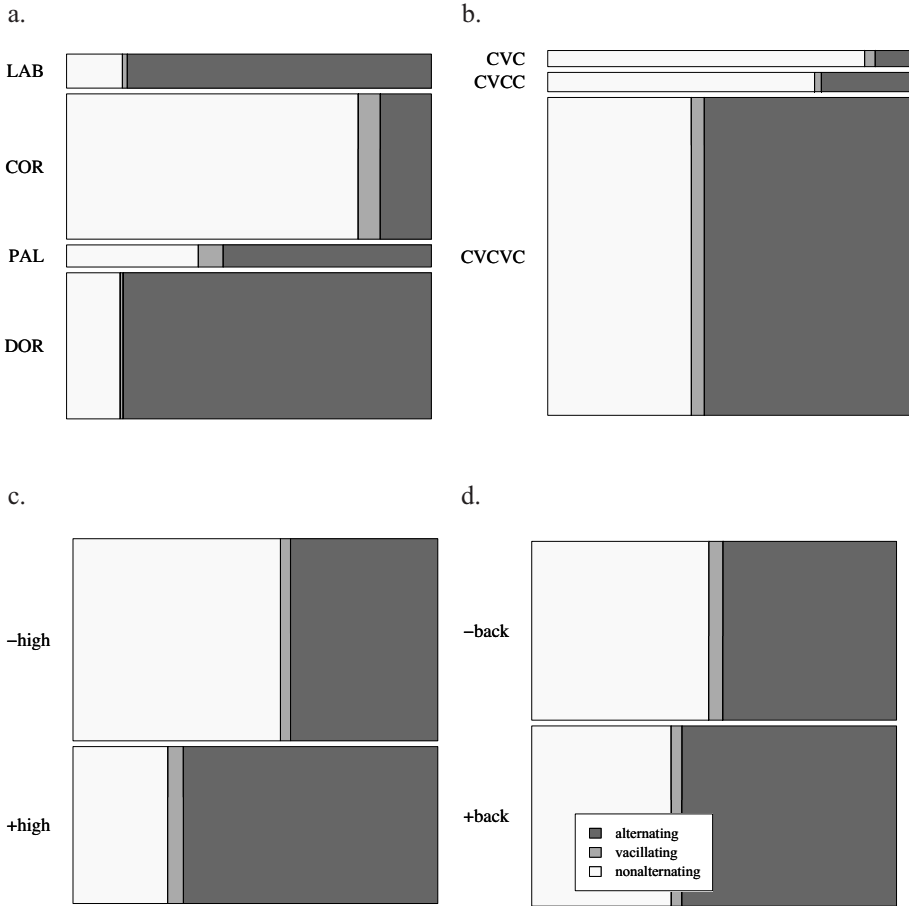


FIGURE 1. Alternation rates in the lexicon, by single features.

very minor size effect, with CVCC words actually having a slightly higher proportion of alternators than either longer or shorter words. Summarizing these interactions, it is not the case that size and place each have a constant effect. Their effect on the distribution of laryngeal alternations cannot be accurately described separately. Anticipating the discussion in §3.3, it is seen that indeed speakers treat each place/size combination separately.

Figure 2b shows that the effect of a back vowel is very small on coronals and dorsals, but it is strong for the palatals, where the proportion of alternating nouns is 30% higher following a back vowel relative to a front vowel. Figures 2c,d show the effect of high vowels: the height effect is concentrated in the coronals and dorsals, and in the polysyllables.

Focusing on one of the three-way interactions, we show in Figure 3 a mosaic plot for the interaction of place, size, and height. It can be seen that the strongest height effect is within the *t*-final polysyllables.

In contrast to the four properties that were examined until now (place, size, high, and back), a phonological property that has but a negligible correlation with the distribution of laryngeal alternations is the rounding of the stem's final vowel: a stem-final stop is merely 2% more likely to alternate following a round vowel compared to a nonround vowel.

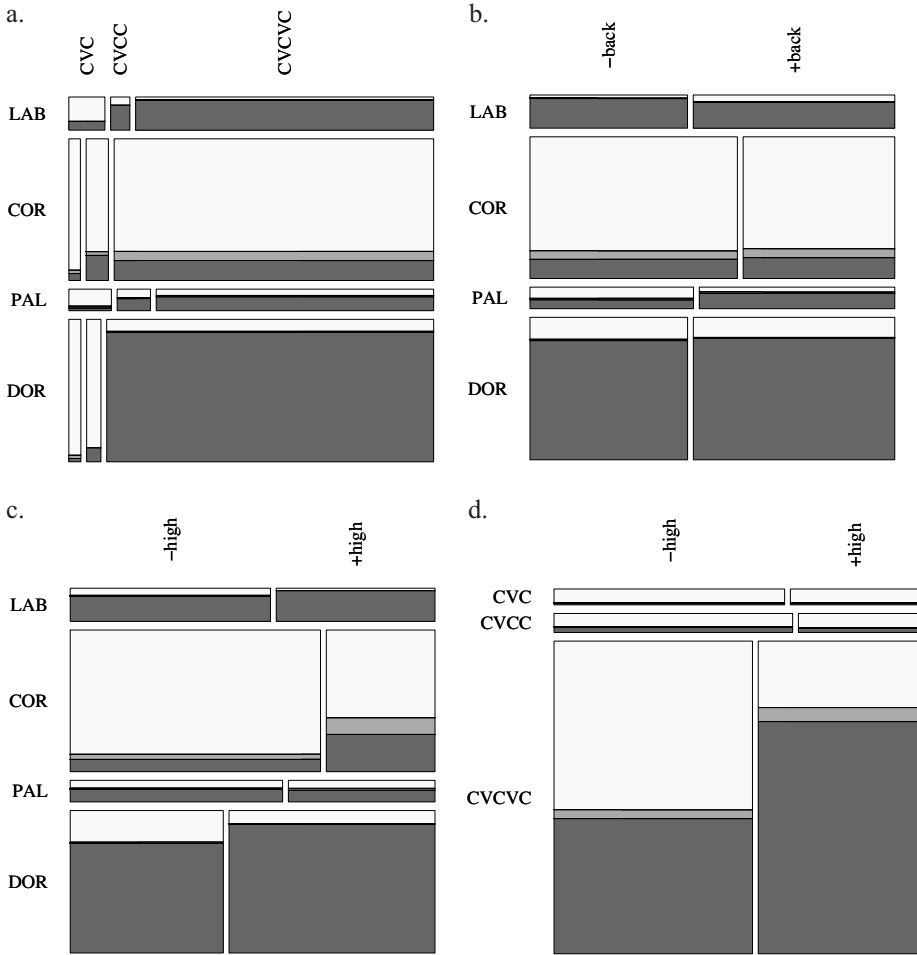


FIGURE 2. Alternation rates in the lexicon, by pairs of features.

A closer examination of vowel rounding is no more revealing, and the details are omitted here for lack of interest. Other phonological properties that were checked and found to be equally unrevealing are the laryngeal features of consonants earlier in the word, such as the closest consonant to the root-final stop, the closest onset consonant, and the closest obstruent.

To sum up the discussion so far, four phonological properties of Turkish nouns were seen to correlate with stem-final laryngeal alternations in Turkish:

- Size: monosyllables alternate less than polysyllables, and among the monosyllables, roots with simplex codas alternate less than roots with complex codas.
- Place (of articulation): stem-final coronals alternate the least, while labials and dorsals alternate the most.
- Vowel height: stem-final stops are more likely to alternate following a high vowel compared to a nonhigh vowel.
- Vowel backness: stem-final stops are more likely to alternate following a back vowel compared to a front vowel.

All of these properties allow greater insight when considered in pairs: size and place have a nonuniform interaction, with CVCC words behaving like CVC words when dor-

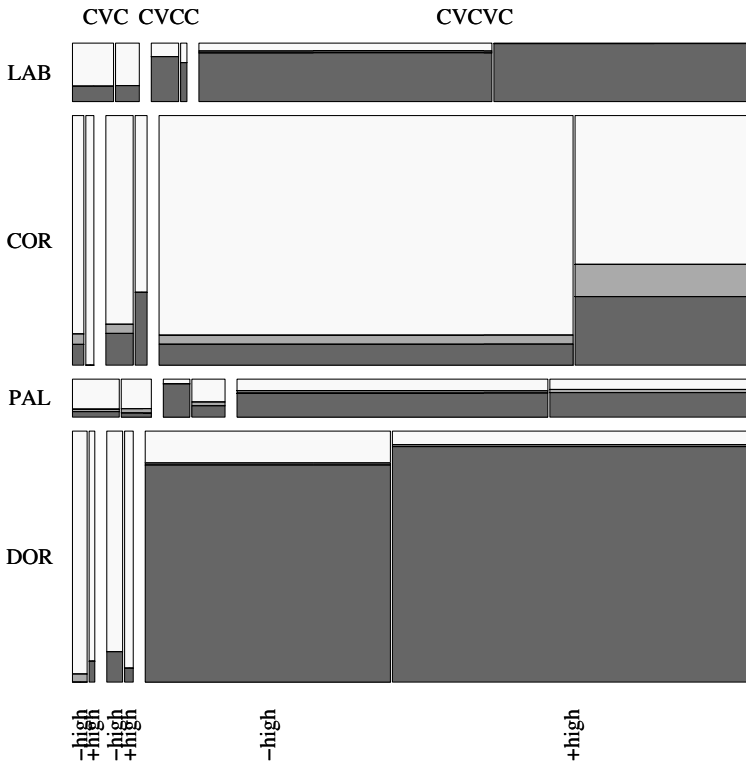


FIGURE 3. Alternation rates in the lexicon: place, size, high.

sal-final and like CVCVC words when labial- or palatal-final. Height and backness interact with place nonuniformly: the correlation with height is concentrated in the coronal-final nouns, while the correlation with backness is concentrated in the palatal-final nouns.

2.2. STATISTICAL ANALYSIS OF TELL. In statistical parlance, the aforementioned properties can be understood as predictors in a regression analysis. Since TELL makes a three-way distinction in stop-final nouns (nouns that do not alternate, nouns that do, and ‘vacillators’, that is, nouns that allow either alternation or nonalternation), an ordinal logistic regression model was fitted to the lexicon using the *lrm()* function in R (R Development Core Team 2007). The dependent variable was a three-level ordered factor, with nonalternation as the lowest level, alternation as the highest level, and vacillation as the intermediate level.⁵

Five independent variables were considered:

- Size: a three-level unordered factor, with levels corresponding to monosyllables with a simplex coda (CVC), monosyllables with a complex coda (CVCC), and polysyllables (CVCVC). CVC was chosen as the base level.
- Place: a four-level unordered factor, with levels corresponding to coronal, palatal, labial, and dorsal. Dorsal was chosen as the base level.

⁵ A referee requested that we run the analysis as a simple two-level logistic regression by assigning half of the vacillators to the voiceless category and the other half to the alternating category. We performed such an analysis, with the results staying almost entirely identical in the absolute values of coefficients and *p*-values, and entirely identical in relative values. The three-level model that we use has a higher model likelihood ratio.

- High, back, and round: each of the three features of the stem-final vowel was encoded as a two-level unordered factor. The base levels chosen were nonhigh, front, and unrounded.

First, each of these five predictors was assessed in its own model, to measure each predictor's overall power in the lexicon (Table 1). This power is measured by Somers's D_{xy} and by the model's likelihood ratio (model LR), which comes with a number of degrees of freedom and a p -value (Baayen 2008:203ff.). It turns out that *place*, *high*, *size*,⁶ and *back* are highly predictive of alternations, in that order, and *round* is not.⁷

	D_{xy}	MODEL LR	df	p
place	.66	1469	3	< 0.001
high	.29	284	1	< 0.001
size	.14	193	2	< 0.001
back	.11	37	1	< 0.001
round	0	0	1	> 0.1

TABLE 1. Strength of individual predictors in TELL.

While *high* has a larger D_{xy} than *size*, the interaction of *high* and *place* is less powerful than the interaction of *size* and *place*. The interactions of *place* with each of *size*, *high*, and *back* were tested in separate models, summarized in Table 2.

	D_{xy}	MODEL LR	df	p
place* <i>size</i>	.74	1920	11	< 0.001
place* <i>high</i>	.73	1621	7	< 0.001
place* <i>back</i>	.67	1496	7	< 0.001

TABLE 2. Strength of two-way interactions in TELL.

When a base model that has *place*size* as a predictor is augmented with *place*high*, D_{xy} goes up to .80. Augmenting the base model with *place*back* only brings D_{xy} up to .77. Finally, a model with all three of the interactions in Table 2 as predictors reaches a D_{xy} of .81, with a model LR of 2078 for nineteen degrees of freedom. This final model is given in Table 3. Each of the intermediate models that led to the model in Table 3 was compared with the next model using sequential ANOVA model comparison, and each step was found to be highly significant.⁸

⁶ We also considered a less linguistically informed size variable that was a simple raw count of the syllables of the stem. This variable was less informative than our size variable, producing a D_{xy} of merely .03 and higher p values, so we have excluded it from the following discussion. One reason raw size is less informative is that alternation rates do not keep going up as the word gets longer, but rather peak with di- and tri-syllables at 64% and 61% respectively, then go down to 40% and 41% for the tetra- and penta-syllables. The difference between the di- and tri-syllables is not significant generally, and only barely reaches significance for the labials ($p = 0.03$). The difference between the tri- and tetra-syllables is significant only without *place* factored in—once the place variable is added, the difference goes away. The vowel effects that we report below come out essentially the same with either size variable. An even more naive size measurement that simply counts segments does worse than either monosyllabicity or syllable count, reaching D_{xy} of merely .02.

⁷ Another method for assessing the predictive power of each feature separately is a TiMBL simulation (Daelemans et al. 2002). Given the data in TELL, this system creates a number called 'information gain' for every predictor that it is given. The system confirmed the verdict in Table 1, assigning the five predictors the following information-gain values, respectively: .367, .071, .047, .009, and .0004.

⁸ The model we report in Table 3 is not the most complex we found. We tried all the possible interactions exhaustively, and found that a model with terms *place*size + place*high*back* reaches slightly higher D_{xy} and model LR, and stands up to validation well. All the tests we report about the model given in Table 3 in the

The final model in Table 3 was validated with the fast backward step-down method of the *validate()* function, and the predictor *back* was the only one deleted. Since the interaction of *back* with *place* was retained, we did not remove *back* from the model, so as not to leave an interaction in the model without its components. In 200 bootstrap runs, seven factors were considered: the three interaction factors, and the four basic factors they were made of. At least five of the seven factors were retained in 197 of the runs, and in the vast majority of the runs, the three interaction factors were among the ones retained. The D_{xy} of the model was adjusted slightly from .81 to .80. An additional step of model criticism was taken with the *pentrace()* function, which penalizes large coefficients. With a penalty of .3, the penalized model was left essentially unchanged from the original model in Table 3, with slight improvements of the *p*-values of the vowel-place interactions at the fourth decimal place.

The model in Table 3 contains few surprises, as it confirms the validity of the observations made earlier in this section. It restates the numerical observations as differences in the propensity to alternate relative to the arbitrarily chosen baseline levels of the predictors, namely CVC size, dorsal place, nonhigh vowels, and front vowels. The size effect is mostly limited to the difference between monosyllables and polysyllables, with polysyllables being significantly more likely to alternate than CVC nouns. In the CVCVC size, the coronal and palatal places alternate significantly less than the baseline dorsal, while labials behave similarly to dorsals. Within the monosyllables, we see that CVCC nouns are more likely to alternate than CVC nouns, but this trend does not reach significance. We see in §3 below that speakers amplify this trend.

There is a main effect of *high* correlating with more alternations, and an additional interaction with coronal place. While there is no main effect of *back*, we see that back vowels correlate with significantly more alternations for the palatals.

	COEF β	SE(β)	WALD z	<i>p</i>
LAB	2.20	0.95	2.31	0.021
COR	-0.10	0.98	-0.10	0.917
PAL	1.25	0.95	1.31	0.189
CVCC	0.78	0.87	0.90	0.367
CVCVC	5.49	0.74	7.47	< 0.001
high	0.87	0.21	4.27	< 0.001
back	0.29	0.20	1.41	0.158
CVCC:LAB	2.02	1.16	1.75	0.081
CVCC:COR	0.70	1.10	0.64	0.523
CVCC:PAL	1.27	1.13	1.12	0.261
CVCVC:LAB	-1.74	0.90	-1.93	0.054
CVCVC:COR	-4.01	0.96	-4.18	< 0.001
CVCVC:PAL	-3.11	0.92	-3.38	< 0.001
LAB:high	0.53	0.54	0.99	0.323
COR:high	0.62	0.25	2.45	0.014
PAL:high	-0.75	0.39	-1.95	0.051
LAB:back	-0.76	0.49	-1.54	0.123
COR:back	0.08	0.25	0.30	0.762
PAL:back	1.17	0.39	2.95	0.003

TABLE 3. Final regression model for laryngeal alternations in TELL. Significant predictors are in boldface.

rest of this article apply equally well to the more complex model. We chose the simpler model because it is easier to interpret, and because we assessed the minuscule differences between the two models not to be worth the added complexity.

Our study assumes that TELL is a good model of the lexica of our speakers. The native speaker who supplied the judgments for TELL is about fifty years older than the average participant in our experiment, but they share a comparably high level of education and socioeconomic background. The *validate()* function that we applied to the model in Table 3 assures that the effects of the predictors are strong and reliable even in lexica that are different from TELL by as much as 37%. We conclude that we have little reason to doubt the usefulness of comparing the TELL data with data from educated younger speakers.

In sum, the quantitative analysis of the proportions of alternating nouns, in the form of a regression model, revealed four factors that are predictive of whether laryngeal alternation will occur: the phonological size of the word, the place of articulation, the height of the preceding vowel, and the backness of the preceding vowel. The first two of these have been previously identified as having an influence on laryngeal alternation in Turkish (Lewis 1967, Inkelas & Orgun 1995, Inkelas et al. 1997), and indeed these two, from a crosslinguistic perspective, are more likely than the other two to have a causal relationship with a stop's laryngeal features.

2.3. NONPHONOLOGICAL TRENDS IN TELL. While our discussion so far has focused exclusively on phonological properties of Turkish nouns, there are two more aspects of the Turkish lexicon to discuss: morphological structure, in particular the presence and distribution of affixes, and lexicon-based/usage-based factors, such as token frequency and neighborhood density.

Regarding morphological structure, we note that TELL includes many polymorphic words, which makes it conceivable that some of the observed trends are created or amplified by affixes, either synchronically or diachronically. For example, many *t*-final nonalternating nouns contain the native nominalizer *-It* or the Arabic feminine suffix *-At*.⁹ Similarly, the suffixes *-Ilk*, *-Ik*, and *-Ak* all yield alternating *k*-final nouns. It is possible, in principle, that the height effect is due to an abundance of nonalternating non-high *-At* and/or an abundance of alternating high *-Ilk*. This is not a concern for the backness effect, as the known *ʃ*-final suffixes all take their backness from the stem. To gauge the importance of morphological structure, we used a morphologically parsed wordlist prepared by Kemal Oflazer at Sabancı University (also used in Pycha et al. 2007), and from it extracted all 1,635 of the stop-final words that were identifiable as monomorphemic nouns. We then crossed this list with TELL, and examined the trends in the 966 items of the crossed list. The same statistical model we presented in Table 3 remained essentially unchanged even in this much smaller list, with the following minor differences: in the shorter list, *high* is no longer significant as a main effect, but its interaction with coronal place is, while *back* came out significant as a main effect in addition to the significant interaction with palatal. Both vowel effects made a significant improvement to a base model that only had *size* and *place*. We conclude that whatever effect affixes may have on the distribution of laryngeal alternations, the monomorphemic list we generated contains these effects just as strongly as TELL.

Turning to lexicon-based measures, we explored the role of token frequency, since all the discussion so far has relied on type frequency. We used Oflazer's corpus, which contains token-frequency information, to extract a list of 12,439 items. We crossed this list with TELL, and arrived at a crossed list of 1,659 items. We ran the same model from

⁹ Here, 'A' stands for a nonhigh vowel that receives its backness from the preceding syllable, while 'I' stands for a high vowel that receives its backness and rounding from the preceding syllable.

Table 3 on this shorter list, and it came out essentially unchanged (though again *high* was no longer a significant main effect, but remained significant in its interaction with coronal). We were now able to add (log) frequency as a predictor to this model, and it emerged highly significant: more frequent items are slightly, but consistently, more likely to be alternators. All of the other trends in the lexicon remained unchanged, however, meaning that even though token frequency does play a role in predicting alternations, this effect is independent from the grammatical effects that we focus on in this article. This is in line with findings in Bybee 1995, Albright & Hayes 2002, and Hay et al. 2004, which suggest that frequency can affect the behavior of individual items, but that overall trends are generally sensitive to the types in the lexicon, rather than being directly affected by token frequencies.

An additional lexicon-based measure we examined is neighborhood density. We counted the number of neighbors each lexical item in TELL has (calculated as in Luce & Pisoni 1998), and if there were any stop-final neighbors, we counted how many were alternators and how many were nonalternators. This yielded three neighborhood-based predictors that we added to separate models. All three predictors came out significant, with raw neighbor count and alternating neighbors correlating with more alternations, and nonalternating neighbors correlating with fewer alternations. We also added the various combinations of the three predictors to the model in Table 3, with the best improvements occurring in one model where the ‘total number of neighbors’ predictor was added, and in a second model where only the two ‘stop-final neighbor’ predictors were added with their interaction. In both of these models, all of our grammatical predictors remained undisturbed, with the exception of *high*, which was still highly significant as a main effect in both of the models we mentioned, but the interaction of *high* with coronal place lost its significance in the second one. In both models, the *back* effect remained as in Table 3.

To summarize, we note that the grammatical effects we are interested in are remarkably robust, and remain highly significant under a variety of manipulations, including serious reductions in the size of the dataset. The only effect that is somewhat less consistent than the others is vowel height, which stayed significant either as a main effect or as an interaction in all of the models that we checked, though not all had both. The backness effect and its interaction with palatal place remained significant in all of the models we examined. While lexicon-based and usage-based measures correlate strongly with laryngeal alternations, they are largely orthogonal to the grammatical measures we explore here.

2.4. SUMMARY OF LEXICAL TRENDS. One characterization of different types of phonotactics makes a distinction between first-order and second-order phonotactics (Warker & Dell 2006): first-order phonotactics regulate the distribution of a particular (set of) phonological feature(s) within a particular position in a syllable or word, whereas second-order phonotactics relate the distribution of a phonological feature in a particular position to some OTHER property of the syllable or word, such as a feature of a neighboring segment. While it is not the case that across the board, first-order phonotactics are more widespread than second-order (for example, vowel harmony is a second-order phonotactic), with respect to the case at hand, namely the distribution of laryngeal features in stops, it is generally the case that only first-order phonotactics matter.

In terms of our above findings on predictors of laryngeal alternations in the Turkish lexicon, Table 1 shows that *high*, an ‘unnatural’ predictor, shows a stronger effect than *size*, a natural predictor, and that both of these are second-order predictors.

The phonological size of a word, as measured here, is a proxy for a fact about the location of the potentially alternating stem-final stop: whether it occurs in the INITIAL SYLLABLE of the word or not. Indeed, as mentioned in the discussion of Turkish phonotactics above, one notorious locus of exceptions to otherwise persistent coda devoicing is in the coda of the initial syllable, as evidenced by words such as *ad* ‘name’ and *eb.k^hem* ‘mute’. This resistance to alternations in monosyllabic words is a result of the fact that in monosyllabic words, the stem-final syllable is the initial syllable. As a consequence, in a word such as *sop^h-u* ‘clan’ (as opposed to *gurub-u* ‘group’) the fact that the stop does not alternate is precisely because of a general resistance to alternations for segments in the initial syllable. Crosslinguistically, initial syllables enjoy greater faithfulness, or resistance to alternation (Beckman 1998). The *size* variable is thus a first-order phonotactic, as it relates the occurrence of particular features (voicing and aspiration) to a particular position in the word (the initial syllable).

The effect of the place of articulation on a stop that potentially undergoes alternation has crosslinguistic support as well. Different places are known to interact differently with laryngeal features (Lisker & Abramson 1964, Ohala 1983, Volatis & Miller 1992), and different relative proportions of alternation rates for different places of articulation were found by Ernestus and Baayen (2003) in their study of the Dutch lexicon. While the relative ranking of alternation rates across places of articulation may differ from language to language, it is a fact that languages exhibit phonotactics in manner features and laryngeal features that are gradient and differential specifically depending on place of articulation. The *place* variable is thus a first-order phonotactic, as it relates the occurrence of a particular set of features (voicing, aspiration, and place).

The effect within the Turkish lexicon of vowel quality (in particular, height and backness) on consonant laryngeal alternation is, by contrast, unexpected given crosslinguistic phonological typology. Interactions between vowel quality (height, backness, rounding) and the laryngeal features of consonants are infrequent, and the handful of documented cases show a causal influence in the opposite direction: the consonant’s laryngeal features can affect the height of a preceding vowel (Kingston 2002), but not vice versa. Consonant voicing and aspiration have been argued to affect vowel height in various languages—for example, in diphthong centralization before voiceless consonants in North American dialects of English, known as ‘Canadian raising’ (Chambers 1973, Moreton & Thomas 2007); in Polish (Gussmann 1980); in Madurese (Stevens 1968); and vowel backness in Northern Sarawak (Blust 2000)—but there is no documented case of a phonological process wherein vowel quality induces a change in consonant voicing or aspiration. Given the fact that interactions of vowel quality and consonantal laryngeal features are second-order phonotactics with little to no crosslinguistic attestation, their existence in Turkish is expected to be accidental rather than principled in nature.

These data therefore raise the question of whether Turkish speakers themselves will take the correlation between vowel quality and consonants’ laryngeal features to be accidental or whether they will take it to reflect an active generalization over their lexicon that they will reproduce. Given that all four of the factors *size*, *place*, *high*, and *back* are statistically reliable predictors of laryngeal alternations in the lexicon, we sought to determine whether speakers actually track and extend these patterns in experimental tasks with novel words.

To summarize the study of the Turkish lexicon, it was found that both *size* and *place* are excellent predictors of the alternation status of nouns. Longer nouns are more likely to alternate, and coronal-final nouns are less likely to alternate. In addition, the height

and backness of final stem vowels are also good predictors in combination with *place*: high vowels promote the alternation of coronals, and back vowels promote the alternation of palatals. All of these generalizations were confirmed to be highly statistically significant in a regression model that was put to several different tests. Put differently, the size of nouns, the place of their final stop, and the height and backness of their final vowels all strongly correlate with laryngeal alternations in a way that is statistically unlikely to be due to chance alone.

3. TESTING SPEAKERS' KNOWLEDGE OF LEXICAL TRENDS. In the previous section, the distribution of laryngeal alternations in the Turkish lexicon was examined and shown to be rather skewed. The distribution of alternating and nonalternating noun-final stops is not uniform relative to other phonological properties that nouns have: size, place, height, and backness were identified as statistically powerful predictors of alternation.

What native speakers of Turkish know about the distribution of laryngeal alternations, however, is a separate question, which is taken on in this section. It turns out that native speakers identify generalizations about the distribution of laryngeal alternations relative to the size of nouns and the place of articulation of their final stops. However, speakers ignore, or fail to reproduce, correlations between the laryngeal features of final stops and the quality of the vowels that precede them. Consequently, a model of the lexicon that LACKS vowel features is a better predictor of the results than a model that has them.

3.1. MOTIVATION FOR A DENEUTRALIZATION TASK. We employed a novel word task (Berko 1958) to find out which statistical generalizations native speakers extract from their lexicon. This kind of task has been shown to elicit responses that, when averaged over several speakers, replicate distributional facts about the lexicon (e.g. Zuraw 2000 and many others).

Recall that whether or not a stop-final noun will fall into the alternating or nonalternating class of words in Turkish is seemingly unpredictable: the unsuffixed noun stem *sop^h* does not alternate when a vowel-initial suffix is added, as in the possessed form *sop^h-u*, but the noun stem *ɟop^h* does: its possessed form is *ɟob-u*. A nonce word like *zop^h*, in which the stem-final consonant appears at the end of the word in coda position, is ambiguous, since the distinction between alternating and nonalternating stops is neutralized. When a speaker is presented with the novel form *zop^h* and asked to form the possessive, they have to undo the neutralization, and decide whether the final stop is of the alternating or nonalternating kind.

This DENEUTRALIZATION task shows a number of parallels with a more general schema of BACKWARD BLOCKING inference, discussed in the literature on causal reasoning and inductive inference. In studies on backward blocking, participants observe an outcome occurring in the presence of two potential causes (A and B). Participants observe that event A independently causes the outcome. Participants are then often less likely to judge B as the cause of the outcome. One example task in which backward-blocking inferences arise is in the 'blicket detector' task of Sobel et al. 2004, in which children were introduced to a blicket-detecting machine that lights up and plays music when certain objects (blickets) are placed on it and were told that 'blickets make the machine go'. In the blicket detector backward-blocking task at hand, A and B are two blocks placed on the blicket detector together, which results in the machine activating. Subsequently, object A is put on the detector alone, again resulting in activation of the machine. Children were then asked whether B was a blicket. As the detection of B's blickethood is neutralized in the presence of A, a known blicket, the 'logical' response rate of whether B is a blicket should have been a 50% rate of guesses that it was.

Nonetheless, in Sobel and colleagues' experiment 3, they showed that four-year-old children were remarkably sensitive to the BASE RATES of whether something was likely to be a blicket, and made use of this information in the face of the logical uncertainty of backward blocking. In this experiment, the authors exposed and familiarized children to a number of nonce objects before introducing them to the blicket detector. There were two conditions. In the 'rare blicket' condition, one out of ten of the objects that the participants were exposed to beforehand were blickets. In the 'common blicket' condition, nine out of ten objects were blickets. The children were then presented with the same task described above: seeing two objects, A and B, seeing that A lights up the blicket detector, and seeing that A and B together light up the blicket detector. The children were then asked if B was a blicket or not. The four-year-olds categorized B as a blicket on average 25% of the time in the rare blicket setup, but 81% of the time in the common blicket setup, showing that they actively employed base-rate information in the deneutralized context of B alone.

The backward-blocking blicket detector task is highly similar in structure to the coda deneutralization task we performed with nonce words in Turkish. Participants observe an outcome (e.g. [p^h] in final position) that occurs in the presence of two potential causes: an alternating paradigm with voicelessness when in coda position, or a uniformly voiceless paradigm. Once it is known that the presence of A alone is sufficient to trigger the outcome (in this case, laryngeal alternations exist in Turkish), then the likelihood that B is playing any role in the outcome should logically be 50%. When Turkish speakers are presented with a word like *zop^h*, however, and asked whether to judge whether the deneutralized form should be *zop^h-u* or *zob-u*, will they take into account the overall likelihood that a word of this shape is in the alternating class? And if so, which base rates do they track, and which do they ignore?

3.2. MATERIALS AND METHOD.

SPEAKERS. Participants were adult native speakers of Turkish ($N = 24$; thirteen males, eleven females, age range: 18–45) living in the United States. Some of the speakers were paid \$5 for their time, and others volunteered their time. The experiment was delivered as an internet questionnaire, with some speakers doing the experiment remotely. For those speakers, reaction times were indicative of the speakers taking the questionnaire in one sitting, with no discernible distractions or pauses.

MATERIALS. A native speaker of Turkish (male, mid-twenties) recorded the bare form and two possible possessive forms for each noun, repeated three times. Each stimulus was normalized for peak intensity and pitch, and inspected by a native speaker and judged to be natural and acceptable. One of the possessive forms was completely faithful to the base, with the addition of a final high vowel that harmonized with the stem, following the regular vowel-harmony principles of the language. In the other possessive form, the stem-final stop was substituted with its voiced counterpart, except for post-vocalic *ks*, which were deleted.

To exemplify all size, place, and vowel-quality combinations would have required creating ninety-six stimuli (four places * three sizes * eight vowel qualities). Since the lexical distribution of laryngeal alternations among palatals and labials is fairly similar, and in the interest of reducing the number of trials, the palatal and labial categories were collapsed into one category, using twelve words of each place, compared to twenty-four for the coronal- and dorsal-final words. The total number of stimuli, then, was seventy-two (three place categories * three sizes * eight vowel qualities).

Additionally, Turkish nouns of native origin allow the rounded vowels {*o*, *ø*} only in initial position. To make the stimuli reflect the native phonology, nonhigh round vowels

in the second syllable of the CVCVC words were replaced with the corresponding high vowels {*u, y*}. The nouns that were used are presented in Table 4. The nonfinal consonants were chosen such that the resulting nouns all sounded plausibly native, with neighborhood densities equalized among the stimuli as much as possible.

Finally, thirty-six fillers were included. All of the fillers ended in either fricatives or sonorant consonants. To give speakers a meaningful task to perform with the fillers, two lexically specific processes of Turkish were chosen: vowel-length alternations (e.g. *ruh* ~ *ru:h-u* ‘spirit’) and vowel ~ \emptyset alternations (e.g. *burun* ~ *burn-u* ‘nose’). Eighteen fillers displayed vowel-length alternations with a CVC base, and the other eighteen displayed vowel ~ \emptyset alternations with a CVCVC base. All of the fillers were chosen from a dictionary of Turkish, some of them being very familiar words, and some being obsolete words that were not familiar to the speakers we consulted.

		CVC		CVCC		CVCVC		
		-high	+high	-high	+high	-high	+high	
p/ɸ	-round	-back	gep ^h	jiɸ ^h	t ^h elp ^h	gintɸ ^h	hevetɸ ^h	ɸisip ^h
		+back	dap ^h	niɸ ^h	p ^h antɸ ^h	dirp ^h	jijap ^h	ma.iɸ ^h
	+round	-back	k ^h ɸɸ ^h	zyp ^h	jɸntɸ ^h	k ^h yrp ^h		bɸlyɸ ^h
		+back	p ^h ɸɸ ^h	t ^h up ^h	solp ^h	munɸ ^h		t ^h ryɸɸ ^h
							k ^h onup ^h	
							gujup ^h	
t	-round	-back	p ^h et ^h	hit ^h	zelt ^h	t ^h int ^h	nik ^h et ^h	gevit ^h
		+back	fat ^h	mit ^h	hant ^h	ɸirt ^h	ja.at ^h	p ^h isit ^h
	+round	-back	sɸt ^h	ɸyt ^h	gɸnt ^h	nyrt ^h		sɸlyt ^h
		+back	jot ^h	nut ^h	ɸɸolt ^h	bunt ^h		bynyt ^h
							t ^h orut ^h	
							mujut ^h	
k	-round	-back	vek ^h	zik ^h	helk ^h	t ^h ink ^h	mesek ^h	p ^h erik ^h
		+back	ɸak ^h	p ^h ik ^h	vank ^h	nirk ^h	t ^h at ^h ak ^h	banik ^h
	+round	-back	hɸk ^h	syk ^h	sɸnk ^h	p ^h yrk ^h		nɸnyk ^h
		+back	mok ^h	nuk ^h	bolk ^h	dunk ^h		dyjyk ^h
							zoruk ^h	
							juluk ^h	

TABLE 4. Experiment items.

The materials were recorded in a sound-attenuated booth into a Macintosh computer at a 44.1 KHz sampling rate. Using Praat (Boersma & Weenink 2008), the token judged best of each suffixed form was spliced and normalized for peak intensity and pitch. Peak intensity was normalized using Praat’s ‘scale peak’ function set to 0.6. For pitch normalization, three points were manually labeled in each affixed form: the onset of the word, the onset of the root’s final segment (the onset of the burst in the case of stops), and the offset of the word. Then, a reversed V-shaped pitch contour was superimposed on the materials, with a pitch of 110 Hz at the onset of the word, 170 Hz at the onset of the root-final segment, and 70 Hz at the offset of the word. These values were chosen in order to best fit most of the speaker’s actual productions, such that changes would be minimal.

Finally, for each stimulus, two .wav files were created by concatenating the two suffixed forms with a 0.8-second silence between the two, once with the voiceless form followed by the voiced form, and once with the voiced followed by the voiceless. A linguist who is a native speaker of Turkish verified that the final materials were of satisfactory quality. While she had some concerns about stress being perceived nonfinally in a few of the filler items, no problems were found with the target stimuli.

PROCEDURE. Before the beginning of the experiment, a familiar nonalternating paradigm ($t^h op^h \sim t^h op^h -u$ ‘ball’) and a familiar alternating paradigm ($\text{ʒep}^h \sim \text{ʒeb} -i$ ‘pocket’) were presented to speakers to remind them that laryngeal alternations are lexically specific. Then, speakers were asked to choose the possessive form of two familiar alternating nouns ($dolap^h$ ‘cupboard’ and $a.atf^h$ ‘tree’), and feedback was given on their choices.

The stimuli were presented in a self-paced forced-choice task. The base form (e.g. fet^h) was presented in Turkish orthography (e.g. ⟨fet⟩), which reflects the relevant aspects of the phonology faithfully. The participants saw an overt possessor with genitive case followed by a blank, to provide the syntactic context for a possessive suffix (e.g. $Ali'nin$ ‘Ali’s’), and they heard two possible possessed forms (e.g. $fet^h -i$ and $fed -i$). Speakers pressed ‘F’ or ‘J’ to choose the first or the second possessive form they heard. Most speakers took fifteen to twenty minutes to complete the experiment.

The order of the stimuli and the order of the choices were randomized. Additionally, the fillers were randomly distributed among the first three quarters of the stimuli.

The results of the experiment are presented in §3.3 below, with qualitative comparisons to the lexicon, followed in §3.4 by a quantitative comparison of the results with the lexical statistics.

3.3. RESULTS. The charts in Figure 4, which parallel the charts in Fig. 1, show that participants preferred alternating stops most often for labials, and least often for coronals. The size effect is clear as well, with more alternating responses for longer items. While there seems to be a height effect in 4c, this is primarily due to the larger number of high vowels in the CVCVC stimuli, due to the phonotactic absence of $\{o, \emptyset\}$ in non-initial syllables. There is no difference to speak of between front and back vowels.

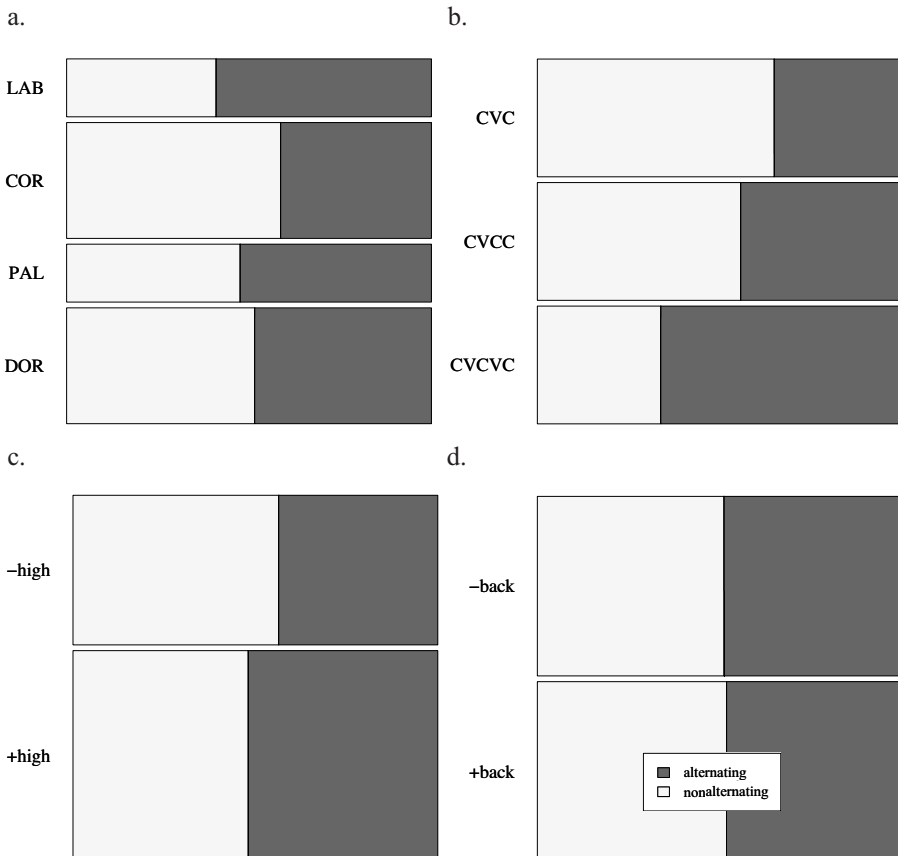


FIGURE 4. Alternating choices for nonce words, by single features.

The interaction in Figure 5a, which parallels the one in Fig. 2a, shows a clear gradation for size for labials and coronals. For palatals, we see CVCC nouns patterning with CVCVCV nouns, while for dorsals, the CVCC nouns pattern with the CVC nouns.

The backness effect, or rather its absence, can be seen in Fig. 5b. In the lexicon the backness effect was concentrated in the palatals, whereas in the experimental results the effect is not just simply absent in the palatals, but is reversed.

Figure 5c shows a small preference for alternations in high vowels in the palatals and dorsals, contrary to the lexicon, where the largest height effect was in the coronals. Figure 5d shows a slight preference for alternations after high vowels in CVCVCV nouns.

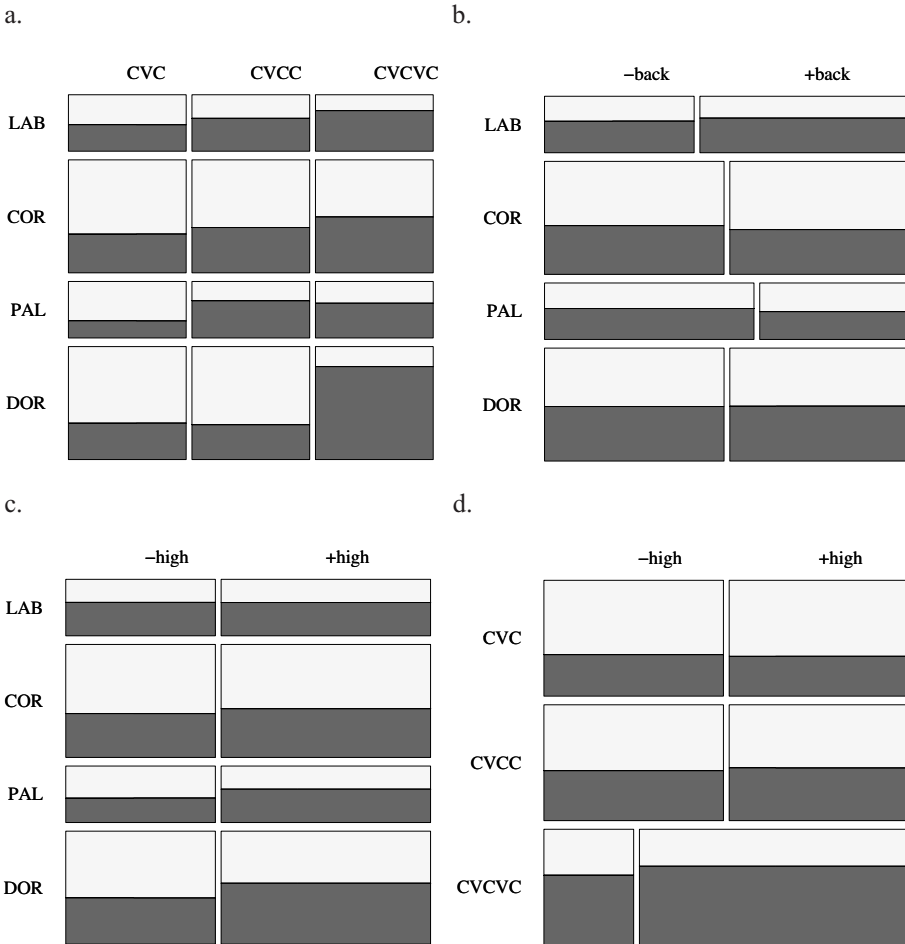


FIGURE 5. Alternating choices for nonce words, by pairs of features.

Finally, Figure 6, which parallels Fig. 3, shows the height effect in the palatals, and more weakly in the dorsals and coronals. The picture is rather noisy, with no overall height effect, nor any consistent effect in any of the three sizes or any of the four places. In the lexicon, the height effect is strongest for coronals and weakest for palatals, whereas the opposite is true in the experimental results—another sign of its rather haphazard nature.

We now turn to a statistical analysis that assesses the reliability of these effects. The results were analyzed with a mixed-effects logistic regression in R (R Development Core Team 2007) using the *lmer()* function of the *lme4* package, with *participant* and

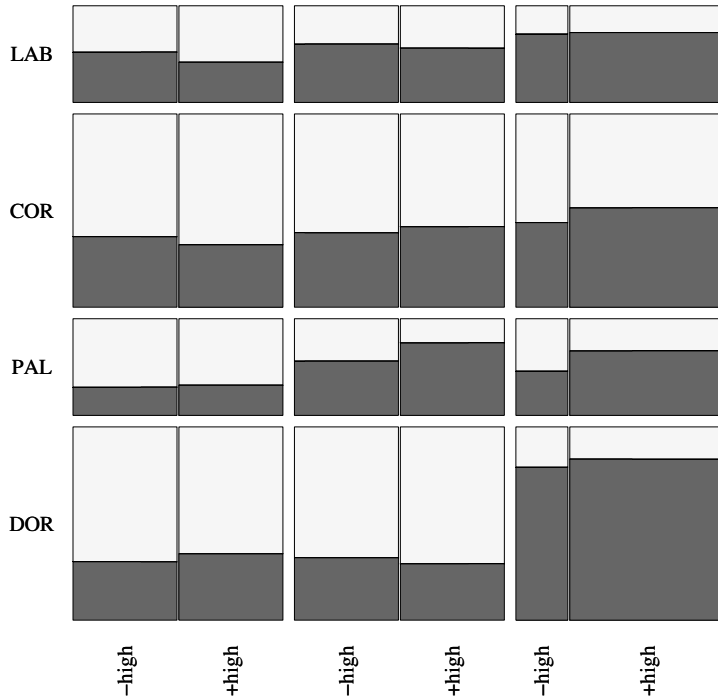


FIGURE 6. Alternating choices for nonces: place, size, high.

item as random-effect variables. The fixed-effect variables were the same ones used in the analysis of the lexicon: *size*, *place*, *high*, *back*, and *round*.

An initial model was fitted to the data using only *size* and *place* as predictors. Adding their interaction to the model made a significant improvement (sequential ANOVA model comparison, $\chi^2(6) = 50.58, p < 0.001$). The improved model with the interaction term is given in Table 5. This model shows that labial place and CVCVC size are more conducive to alternating responses than the baseline dorsal place and CVC size, respectively, which is exactly what we found in the lexicon model in Table 3. For the CVCC size, palatal place is significantly more conducive to voicing than the baseline dorsal place—here we see that speakers amplified and sharpened what was a mere trend in Table 3. Additionally, in the CVCVC size, all places are less conducive to alternating responses than the baseline dorsal place with the same CVCVC size, which closely matches Table 3.

The addition of any vowel feature to the baseline model (*high*, *back*, or *round*) made no significant improvement ($p > 1$). No vowel feature approached significance, either on its own or by its interaction with *place*. For example, adding the interaction *place*high* to the model in Table 5 yields a new model where the interaction of coronal place and *high* is almost exactly at chance level ($p = 0.981$). Adding *place*back* to the baseline model gives an interaction of palatal place and *back* that is nonsignificant ($p = 0.661$) and its coefficient is negative, that is, going in the opposite direction from the lexicon, where palatal place and backness are positively correlated.

In other words, *size* and *place* had statistically significant power in predicting the participants' choice of alternation vs. nonalternation of stem-final stops. Crucially, however, none of the vowel features had an effect on the participants' choices, not even as a trend.

	ESTIMATE β	SE(β)	z	p
LAB	0.74	0.30	2.45	0.014
COR	0.11	0.26	0.43	0.665
PAL	-0.12	0.32	-0.37	0.710
CVCC	-0.09	0.26	-0.34	0.733
CVCVC	2.69	0.29	9.47	< 0.001
CVCC:LAB	0.64	0.43	1.49	0.137
CVCC:COR	0.39	0.36	1.07	0.287
CVCC:PAL	1.87	0.45	4.17	< 0.001
CVCVC:LAB	-1.44	0.46	-3.15	0.002
CVCVC:COR	-1.94	0.38	-5.14	< 0.001
CVCVC:PAL	-1.13	0.46	-2.46	0.014

TABLE 5. Final regression model for alternating choices of nonce words.

3.4. COMPARISON OF THE EXPERIMENT WITH THE LEXICON. In this section, we compare how closely the experimental results match models of the lexicon that are based on TELL. We quantify the matches that each lexicon model makes, and assess the contribution of the grammatical predictors we use (*place*, *size*, *high*, and *back*). We show that the lexicon model that takes ONLY place and size into account is the one that matches the experimental results best, and it makes significantly better predictions than the model that adds vowel features. Before we delve into the statistics, however, we start by visualizing the contribution of our four grammatical factors.

The effect of individual factors is shown in Figure 7, where the four charts parallel those in Figs. 1 and 4. Each chart was made by fitting an ordinal logistic regression model to TELL, as described in §2, and using the *predict()* function to generate the expected alternation rate. In such a simple model, this is just the average for the relevant items, for example, 20% for *t*-final nouns and 63% for *tf*-final nouns. The prediction of each model was merged with the experimental results, and thus plotted against the average experimental response for each item. We see a strong correlation between the lexicon and the experimental results for *place* and *size*, and no effect for *back*. We do see an effect of *high*, but this is again a reflection of the *size* effect, since there are more CVCVC stimuli with high vowels than CVCVC stimuli with nonhigh vowels.

The charts in Figure 8 show the predictions of lexicon models that use pairs of predictors with their interactions, paralleling those in Fig. 2 and Fig. 5. Here it becomes clearer that *place*size* is a powerful combination that offers an excellent prediction of the experimental results. The correlation is much more modest when *place* is augmented by *back* or *high*. Comparing Figs. 8c and 8d, we see that *high* has a rather a modest effect on the correlation, and the difference between the charts is attributed to the stronger effect of *size* relative to *place*.

Finally, Figure 9 compares a model of the lexicon that uses only *size*place* (on the left) with a model that uses *size*place* + *high*place* + *back*place* (on the right). This right-hand model is identical to the one in Table 3. While the data points are more evenly distributed on the x-axis in Fig. 9b, we see no improvement in correlation relative to Fig. 9a.

Moving beyond the visualization in Fig. 9, we employed model comparison tests to show that Fig. 9a offers a superior match to the experimental results. To test the effect of *high* separately from *back*, we define four models, shown in Table 6. The BASE model has no vowel features in it, HIGH and BACK have one additional feature each, and FULL is identical to the model in Table 3.

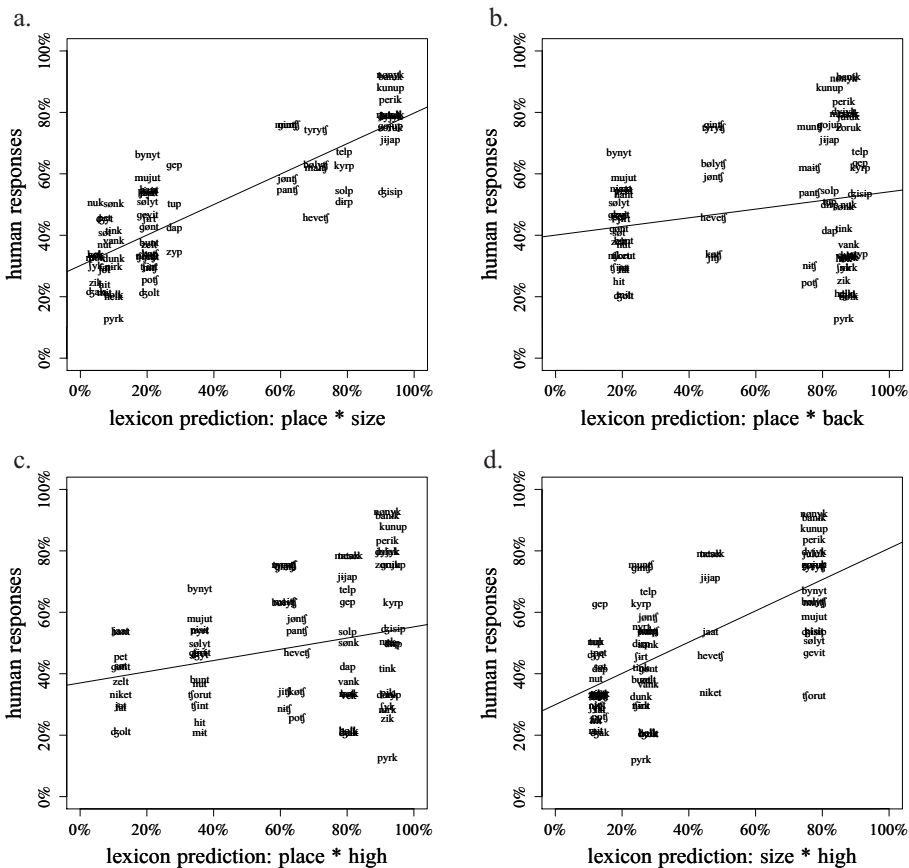


FIGURE 8. Lexicon vs. nonce words, by pairs of features.

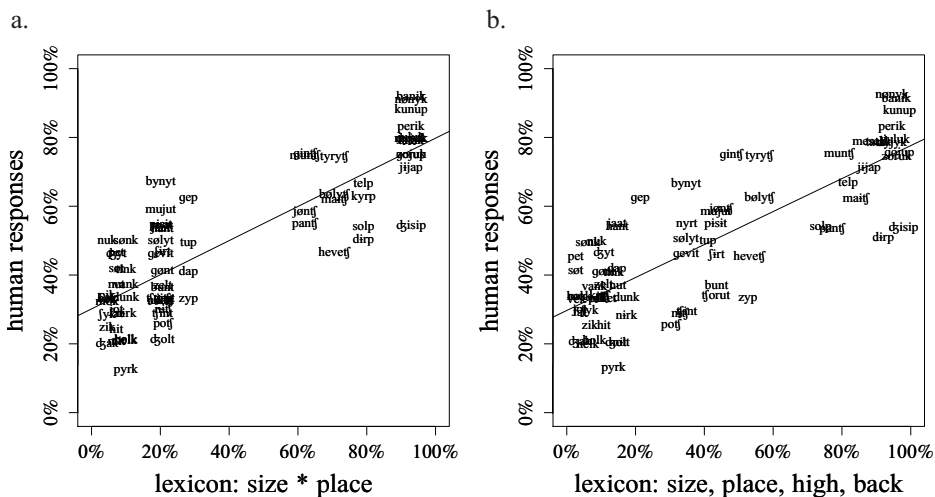


FIGURE 9. Lexicon vs. nonce words, without and with vowel features.

BASE	place*size
HIGH	place*size + place*high
BACK	place*size + place*back
FULL	place*size + place*high + place*back

TABLE 6. Our four models.

	FIT TO LEXICON		FIT TO EXPERIMENTAL RESULTS	
	MODEL LR	D_{xy}	MODEL LR	D_{xy}
BASE	1920	.744	196	.360
BACK	1948	.757	187	.349
HIGH	2049	.800	181	.345
FULL	2078	.810	175	.349

TABLE 7. Model fit.

To summarize the findings, Turkish speakers reproduced the distribution of laryngeal alternations in the lexicon by paying attention to the size of the nouns and the place of the final stops. The lexicon model that lacks vowel features (BASE) is a significantly better predictor of the human responses than models that use vowel features, even though vowel features are strongly predictive of laryngeal alternations in the lexicon.

3.5. EVALUATING POTENTIAL CONFOUNDS IN THE EXPERIMENTAL STIMULI. Before proceeding with the implications of our findings for phonological theory, we report on the results of two post-hoc evaluations designed to ensure that our results were not due to confounding properties of the stimuli. Specifically, we examined the naturalness of the acoustic properties of each stimulus using a post-hoc norming study, and we assessed the potential role of neighborhood density in explaining the alternation results. Neither turned out to have any confounding effect.

The potential effect of purely auditory or other non-task-related properties in biasing speakers' choice of alternating or nonalternating response was addressed by conducting a post-hoc norming study. We asked nine native speakers to listen to all of the suffixed forms, in random order. Each sound file was presented together with its orthographic representation, and speakers were asked to 'rate the clarity of pronunciation' of each word on a scale from 1 to 7. Speakers were told that all the forms they will hear are possessives. The study was done via the internet, using an interface very similar to the one we used for the main study. Overall, the speakers found the materials to be of high quality, with a mean of 6.2 out of 7 ($SD = 1.2$, 56% '7' responses, 23% '6', 12% '5'). We then used the average rating per item as a predictor of the experimental results, and found no significant predictive power. We conclude that our stimuli did not contain any acoustic properties that could shape the speakers' responses in any measurable way.

Recall that in §2.3 we found that existing lexical items are more likely to alternate if they are frequent and/or if they have many alternating neighbors (while, of course, being independent from the grammatical effects we study). Frequency is not a concern for nonce items, since by definition their frequency is zero. To make sure that neighborhood density is not masking the grammatical effects, we calculated the neighborhood density of the nonce items, and used it to predict the experimental results. We found that adding neighborhood density as a predictor into the analysis in Table 5 made no noticeable change, as confirmed by an ANOVA model comparison ($\chi^2(1) = .260, p > 0.1$).

3.6. DISCUSSION. The experimental results show that Turkish speakers generalize their knowledge of the laryngeal alternations in their lexicon. Not contenting themselves with memorizing the alternating or nonalternating status of single nouns, speak-

ers have access to the relative proportion of alternating nouns categorized by size and place. Using size and place as factors, speakers must somehow project their lexical statistics onto novel items. Although the height and backness of stem-final vowels are strongly correlated with alternations in the lexicon, we see no effect of these vowels on speakers' choices. In fact, a model that ignores the vowel features can predict the speakers' responses significantly better than a model that uses vowel features.

Speakers failed to reproduce the correlation between vowels and laryngeal alternations in spite of an abundance of overt evidence, while learning the size and place effects even with very little evidence. For instance, the difference in alternation rates between *ʃ*-final CVC and CVCC nouns was successfully reproduced in the experiment results, even though the evidence comes from twenty-three and eighteen nouns, respectively, and this is a nonsignificant trend in the lexicon. The evidence for the vowel effects, however, comes from hundreds of nouns.

The proposal advanced here is that the results are best understood in light of a theory of universally possible phonological interactions, as encoded in a set of universal constraints. Only factors that can be expressed in terms of constraint interaction can be identified by language learners, with other lexical generalizations going unnoticed.

4. REPRESENTING TURKISH LARYNGEAL ALTERNATIONS IN THE GRAMMAR. In this section we offer a model of the grammar of Turkish that incorporates constraints into the determination of laryngeal alternations, so as to prevent unrestricted access to the lexicon. In order to model the experimental results in which not all statistically robust lexical predictors of alternation are extended in the novel word task, we propose a grammatical filter through which such trends can be extended. In essence, the representation of alternation predictions must be mediated by a grammar with formal primitives that can be used to encode natural interactions. In order to have a concrete implementation, we present our analysis of Turkish in §4.3, which uses an optimality-theoretic grammar with lexically specific rankings. However, we first review the phonetics and phonology of Turkish laryngeal contrasts in §4.1. We then show why the difference between alternating and non-alternating nouns must not be encoded in the underlying representation of roots (§4.2) if one is to formulate a grammatical explanation for our experimental results.

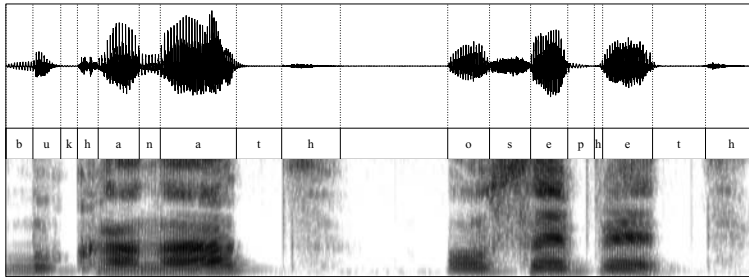
4.1. LARYNGEAL CONTRASTS IN TURKISH. The literature on Turkish (at least since Lees 1961) agrees that Turkish contrasts two stops in each place of articulation on the surface (3), but that stem-final stops display three kinds of behavior under affixation: they are either pronounced the same in the base and in the affixed form (4a,b), or they alternate (4c). It is also known that final voiced stops, as in 4b, are rare in the language.

- (3) Two-way surface distinction in roots
- | | | |
|----|--------------------------|------------------------------|
| | initially | intervocally |
| a. | t ^h in 'soul' | at ^h a 'ancestor' |
| b. | din 'religion' | ada 'island' |
- (4) Three-way contrast finally
- | | | |
|----|--------------------------------|-----------------------------|
| | bare stem | possessive |
| a. | at ^h | at ^h -i 'horse' |
| b. | ad | ad-i 'name' |
| c. | t ^h at ^h | t ^h ad-i 'taste' |

In Turkish orthography, the surface distinction is represented by the letters ⟨p, t, ç, k⟩ and ⟨b, d, c, g⟩, and the distinction has been taken to be one of voicing by much of the literature on Turkish (Lees 1961, Inkelas & Orgun 1995, Inkelas et al. 1997, and many others).

More recently, Kallestinova (2004) and Petrova and colleagues (2006) have shown that the voiceless stops of Turkish are in fact aspirated in onset position.¹⁰ While these authors do not commit to the surface realization of word-final stops, it is known that final stops are consistently released with an audible voiceless burst. Crucial evidence for considering this audible release as aspiration comes from Kopkallı (1993), who shows that the release of word-final stops is as long as the duration of aspiration on intervocalic voiceless stops, suggesting that speakers treat these as a consistent phonetic category. (For further discussion of laryngeal features in Turkish, see Jannedy 1995, and for a broader perspective, see Avery 1996, Beckman & Ringen 2004, Vaux & Samuels 2005, and Nicolae & Nevins 2010.)

The spectrogram in Figure 10 exemplifies the finding in Kopkallı 1993, showing a clear, voiceless burst at the end of both the alternating *k^hanat^h* and the nonalternating *seph^het^h*. In fact, this token, spoken by a thirty-year-old male speaker from Istanbul, happens to have an even stronger burst for *k^hanat^h*, although Kopkallı shows that there is no significant difference in the duration of the final burst between alternating and nonalternating nouns.



[bu k^hanat^h o seph^het^h] 'This is a wing; that is a basket' (lit. 'this wing; that basket')

FIGURE 10. Waveform and spectrogram for one alternating and one nonalternating stop.

For the purposes of the analysis we offer in §4.3, the exact details of Turkish laryngeal features are not crucial. What is crucial is that all stop-final nouns fall into one of two groups: in one group, the suffixed form is faithful to the base (such that faithfulness to laryngeal features ranks over any relevant markedness constraints), and in the other group, the suffixed form is unfaithful (and markedness ranks over any relevant faithfulness constraints). As we show below, the inconsistent ranking arguments allow speakers to build lexical information into their grammars, and thus learn the distribution of the laryngeal alternations in grammatical terms. In this article, we use the more accurate transcription, which marks aspiration.

Under this view, Turkish stops surface either voiced or aspirated. Any hypothetical underlyingly voiceless unaspirates map unfaithfully either to voiced or to aspirated stops due to a high-ranking constraint that requires a laryngeal specification on every stop (Petrova et al. 2006). Additionally, barring a few exceptional native words and some loanwords, word-final stops are regularly required to be aspirated, as has been shown for German, Kashmiri, and Klamath (Iverson & Salmons 2007).

¹⁰ The aspiration is consistent in roots. In affixes that show laryngeal alternations, such as the locative *-ta/-da* and the ablative *-tan/-dan*, the voiceless variant is unaspirated. In affixes that do not alternate, like the relativizer *-k^hi* and adverbial *-k^hen*, voiceless stops are aspirated just like root stops.

4.2. ENCODING (NON)ALTERNATION WITH CONSTRAINT RANKINGS. The existing analyses of Turkish laryngeal alternations, either in terms of voicing (Lees 1961, Inkelas & Orgun 1995, Inkelas et al. 1997) or in terms of aspiration (Avery 1996, Kallestinova 2004, Petrova et al. 2006), share the same architecture that attributes the different behavior of final stops to different underlying representations of laryngeal features. In this section, we demonstrate that such an analysis is insufficient on its own, showing that speakers must use grammatical mechanisms instead of underlying representations to generalize over the distribution of laryngeal alternations.

The traditional analysis along the lines of Inkelas et al. 1997 is shown in 5. In this analysis, nouns that surface with a voiceless (aspirated) stop throughout the paradigm have a voiceless (aspirated) stop underlyingly, while stops that alternate have an underlying stop that is unspecified for laryngeal features. Identity to laryngeal features assures that underlyingly specified stops surface faithfully in all positions, while a constraint against intervocalic voiceless stops causes alternation when faithfulness is not at issue (UR = underlying representation).

- (5) a. The URs of [at^h] and [t^hat^h] are /at^h/ and /t^haD/
 b. The UR of the possessive is /I/ (a high vowel)
 c. /at^h + I/ → [at^h-i] **requires** IDENT(lar) >> *VtV

	/at ^h + I/	IDENT(lar)	*VtV
a. ☞	at ^h -i		*
b.	ad-i	*!	

- d. /t^haD + I/ → [t^had-i] is **consistent** with IDENT(lar) >> *VtV

	/t ^h aD + I/	IDENT(lar)	*VtV
a.	t ^h at ^h -i		*!
b. ☞	t ^h ad-i		

In this theory, IDENT(lar) dominates any relevant markedness constraints, and alternating stops have underspecified underlying representations that escape faithfulness. The deletion of dorsals can be encoded using another representational mechanism, that of ‘floating segments’, or segments whose absence from the output does not violate the regular MAX (as in e.g. Zoll 1996).

The crucial element of such an analysis is that both rankings in 5 are consistent. Alternating nouns like t^hat^h and nonalternating nouns like at^h do not require different grammatical factors that point to their alternation, and thereby cannot situate alternation itself as something specifically interacting with the phonological grammar of the language. Rather, the behavior of different nouns is encoded in the lexicon, outside the purview of grammar. The same is true of Avery 1996, Kallestinova 2004, and Petrova et al. 2006.

If all of the nouns of Turkish are derived using a single grammar, as in 5, then the trends in the lexicon are not available to the grammar. One could imagine that a speaker will find the relevant lexical statistics by going directly to the lexicon and extracting the relevant information from it. When going to the lexicon directly, however, the speaker will not be biased by UG to find only grammatically principled generalizations, contrary to the observed facts.

We propose that the status of a word as alternating or nonalternating must be represented by lexically specific grammatical rankings, instead of (or in addition to) an underlying difference. The only way to make sense of the grammatical biases against extending all lexical statistics is to encode the alternating or nonalternating status of a word as a GRAMMATICAL phenomenon.

The analysis we offer in §4.3, summarized in 6 below, posits the bare forms of nouns as their underlying representations. As 6c shows, differential alternation for the same UR forces the speaker to find conflicting ranking arguments for the different classes of lexical items. The result of aggregating such ranking arguments becomes the vehicle for encoding lexical statistics in the grammar.

- (6) Overview of conflicting ranking analysis of alternating vs. nonalternating nouns
- a. The URs of [at^h] and [t^hat^h] are /at^h/ and /t^hat^h/
 - b. The UR of the possessive is /I/ (a high vowel)
 - c. /at^h + I/ → [at^h-i] **requires** IDENT(lar) >> *VtV
/t^hat^h + I/ → [t^had-i] **requires** *VtV >> IDENT(lar)

In effect, alternating vs. nonalternating status is not captured in terms of underlying representations (which are, by hypothesis, the same for both classes of items), but rather in terms of lexically specific constraint rankings. This assures that lexical trends are identified in terms of constraints, and thereby are captured in phonological terms, using the variety of phonological primitives that constraints are sensitive to, such as marked combinations of features, preferred alignments of phonological elements, positional faithfulness, and other formal properties of the theory.

As should be clear, the proposal made here ‘reverses’ the effect of the phonology: instead of assigning the hidden aspects of bases to their underlying representation, and then neutralizing them in the unaffixed form, as is done traditionally, we propose that for Turkish, the surface forms of bases are assumed as their underlying form, and any properties of the base that emerge only in suffixed forms are achieved by constraint interaction. In the simple case of Turkish, where the only hidden property of nominal roots is the laryngeal specification of their final stop, the analysis in terms of lexically specific constraint rankings is not only clearly feasible, but it is also the only analysis that allows speakers to grammatically capture the variety of lexical trends that the language has. We demonstrate this in more detail in the next section.

The idea that the surface form of the base may be preferred to an abstract UR finds previous support in the work of Hayes (1995, 1999), and a similar approach is pursued in Albright 2008b. These authors discuss a range of languages where roots had been claimed to have abstract URs, as in Korean, and analyze them in grammatical terms. Assuming the base form as the underlying representation has the added benefit of obviating the search for non-surface-true underlying representations. This search requires a significant amount of computation, as shown by Tesar (2006) and Merchant (2008), and in parallel lines of work, also by Boersma (2001) and by Jarosz (2006), who specifically look at Turkish-like languages where the behavior of root-final stops is hidden in the bare form of the root. A full comparison of the computational complexity of these approaches and our approach goes beyond the scope of this article, however.

4.3. ANALYSIS WITH CLONED CONSTRAINTS. As established in §3, when asked to choose a possessive form for a novel item like *gevit^h*, Turkish speakers match the probability of alternation for items that share the relevant phonological properties, in this case, *place* (*t*-final) and *size* (polysyllabic). They do not perform an unconstrained comparison to their lexicon, however, which would wrongly predict an effect of the stem-final vowel. In this section, we present an OT-based model that uses the constraints in CON to store information about lexical items and project the trends that these items create onto novel items. The proposal is outlined here, with further details available in Becker 2009.

As explained in §4.2, we assume the bare stems of Turkish nouns as their underlying forms. Nouns with nonalternating stops require FAITHFULNESS >> MARKEDNESS, while nouns with alternating stops require MARKEDNESS >> FAITHFULNESS, and these conflicting demands lead to two conflicting grammars. Constraint cloning (Pater 2006, 2009) is a mechanism that uses the constraints in CON to categorize and list lexical items that have unpredictable behavior. Aggregating the pattern of alternating vs. nonalternating items specifically in terms of constraint rankings ensures that they are only categorized based on universal grammatical principles that are found in the inventory of universal grammar. The tableaux in 7 and 8 exemplify the conflicting grammars required by the nonalternator *anatf^h* ‘cub’ and the alternator *amatf^h* ‘goal’.

(7)

	/anatf ^h + i/	IDENT(lar)	*VtjV
a.	anatf ^h -i		*
b.	anaɖɟ-i	*!	

(8)

	/amatf ^h + i/	*VtjV	IDENT(lar)
a.	amaɖɟ-i		*
b.	amatf ^h -i	*!	

When the learner encounters a conflict as in 7 and 8, he or she can no longer maintain one grammar for the entire language. One constraint is chosen for cloning, meaning that two copies of the constraint are created, and each copy is associated with the same lexical items. If IDENT(lar) is chosen, the resulting grammar is that in 9.

$$(9) \text{IDENT(lar)}_{\text{anatf}} \gg *VtjV \gg \text{IDENT(lar)}_{\text{amatf}}$$

To replicate the effect that *place* has over the distribution of laryngeal alternations, the language learner must separately keep track of words that end in different stops. The fact that laryngeal features affect stops of different places of articulation differently is well documented (e.g. Lisker & Abramson 1964, Ohala 1983, Volatis & Miller 1992). Additionally, the lenition of voiceless stops to voiced stops between vowels is also very well documented (for an overview, see Kirchner 1998). These effects quite plausibly give rise to a family of constraints that penalize voiceless stops between vowels: *VpV, *VtV, *VtjV, *VkV. The interaction of each of these constraints with faithfulness will allow the speaker to discover the proportion of the stop-final nouns of Turkish that alternate in each place of articulation. When any of these markedness constraints rank over the faithfulness constraint IDENT(lar), the stem-final stop surfaces voiced. When *VkV outranks the anti-deletion constraint MAX, a stem-final dorsal deletes.¹¹

Speakers also replicate the *size* effect of the lexicon, with monosyllabic nouns recorded separately from polysyllabic nouns. We attribute this to initial-syllable faithfulness (Beckman 1997, 1998, Casali 1998). In a monosyllabic alternator like *tatf* ~ *taɖɟi* ‘crown’, the alternation impacts the initial syllable of the base, and thus *VtjV must outrank both IDENT(lar)_{σ1} and the general IDENT(lar). In a polysyllabic alternator like *amatf* ~ *amaɖɟi* ‘goal’, the initial syllable is not disturbed, and *VtjV is only required to outrank IDENT(lar). Listing items with these different constraint rankings naturally separates the monosyllables from the polysyllables. Similarly, the availability of

¹¹ The full analysis of dorsal deletion also involves a constraint against postvocalic [g] and appropriate ranking of IDENT(lar). These details are covered in Becker 2009.

MAX_{σ1} in addition to MAX allows the deletion of stem-final *k* to be learned separately for monosyllables and polysyllables.

Adding *atf* ~ *atfi* ‘hunger’ and *gytʰ* ~ *gyɟ-y* ‘force’ to the grammar fragment in 9 yields the grammar fragment in 10, which separates monosyllables from polysyllables.

(10) IDENT(lar)_{σ1atf}, IDENT(lar)_{anatf} >> *VtʃV >> IDENT(lar)_{σ1tatf}, IDENT(lar)_{amatf}

The complete grammar lists all the stop-final items that speaker knows. Since each lexical item must be recorded in terms of its lexically specific ranking of markedness (*VpV, *VtV, *VtʃV, *VkV) and faithfulness (IDENT(lar), IDENT(lar)_{σ1}, MAX, MAX_{σ1}), lexical items are naturally separated according to their size and place. The relative numbers of items in each group constitute the lexical statistics, which become available to the speaker in terms of rankings of universal constraints. The height and backness effects that are in the lexicon are not learned, because no known constraints relate vowel quality to the voicing of a following consonant; it is assumed that no constraints such as *[+high]tV exist that would encourage alternations following a high vowel.

Now suppose a speaker encounters a novel noun in its bare form, such as *hevetʰ*, and is required to produce its possessive form. The learner has a choice between two grammars that can apply to this item: one that has IDENT(lar) outranking *VtʃV, and one that has the reverse ranking. The clone of IDENT(lar) that ranks below *VtʃV has more items associated with it, and it therefore has a stronger effect on novel words, making the possessive of *hevetʰ* more likely to be *hevetɟ-i* than *hevetʰ-i*. In this theory, the grammar and the lexicon are intertwined, with the behavior of known items recorded in terms of constraint ranking, and it is this aspect of the grammar that allows it to project the trends in the lexicon onto novel items.

We note that cloning is designed specifically to keep track of paradigmatic alternations. While extending this approach to phonotactics, and especially to exceptional phonotactics, is left for future work, a recent attempt in this direction can be found in Coetzee 2008. Another potentially relevant mechanism for connecting the grammar and the lexicon is found in Zuraw’s (2000) USELISTED approach. The cloning approach we use here is designed to regulate paradigmatic relations and provide a way to encode patterns of alternations in the grammar. As such, it is regulated by the inventory of extant and plausible UG constraints. Since, by hypothesis, UG does not contain constraints that relate vowel height and vowel backness to laryngeal features of neighboring consonants, such relations cannot be encoded in terms of constraint ranking or cloning. The technical details of the analysis are worked out in Becker 2009; here, we wish to focus on the crux of the analysis, which is the encoding of lexical trends in terms of a grammar. It is the mediation of the grammar in the acquisition and extension of lexical trends that connects our results in Turkish to the broader picture of phonological patterns in the world’s languages.

5. GENERAL DISCUSSION. This article presented a study of Turkish laryngeal alternations that contrasted trends found in the Turkish lexicon with the knowledge that speakers have about it, showing that speakers are biased to reproduce certain trends but not others. The experimental finding, that speakers do not adopt an omnivorous model of statistical generalization when it comes to vowel-consonant interactions, falls under a more general set of conclusions about the phonetic basis for phonotactic interactions. Taken together, these results suggest a more general implication for realistic models of inductive generalization from linguistic regularities: the need for a balanced interaction between the power of tracking statistical information and the constraints of linguistically specific filters that guide the learner’s analysis and acquisition of phonotactic patterns.

5.1. GENERAL-PURPOSE LEARNING WITH THE MGL. In this subsection we compare a bi-ased learner (i.e. one that excludes interactions of vowel height and backness with consonantal laryngeal features) with a model of learning that has no substantive biases. We report on the results of such a simulation to demonstrate that unbiased models erroneously predict extension of V-C interactions in the experimental results.

The MINIMAL GENERALIZATION LEARNER (MGL) of Albright & Hayes 2002, 2003, 2006 is an information-theoretic algorithm that generalizes patterns over classes of words that undergo similar alternations. The MGL provides a reflection of trends in the lexicon and has the potential to generalize them to novel outputs. The MGL has been shown to successfully model humans’ experimental results in novel word-formation tasks with the past tense in English and with similar tasks in other languages, and is thus a good representative of a class of models that access lexical patterns without any bias against generalizing from phonologically unnatural trends.

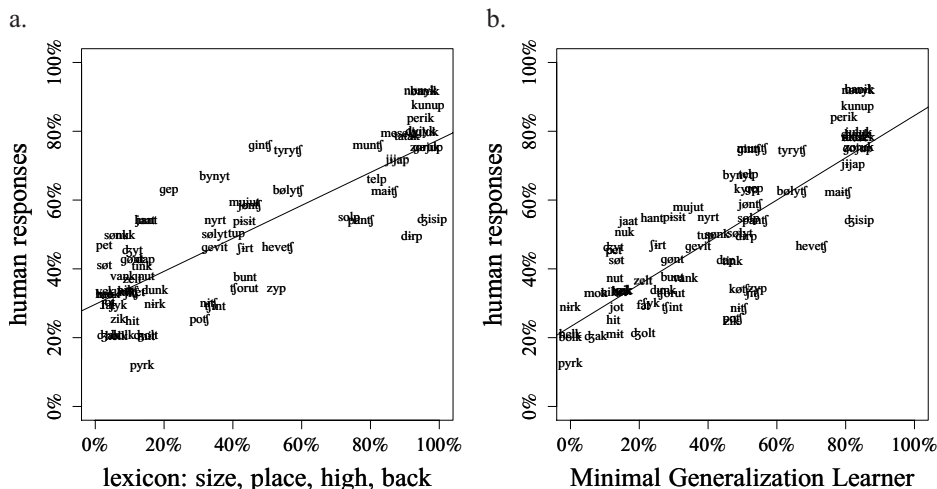


FIGURE 11. Predicting alternations in nonce words, regression model vs. MGL.

We supplied the MGL with the lexical items in TELL and assessed its predictions for the experimental items. The results are plotted in Figure 11, where Fig. 11a is repeated from Fig. 9b, and Fig. 11b shows the MGL predictions. We see similarly impressive fits in both.

We assessed the ability of the MGL to predict the experimental results, comparing the predictions to the four models we used in §3.4, using comparison of nested models. The MGL consistently achieves lower χ^2 values than the four regression models we use in §3.4, and its overall fit is closest to the FULL model, indicating that the MGL is most similar to the statistical model that uses vowel-quality information and least similar to the model that uses no vowel information.¹²

The MGL results show that a learning model unequipped with a set of biases that determine the universal range of phonological interactions will be unable to successfully mimic human behavior and ignore accidental regularities in a lexicon.

¹² Comparison of MGL with the BASE model: 5.61 vs. 19.25; with the HIGH model: 8.29 vs. 16.25; with the BACK model: 9.09 vs. 15.91; with the FULL model: 12.46 vs. 14.77. All tests are significant ($p < 0.05$).

The MGL results are representative of a wider range of learning algorithms, such as CART (Breiman et al. 1984), C4.5 (Quinlan 1993), or TiMBL (Daelemans et al. 2002), which use purely distributional properties of a lexicon to model human behavior. The MGL's advantage over these other models is that it is not given a list of possible generalizations to explore in advance, but rather generates its own set of hypotheses. The MGL simulation is informative specifically because it is given whole words to deal with, without additional information about which generalizations to attend to. The finding that the MGL replicates the FULL model's fit to the experimental results confirms the fact that a biased learner is necessary. With models other than the MGL, the lack of vowel effects could be hard-wired by not supplying the model with information about vowel quality (although this would be an implausible move, given that vowel information is necessary to correctly predict the harmonizing vowel of the possessive suffix). Since these models do not have any information about natural phonological interactions, such an exercise would offer little insight into the problem at hand.

5.2. PHONETIC FEATURES AS A BASIS FOR SECOND-ORDER PHONOTACTICS. We claim that speakers are attuned to certain factors and ignore others, and furthermore, that the choice is based on a principled inventory of universally possible phonological interactions. Among these are the fact that the size of a word and the place of articulation of an alternating stop are reasonable determinants of phonotactic distributions to consider in whether a stop will undergo a laryngeal alternation, but that the height or backness of a preceding vowel are factors that learners are not biased to consider in tracking laryngeal alternations.

The size effect can be traced to a well-known initial-syllable effect. Crosslinguistically, initial syllables enjoy greater faithfulness, or resistance to alternation (Beckman 1998). The initial syllable plays a central role in Turkish phonology: native Turkish nouns allow voiced codas only in the initial syllable (e.g. *ab.la* 'elder sister', *ad* 'name'), and initial syllables serve as starting points for vowel harmony. Nakipoğlu and Ketrez (2006) find that children quickly master suffixal allomorphy for the aorist, which is based on syllable count. Ketrez (2007) finds that child metathesis errors involving labials (e.g. *k^{hit}ap^h → k^{hip}at^h* 'book') do not occur with monosyllables (e.g. *yap^h*) and attributes this to protection of the initial syllable. In addition, Barnes (2001) finds significantly longer duration for initial syllables in Turkish. Hence, a predicate such as 'within initial syllable' is likely to be a salient factor for Turkish learners, and thus biases attention to alternation rates correlated with this factor.

The place of articulation of stem-final stops is also very likely to influence alternation rates. Different places are known to interact differently with laryngeal features (Lisker & Abramson 1964, Ohala 1983, Volatis & Miller 1992). Specifically in Turkish, dorsal stops delete rather than undergo voicing intervocalically, supplying a cue to learners that the behavior of at least one place must be learned separately. Indeed, Nakipoğlu and Üntak (2006), studying alternations in real words, show that Turkish-learning children (at ages 4;1, 6;1, 6;11) are sensitive to the differential behavior of the different places of articulation, replicating the patterns of adult alternation rates for these places.

By contrast to size and place, the vowel that precedes the stem-final stop is not likely to play any causal role in stop alternations, and hence we argue that learners ignore this factor. Although consonantal laryngeal features have been argued to affect vowel height in various languages, as in Canadian raising (Chambers 1973, Moreton & Thomas 2007) and Polish (Gussmann 1980)—in many cases due to the historical development

of quality alternations from a preexisting vowel-length contrast in closed syllables—there is no report of vowel height or backness inducing a change in laryngeal features in a following obstruent.

We argue that this typological gap reflects a principled lacuna in the inventory of possible phonological interactions, and specifically that phonological grammars lack any constraint-based or rule-governed process of vowel quality affecting adjacent consonantal laryngeal features. In fact, Moreton (2008), in an attempt to teach an artificial language pattern with height-laryngeal interactions (i.e. in which VC sequences were always high vowel followed by voiced consonant or nonhigh vowel followed by voiceless consonant), found that participants were biased against generalizing this pattern. Importantly, Moreton's subjects were able to learn a comparably complex vowel-to-vowel interaction, suggesting that the failure to learn the height-laryngeal pattern was truly due to an analytic bias.

While studies of phonotactic typology and the predictions of phonological theory make clear that relations between vowel height or vowel backness and the laryngeal features of a following stop are not possible phonological interactions, it is not the case that all vowel-consonant interactions are disfavored in natural language; on the contrary, such interactions can be quite commonplace. For example, front high vowels force a change of the place of articulation in an adjacent obstruent consonant in a number of languages, leading to phonotactic bans against sequences such as *ti*, *si*, or *ki* as opposed to *tʃi* or *ʃi*; such palatalization processes are found in Japanese, Italian, Finnish, and Korean, among many other languages (Bhat 1978, Hall & Hamann 2006). In Turkish itself, velars are fronted following front vowels (*ek^{jh}* 'affix' vs. *ak^h* 'white'). Similarly, consonants can affect the distribution of adjacent vowels, as in the case of nasalization in Brazilian Portuguese, in which a stressed vowel must be nasalized before a nasal consonant, leading to phonotactic bans against sequences such as *ana* as opposed to *ãna* (Wetzels 1997). Importantly, these cases of consonant-vowel assimilatory interactions are mediated by the fact that the phonetic feature in the consonant that triggers the change is identical to the changed feature on the vowel (or vice versa): for example, the palatal place of articulation of high front vowels is identical to the palatal place of articulation of the consonant affected by palatalization, and the phonological representation of the place of articulation of [i] and [ʃ] has been argued to be identical (Hume 1994). Similarly, nasal consonants and nasalized vowels share a common phonetic articulation, [+nasal], required in the production of sounds that allow airflow through the nose (Cohn 1993).

The cases of palatalization and nasalization discussed above are processes in which vowel-consonant interaction is mediated by a common supralaryngeal phonetic feature. There are also, in fact, cases of vowel-consonant assimilatory interactions involving laryngeal features. One such phonotactic restriction involves voicing of obstruents, in which a high tone on a vowel can affect the voicing of an adjacent consonant (i.e. a high tone on a vowel implies voiceless consonants, or vice versa), as found in Shanghaiese or Jabem (Poser 1981). This vowel-consonant phonotactic interaction, however, also involves a common phonetic feature in both the trigger and target: high tone in vowels and voicelessness in obstruents are both controlled by the laryngeal property of stiffened vocal folds (Halle & Stevens 1971, though see Tang 2008 for an alternative view).

Phonotactic interactions between vowels and consonants are thus possible and indeed quite common when the nature of the phonotactic restriction involves a phonetic feature shared by the vowel and consonant. The phonetic basis for this phonotactic interaction can be either a laryngeal feature that both the vowel and consonant share, such

as stiffened vocal folds, or a supralaryngeal feature that the vowel and consonant share, such as place of articulation in the vocal tract. The putative interaction of vowel height with consonantal laryngeal features does not even remotely fit within this rubric, however, since vowel height is a supralaryngeal feature, and the two have thus virtually nothing to do with each other, either phonetically or in terms of their phonological representations.

The SAME-FEATURE CONSTRAINT on vowel-consonant interactions is thus an ‘overhypothesis’ in the sense of Goodman 1955 and Kemp et al. 2007: a meta-level hypothesis that constrains the form of possible specific hypotheses and generalizations induced from the data. The same-feature constraint figures prominently in the phonological literature, most recently in Moreton 2010, and in earlier work such as Pycha et al. 2003, Peperkamp et al. 2006, and others. Whether the same-feature constraint on vowel-consonant phonotactics is innate, or perhaps itself induced in parallel, for example, through use of a hierarchical Bayesian model (Good 1980, Kemp et al. 2007), is not something that our experimental results speak to directly, but is an important question for modeling how it is that the vowel-quality/obstruent laryngeal feature phonotactic of Turkish is ignored.

5.3. PRIOR ANALYTIC BIASES FILTER STATISTICAL REGULARITIES. A number of current phonological theories adopt a constrained theory of possible phonological processes. Optimality theory posits a universal inventory of possible phonological interactions that can be expressed as the result of the interactions among a universal set of constraints (see Kager 1999, McCarthy 2002). Parametric models of phonological rules express constraints on what can be a possible phonological interaction as a property of the space created by a given parametric system (e.g. Dresher & Kaye 1990, Archangeli & Pulleyblank 1994, Cho 1999). Both the theories of universal constraint inventories and the theories of parameterized rules of assimilation can require that the feature dictating a vowel-consonant interaction must be shared by both the consonant and the vowel. These models thus adopt a specific set of analytic biases, often called universal grammar, that the language learner brings to the task of extracting phonotactic generalizations from the lexicon, and that constrain possible generalizations that learners will make. The possibility of consonantal laryngeal features being determined or affected by vowel height or vowel backness is excluded, or highly disfavored to the point that even significant evidence for such a relationship in the lexicon is not enough. Computational modeling studies of phonological rule induction, such as Gildea & Jurafsky 1996, have converged on the conclusion that abstract learning biases lead to more compact, more accurate, and more general finite-state transducers for generating morphophonemic alternations.

If these phonetically unmotivated patterns are never used and in fact excluded or disfavored by learning biases, why do they exist in the Turkish lexicon in the first place? The existence of a statistically significant trend for high vowels or for back vowels to be followed by alternating stops in the Turkish lexicon is arguably tied to the fact that the Turkish lexicon represents an accumulation of several centuries worth of language contact. Many of the lexical trends that were identified in our quantitative lexicon analysis are ultimately traceable to extensive lexical borrowing from Arabic, to much the same degree that many of the lexical trends found in English phonotactics, such as the existence of more words that begin with [ʤ] than [ʒ], are ultimately traceable to lexical borrowing from French centuries ago, when Old French had [ʤ] but not [ʒ] word-initially. In Turkish borrowings of words with voiced stops in the source language, final devoicing in the bare stem but not in the forms with vowel-initial suffixes causes a noun to become alternating (e.g. Arabic *burǧ* ‘sign’ > Turkish *burf*^h ~ *burǧ-u*), whereas source

words that end in a voiceless stop are nonalternating across the paradigm. Arabic lacks the consonants [p] and [tʃ] and has many nouns that end in [b] and [ɟ], and as a consequence, the lexicon's overall alternation rates are boosted for those places of articulation. In addition, the existence of many Arabic nouns with the feminine suffix *-at/-et* boosted the number of nonalternating, nonhigh vowel, coronal-final nouns. Ultimately, however, the historical explanation for these lexical trends is completely inaccessible to speakers that are not experts in historical linguistics, many of whom (like the English speakers who know the word *judge* but not its origin) do not even know that there was a source language that provided this borrowed word, well integrated into the phonotactics for centuries.

If indeed the skewed distribution of the Turkish laryngeal alternations is largely due to massive borrowing from Arabic, it is instructive that Turkish speakers synchronically generalize the historically accidental place effect, but discard the equally accidental height effect. History has dealt Turkish speakers a certain hand, and they use universal grammar to pick the cards they want to keep. This view contrasts with the proposal by Hayes (1999), who claims that when history creates nonuniversal patterns, speakers are able to complement their universal grammar with arbitrary generalizations.

In Turkish, the distribution of laryngeal alternations is not known to correlate with the native or borrowed status of roots (and as mentioned in the introduction, loanwords such as *group* > *gurub-u* conform to the polysyllabic-as-alternating generalization). Thus, the sources of some of the unprincipled statistical regularities are arguably historical in nature, yielding phonetically ungrounded synchronic patterns that are simply ignored.

The result that Turkish speakers reliably extend base rates for laryngeal alternations based on place of articulation and size of the word, but not based on preceding vowel quality, arguably due to an analytic bias against learning such arbitrary interactions, strengthens the finding of Moreton (2008) that English speakers were less successful learning an artificial language pattern with height-voicing interactions, and more successful learning nonadjacent V-V interactions, in which high vowels were followed by high vowels in the adjacent syllable. In Turkish, the case is even more striking: a lexical generalization is staring Turkish speakers in the face, but they do not generalize it productively in experimental contexts. The results provide support for an analytically biased mechanism of filtering lexical statistics, one in which phonologically implausible interactions are not actively incorporated into phonotactic knowledge. There is by now a general consensus that statistical information is indispensable in arriving at phonotactic generalizations, a fact that our experimental results confirm. At the same time, accurate models of the acquisition of phonological knowledge need to build in a set of linguistically specific priors that constrain and restrict the learning of statistical patterns. Apparently, given a surfeit of the stimulus, not every statistical fact about the lexicon is used or kept track of.

The analysis we offer in §4 uses OT constraints to organize lexical items according to their behavior, meaning that the constraints act as 'priors' on what data is to be used in forming grammatical hypotheses. This implicates an analytic bias that, in this case, ignored the correlation between vowel quality and consonantal laryngeal features thanks to the absence of constraints that relate the two, thus closely modeling the pattern produced by native speakers.

5.4. SURFEIT PHENOMENA AND NATURAL VS. UNNATURAL EFFECTS. Our results strongly suggest the existence of a surfeit-of-the-stimulus effect: a phonotactic pattern is readily available in the lexicon for speakers, and in an experimental task sensitive enough to probe such knowledge, they do not extend such a pattern to novel items.

Not all experiments of this sort have found surfeit effects. Given the publication of Hayes et al. 2009, it is worth discussing the fact that unnatural constraints are in fact sometimes learned. Hayes and colleagues, despite finding generalization of consonantal effects in Hungarian vowel harmony, nonetheless remark:

We found that unnatural constraints were underlearned, giving modest support to the idea (Wilson [2003], Moreton 2008) that people show a learning bias against unnatural constraints. We also found underlearning for some of the natural constraints, however, namely those responsible for the count effect and part of the height effect. This suggests perhaps a role for a simplicity bias as well. (Hayes et al. 2009:856)

We maintain that our findings provide another result to add to the growing pool of research in the area of ‘underlearning’ of statistical patterns, of which we provide a summary below. We note, interestingly, that many researchers have not cast their results in terms of surfeit effects, but rather as ‘failures’ to find certain phonotactic knowledge. We hope our framework can provide a lens through which to view these results as the effect of the grammar stepping in and filtering out patterns that are ‘unnatural’.

One of the earliest studies in this vein (also, as it turns out, conducted on Turkish), was by Zimmer (1969), who investigated the extension of phonotactic constraints to novel items. He found that all Turkish speakers extended a preference for labial and palatal vowel harmony to novel roots, but that not all of them demonstrated an extension of the phonotactic of ‘labial attraction’ to these novel items. Labial attraction in Turkish is the pattern whereby *aMu* sequences, where M is a labial consonant, greatly outnumber *aMi* sequences in the lexicon. This is clearly a complex and somewhat unnatural phonotactic, both in terms of the nonlocality of environment and the conjunction of features from two distinct triggers, and it is therefore a welcome result that not all speakers readily encoded it into a generalizable constraint.

In a study of laryngeal alternations in Dutch, using a methodology similar to ours, Ernestus and Baayen (2003) show that speakers project the rate of alternation of different stops based on their place of articulation, just like the Turkish speakers. Ernestus and Baayen’s (2003) report of the vowel effects is instructive: in the lexicon, stops alternate more following long vowels and less after short vowels. Following the high vowels of Dutch, which are all short, stops have an intermediate rate of alternation. In their experiment, however, speakers projected and strengthened the vowel-length effect, preferring more alternations after long vowels. Speakers did not project the vowel-height effect, choosing alternations equally frequently after short vowels that are either high or nonhigh. Given our proposal, this result is not surprising: as mentioned above, vowel height is universally not expected to interact with laryngeal features. The preference for longer vowels before voiced consonants, however, is well attested (Denes 1955, Peterson & Lehiste 1960, Chen 1970, among others). The absence of observed lengthening before voiced consonants in some languages lends support to the view that vowel lengthening before voiced stops is not an anatomical necessity, but is rather controlled by the grammar (Keating 1985, Buder & Stoel-Gammon 2002) and thus can enter into speakers’ learning of lexical trends. The Dutch results, then, strengthen the evidence for a formal bias against extending featural V-C interactions to novel items.

In a different study, Kager and Pater (2010) explore Dutch speakers’ phonotactic knowledge that long vowels are rarely followed by a consonant cluster in which the second consonant is noncoronal (e.g. *mɛlk* ‘milk’ vs. **me:lk*). While the generalization is strong in the Dutch lexicon both in monosyllables and in polysyllables, Kager and Pater found that in a novel word task, the constraint was applied more weakly to polysyllables than to monosyllables. This finding can be straightforwardly interpreted in

terms of a limit on the complexity of generalization, in this case a nonlocal interaction between vowel length and the noncoronality of a nonadjacent following consonant. A different notion of complexity arises from Zhang et al. 2006 in a study of Taiwanese tone sandhi. In this language, nonfinal syllables in compounds change tone: 51 → 55, 55 → 33, 33 → 21, and 21 → 51, effectively creating a ‘tone circle’. Zhang and colleagues (2006) find that only some of these changes, such as 33 → 21, are applied to novel nouns and novel combinations of existing nouns, while other changes, such as 51 → 55, are hardly applied at all. In this case, the complexity of the generalization is due to the tone sandhi alternations involving a circular chain shift, which is too opaque to be encoded grammatically.

More broadly, experimental work has revealed effects of phonological knowledge (or lack thereof) that is independent of (and sometimes contrary to) patterns available through frequency counts in the lexicon. Moreton (2002) found that given two consonant cluster sequences in English, *tl-* and *bw-*, both with an attested frequency near zero, speakers nonetheless greatly prefer the latter in a rating task. Davidson (2006) found that English speakers accurately produce illicit nonnative consonant clusters based on their featural composition and not on their frequency in the lexicon, showing that grammatical factors shape phonotactic knowledge.

As domain-specific biases involved in language learning, these grammatical filters on intake may not kick in, or may not even NEED to kick in, when faced with data that is so compact, or stimulus presentation that may not be recognized as part of a grammatical system. Results such as those in Onishi et al. 2003, which demonstrate participants’ ability to generalize over arbitrary patterns while encountering a small amount of data under brief exposure, may fall into the category of such cases. When miniature artificial languages have a structure in which there are very few competing hypotheses to have, and very little potentially ambiguous data, covering it with a single arbitrary generalization may suffice. Such questions—for example, how complex an artificial language, or indeed a corpus of primary language data, needs to be before analytical filters on data intake are necessary—clearly lead to many experimental possibilities for future testing: Under which conditions do humans detect a surfeit of the stimulus and deploy linguistically unnatural mechanisms to help?

Casting the failure of incorporating phonotactic knowledge as we have above may also invite revisiting past sets of experimental findings that have otherwise stayed in the proverbial file cabinet. Inspection and reflection might find that many ‘null results’ in phonotactic knowledge tasks with novel items are in fact surfeit effects. A null result in wug tests or similar tasks may reveal that experimenters were looking to find evidence for a pattern that grammars never bothered to keep. In practice, one would diagnose surfeit effects by pitting different predictions from the lexicon against each other and seeing which one is the better predictor of the experimental results. In our Turkish case, our model of the lexicon predicted an effect of vowel height and vowel backness; it was this prediction that caused the lack of effect in the experimental results to be interpretable as ‘underlearning’, or a surfeit effect. More broadly, such surfeit effects can inform our understanding of generalizations that are or are not easily formally expressible in terms of the primitives of phonological theory, either due to their naturalness and/or due to their formal simplicity.

The ability of humans and other animals to track frequency patterns in a range of modalities and domains of cognition (sequential presentation, simultaneous presentation, visual, auditory) is impressive and undisputable. Our interpretation of the findings of the current experiment, however, are that this particular skill is not freely imported

into the construction and refinement of knowledge of what constitutes a well-formed linguistic expression in a given language: each hypothesis that is adopted toward this end represents a balance of how well it covers the data and how likely it is as a hypothesis about language (its ‘prior’). As Pearl and Lidz (2009:256) observe, ‘a domain-general learning procedure can be successful ... but, crucially, only when paired with domain-specific filters on data intake’. Similarly, as the thinker Shunryu Suzuki remarked, ‘In the beginner’s mind there are many possibilities; in the expert’s mind there are few’. Humans are above all else expert language learners, and as such they do not consider every possibility when going from lexicon to grammar.

REFERENCES

- ALBRIGHT, ADAM. 2008a. A restricted model of UR discovery: Evidence from Lakhota. Cambridge, MA: MIT, MS.
- ALBRIGHT, ADAM. 2008b. Explaining universal tendencies and language particulars in analogical change. *Language universals and language change*, ed. by Jeff Good, 144–84. Oxford: Oxford University Press.
- ALBRIGHT, ADAM; ARGELIA ANDRADE; and BRUCE HAYES. 2001. Segmental environments of Spanish diphthongization. *UCLA Working Papers in Linguistics (Papers in phonology 5)* 7.117–51.
- ALBRIGHT, ADAM, and BRUCE HAYES. 2002. Modeling English past tense intuitions with minimal generalization. *Proceedings of the sixth meeting of the ACL Special Interest Group in Computational Phonology*, ed. by Michael Maxwell, 58–69. Philadelphia: Association for Computational Linguistics.
- ALBRIGHT, ADAM, and BRUCE HAYES. 2003. Rules vs. analogy in English past tenses: A computational experimental study. *Cognition* 90.119–61.
- ALBRIGHT, ADAM, and BRUCE HAYES. 2006. Modeling productivity with the gradual learning algorithm: The problem of accidentally exceptionless generalizations. *Gradience in grammar*, ed. by Gisbert Fanselow, Caroline Féry, Ralf Vogel, and Matthias Schlesewsky, 185–204. Oxford: Oxford University Press.
- ANTTILA, ARTO. 2002. Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory* 20.1–42.
- ARCHANGELI, DIANA, and DOUGLAS PULLEYBLANK. 1994. *Grounded phonology*. Cambridge, MA: MIT Press.
- AVERY, PETER. 1996. *The representation of voicing contrasts*. Toronto: University of Toronto dissertation.
- BAAYEN, R. HARALD. 2008. *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press.
- BAILEY, TODD, and ULRIKE HAHN. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44.568–91.
- BARNES, JONATHAN. 2001. Domain-initial strengthening and the phonetics and phonology of positional neutralization. Paper presented at the 32nd conference of the North East Linguistic Society (NELS), CUNY, October 2001.
- BECKER, MICHAEL. 2009. *Phonological trends in the lexicon: The role of constraints*. Amherst: University of Massachusetts, Amherst dissertation.
- BECKER, MICHAEL, and ANDREW NEVINS. 2009. Initial-syllable faithfulness as the best model of word-size effects in alternations. Paper presented at the 40th conference of the North East Linguistic Society (NELS), MIT, November 2009. Online: http://becker.phonologist.org/initialsyllfaith/becker_nevins_nels_handout.pdf.
- BECKMAN, JILL N. 1997. Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* 14.1–6.
- BECKMAN, JILL N. 1998. *Positional faithfulness*. Amherst: University of Massachusetts, Amherst dissertation.
- BECKMAN, JILL N., and CATHERINE O. RINGEN. 2004. Contrast and redundancy in OT. *West Coast Conference on Formal Linguistics (WCCFL)* 23.85–98.
- BERKO, JEAN. 1958. The child’s learning of English morphology. *Word* 14.150–77.

- BHAT, D. N. S. 1978. A general study of palatalization. *Universals of human language*, vol. 2: *Phonology*, ed. by Joseph H. Greenberg, 47–92. Stanford, CA: Stanford University Press.
- BLUST, ROBERT. 2000. Low-vowel fronting in northern Sarawak. *Oceanic Linguistics* 39.285–319.
- BOERSMA, PAUL. 2001. Phonology-semantics interaction in OT, and its acquisition. *University of Alberta Papers in Experimental and Theoretical Linguistics* 6.24–35.
- BOERSMA, PAUL, and DAVID WEENINK. 2008. Praat: Doing phonetics by computer (version 5.0.24). Online: <http://www.praat.org>.
- BONATTI, LUCA; MARCELA PEÑA; MARINA NESPOR; and JACQUES MEHLER. 2005. Linguistic constraints on statistical computations. *Psychological Science* 16.6.451–59.
- BREIMAN, LEO; JEROME FRIEDMAN; R. A. OLSHEN; and CHARLES J. STONE. 1984. *Classification and regression trees*. Belmont, CA: Wadsworth International.
- BUDER, EUGENE H., and CAROL STOEL-GAMMON. 2002. American and Swedish children's acquisition of vowel duration: Effects of vowel identity and final stop voicing. *Journal of the Acoustical Society of America* 111.1854–64.
- BURZIO, LUIGI. 2002. Surface-to-surface morphology: When your representations turn into constraints. *Many morphologies*, ed. by Paul Boucher, 142–77. Somerville, MA: Cascadilla.
- BYBEE, JOAN. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10.425–55.
- CASALI, RODERIC. 1998. *Resolving hiatus*. New York: Garland.
- CHAMBERS, J. K. 1973. Canadian raising. *Canadian Journal of Linguistics* 18.113–35.
- CHATER, NICK, and CHRISTOPHER D. MANNING. 2006. Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences* 10.7.335–44.
- CHEN, MATTHEW. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22.125–59.
- CHO, YOUNG-MEE YU. 1999. *Parameters of consonantal assimilation*. Munich: LINCOM Europa.
- COETZEE, ANDRIES W. 2008. Grammaticality and ungrammaticality in phonology. *Language* 84.218–57.
- COHN, ABIGAIL. 1993. *Phonetic and phonological rules of nasalization*. Los Angeles: University of California, Los Angeles dissertation.
- COLEMAN, JOHN, and JANET PIERREHUMBERT. 1997. Stochastic phonological grammars and acceptability. *Computational phonology: Proceedings of the third meeting of the ACL Special Interest Group in Computational Phonology*, 49–56. Somerset, NJ: Association for Computational Linguistics.
- DAELEMANS, WALTER; JAKUB ZAVREL; KO VAN DER SLOOT; and ANTAL VAN DEN BOSCH. 2002. TiMBL: Tilburg memory based learner reference guide, version 4.2. Tilburg: Computational Linguistics, Tilburg University.
- DAVIDSON, LISA. 2006. Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics* 34.104–37.
- DENES, P. 1955. Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America* 27.761–64.
- DRESHER, B. ELAN, and JONATHAN KAYE. 1990. A computational learning model for metrical phonology. *Cognition* 34.2.137–95.
- ERNESTUS, MIRIAM, and R. HARALD BAAYEN. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79.5–38.
- FLACK, KATHRYN. 2007. *The sources of phonological markedness*. Amherst: University of Massachusetts, Amherst dissertation.
- GILDEA, DANIEL, and DANIEL JURAFSKY. 1996. Learning bias and phonological-rule induction. *Computational Linguistics* 22.4.497–530.
- GOOD, IRVING JOHN. 1980. Some history of the hierarchical Bayesian methodology. *Bayesian statistics*, ed. by José M. Bernardo, Morris H. DeGroot, Dennis V. Lindley, and Adrian F. M. Smith, 489–519. Valencia: Valencia University Press.
- GOODMAN, NELSON. 1955. *Fact, fiction, and forecast*. Cambridge, MA: Harvard University Press.
- GUSSMANN, EDMUND. 1980. *Studies in abstract phonology*. Cambridge, MA: MIT Press.

- HALL, TRACY, and SILKE HAMANN. 2006. Towards a typology of stop assibilation. *Linguistics* 44.6.1195–236.
- HALLE, MORRIS, and KENNETH N. STEVENS. 1971. A note on laryngeal features. *MIT Quarterly Progress Report* 11.198–213.
- HAY, JENNIFER, and R. HARALD BAAYEN. 2005. Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences* 9.342–48.
- HAY, JENNIFER; JANET PIERREHUMBERT; and MARY BECKMAN. 2004. Speech perception, well-formedness and the statistics of the lexicon. *Phonetic interpretation: Papers in laboratory phonology 6*, ed. by John Local, Richard Ogden, and Rosalind Temple, 58–74. Cambridge: Cambridge University Press.
- HAYES, BRUCE. 1995. On what to teach the undergraduates: Some changing orthodoxies in phonological theory. *Linguistics in the morning calm* 3, ed. by Ik-Hwan Lee, 59–77. Seoul: Hanshin.
- HAYES, BRUCE. 1999. Phonological restructuring in Yidin and its theoretical consequences. *The derivational residue in phonology*, ed. by Ben Hermans and Marc van Oostendorp, 175–205. Amsterdam: John Benjamins.
- HAYES, BRUCE. 2004. Phonological acquisition in optimality theory: The early stages. *Constraints in phonological acquisition*, ed. by René Kager, Joe Pater, and Wim Zonneveld, 158–203. Cambridge: Cambridge University Press.
- HAYES, BRUCE, and ZSUZSA LONDE. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23.59–104.
- HAYES, BRUCE, and COLIN WILSON. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39.379–440.
- HAYES, BRUCE; KIE ZURAW; PÉTER SIPTÁR; and ZSUZSA LONDE. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85.822–63.
- HUME, ELIZABETH. 1994. *Front vowels, coronal consonants and their interaction in nonlinear phonology*. New York: Garland.
- INKELAS, SHARON; AYLIN KUNTAY; JOHN LOWE; CEMIL ORHAN ORGUN; and RONALD SPROUSE. 2000. Turkish Electronic Living Lexicon (TELL). Online: <http://linguistics.berkeley.edu/TELL/>.
- INKELAS, SHARON, and CEMIL ORHAN ORGUN. 1995. Level ordering and economy in the lexical phonology of Turkish. *Language* 71.763–93.
- INKELAS, SHARON; CEMIL ORHAN ORGUN; and CHERYL ZOLL. 1997. The implications of lexical exceptions for the nature of the grammar. *Derivations and constraints in phonology*, ed. by Iggy Roca, 393–418. Oxford: Clarendon.
- ITÔ, JUNKO, and ARMIN MESTER. 1995. The core-periphery structure of the lexicon and constraints on reranking. *Papers in optimality theory* (University of Massachusetts occasional papers in linguistics 18), ed. by Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, 181–210. Amherst: GLSA Publications.
- IVERSON, GREGORY, and JOSEPH SALMONS. 2007. Domains and directionality in the evolution of German final fortition. *Phonology* 24.121–45.
- JANNEDY, STEFANIE. 1995. Gestural phasing as an explanation for vowel devoicing in Turkish. *Ohio State University Working Papers in Linguistics* 45.56–84.
- JAROSZ, GAJA. 2006. *Rich lexicons and restrictive grammars: Maximum likelihood learning in optimality theory*. Baltimore: Johns Hopkins University dissertation.
- KAGER, RENÉ. 1999. *Optimality theory*. Cambridge: Cambridge University Press.
- KAGER, RENÉ, and JOE PATER. 2010. Phonotactics as phonology: Knowledge of a complex constraint in Dutch. Utrecht: University of Utrecht, and Amherst: University of Massachusetts, Amherst, MS. Online: <http://people.umass.edu/pater/kager-pater.pdf>.
- KALLESTINOVA, ELENA. 2004. Voice and aspiration of stops in Turkish. *Folia Linguistica* 38.117–43. (Special issue on voice, ed. by Grzegorz Dogil.)
- KEATING, PATRICIA. 1985. Universal phonetics and the organization of grammars. *Phonetic linguistics: Essays in honor of Peter Ladefoged*, ed. by Victoria A. Fromkin, 115–32. New York: Academic Press.
- KEMP, CHARLES; AMY PERFORIS; and JOSHUA TENENBAUM. 2007. Learning overhypotheses with hierarchical Bayesian models. *Developmental Science* 10.3.307–21.
- KETREZ, NIHAN. 2007. Alignment versus linearity constraints in a Turkish child's speech. *Dilbilim Araştırmaları* 2007.77–88.
- KINGSTON, JOHN. 2002. Keeping and losing contrasts. *Berkeley Linguistics Society* 28.155–76.

- KIRCHNER, ROBERT. 1998. *An effort-based approach to consonant lenition*. Los Angeles: University of California, Los Angeles dissertation.
- KOPKALLI, HANDAN. 1993. *A phonetic and phonological analysis of final devoicing in Turkish*. Ann Arbor: University of Michigan dissertation.
- LEES, ROBERT. 1961. *The phonology of Modern Standard Turkish*. Bloomington: Indiana University Press.
- LEWIS, GEOFFREY L. 1967. *Turkish grammar*. Oxford: Clarendon.
- LISKER, LEIGH, and ARTHUR ABRAMSON. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20.384–422.
- LUCE, PAUL A., and DAVID B. PISONI. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19.1–36.
- MCCARTHY, JOHN J. 2002. *A thematic guide to optimality theory*. Cambridge: Cambridge University Press.
- MERCHANT, NAZARRÉ. 2008. *Discovering underlying forms: Contrast pairs and ranking*. New Brunswick, NJ: Rutgers University dissertation.
- MORETON, ELLIOTT. 2002. Structural constraints in the perception of English stop-sonorant clusters. *Cognition* 84.55–71.
- MORETON, ELLIOTT. 2008. Analytic bias and phonological typology. *Cognition* 25.83–127.
- MORETON, ELLIOTT. 2010. Connecting paradigmatic and syntagmatic simplicity bias in phonotactic learning. Talk given at MIT, April 9.
- MORETON, ELLIOTT, and ERIK THOMAS. 2007. Origins of Canadian raising in voiceless-coda effects: A case study in phonologization. *Laboratory phonology 9*, ed. by Jennifer Cole and José Ignacio Hualde, 37–64. Berlin: Mouton de Gruyter.
- NAKIPOĞLU, MINE, and NIHAN KETREZ. 2006. Children's overregularizations and irregularizations of the Turkish aorist. *Proceedings of the Boston University Conference on Language Development* 30.399–410.
- NAKIPOĞLU, MINE, and ASLI UNTAK. 2006. What does the acquisition of stems that undergo phonological alternation reveal about rule application? Paper presented at the International Conference on Turkish Linguistics, Uppsala, Sweden.
- NICOLAE, ANDREEA, and ANDREW NEVINS. 2010. Underlying laryngeal specifications, fricative alternations, and word-size effects. Cambridge, MA: Harvard University, MS.
- OHALA, JOHN. 1983. The origin of sound patterns in vocal tract constraints. *The production of speech*, ed. by Peter F. MacNeilage, 189–216. New York: Springer.
- ONISHI, KRISTINE; KYLE CHAMBERS; and CYNTHIA FISHER. 2003. Learning phonotactic constraints from brief auditory experience. *Cognition* 83.B13–B23.
- PATER, JOE. 2006. The locus of exceptionality: Morpheme-specific phonology as constraint indexation. *Papers in optimality theory 3* (University of Massachusetts occasional papers in linguistics 32), ed. by Leah Bateman and Adam Werle, 1–36. Amherst: GLSA Publications.
- PATER, JOE. 2009. Morpheme-specific phonology: Constraint indexation and inconsistency resolution. *Phonological argumentation: Essays on evidence and motivation*, ed. by Steve Parker, 1–33. London: Equinox.
- PEARL, LISA, and JEFFREY LIDZ. 2009. When domain-general learning fails and when it succeeds: Identifying the contribution of domain specificity. *Language Learning and Development* 5.235–65.
- PEPERKAMP, SHARON; KATRIN SKORUPPA; and EMMANUEL DUPOUX. 2006. The role of phonetic naturalness in phonological rule acquisition. *Proceedings of the Boston University Conference on Language Development* 30.464–75.
- PETERSON, GORDON E., and ILSE LEHISTE. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32.693–703.
- PETROVA, OLGA; ROSEMARY PLAPP; CATHERINE RINGEN; and SZILÁRD SZENTGYÖRGYI. 2006. Voice and aspiration: Evidence from Russian, Hungarian, German, Swedish, and Turkish. *The Linguistic Review* 23.1–35.
- POSER, WILLIAM J. 1981. On the directionality of the tone-voice correlation. *Linguistic Inquiry* 12.483–88.
- PRINCE, ALAN. 2002. Arguing optimality. *Papers in optimality theory 2* (University of Massachusetts occasional papers in linguistics 26), ed. by Andries Coetzee, Angela Carpenter, and Paul de Lacy, 269–304. Amherst: GLSA Publications.
- PRINCE, ALAN, and PAUL SMOLENSKY. 2004 [1993]. *Optimality theory: Constraint interaction in generative grammar*. Oxford: Blackwell.

- PYCHA, ANNE; SHARON INKELAS; and RONALD SPROUSE. 2007. Morphophonemics and the lexicon: A case study from Turkish. *Experimental approaches to phonology*, ed. by Maria-Josep Solé, Patrice Speeter Beddor, and Manjari Ohala, 369–85. Oxford: Oxford University Press.
- PYCHA, ANNE; PAWEŁ NOWAK; EURIE SHIN; and RYAN SHOSTED. 2003. Phonological rule-learning and its implications for a theory of vowel harmony. *West Coast Conference on Formal Linguistics (WCCFL)* 22.423–35.
- QUINLAN, J. ROSS. 1993. *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufman.
- R DEVELOPMENT CORE TEAM. 2007. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
- SAFFRAN, JENNY R. 2003. Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science* 12.4.110–14.
- SAFFRAN, JENNY R., and ERIK D. THIESSEN. 2003. Pattern induction by infant language learners. *Developmental Psychology* 39.3.484–94.
- SEZER, ENGIN. 1981. The k/∅ alternation in Turkish. *Harvard studies in phonology*, ed. by George N. Clements, 354–82. Bloomington: Indiana University Linguistics Club.
- SOBEL, DAVID; JOSHUA TENENBAUM; and ALISON GOPNIK. 2004. Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science* 28.303–33.
- STEVENS, ALAN M. 1968. *Madurese phonology and morphology*. (American oriental series 52.) New Haven, CT: American Oriental Society.
- TANG, KATRINA. 2008. *The phonology and phonetics of consonant-tone interaction*. Los Angeles: University of California, Los Angeles dissertation.
- TESAR, BRUCE. 1998. Using the mutual inconsistency of structural descriptions to overcome ambiguity in language learning. *North East Linguistic Society (NELS)* 28.469–83.
- TESAR, BRUCE. 2006. Learning from paradigmatic information. *North East Linguistic Society (NELS)* 36.619–38.
- TESAR, BRUCE, and ALAN PRINCE. 2006. Using phonotactics to learn phonological alternations. *Chicago Linguistic Society* 39.2.241–69.
- TESAR, BRUCE, and PAUL SMOLENSKY. 1998. Learnability in optimality theory. *Linguistic Inquiry* 29.229–68.
- TESAR, BRUCE, and PAUL SMOLENSKY. 2000. *Learnability in optimality theory*. Cambridge, MA: MIT Press.
- USSISHKIN, ADAM, and ANDREW WEDEL. 2009. Lexical access, effective contrast, and patterns in the lexicon. *Perception in phonology*, ed. by Paul Boersma and Silke Hamann, 267–92. Berlin: Mouton de Gruyter.
- VAUX, BERT, and BRIDGET SAMUELS. 2005. Laryngeal markedness and aspiration. *Phonology* 22.395–436.
- VOLATIS, LYDIA, and JOANNE MILLER. 1992. Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America* 92.723–35.
- WARKER, JILL, and GARY DELL. 2006. Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory and Cognition* 32.2.387–98.
- WEDEL, ANDREW. 2002. Phonological alternation, lexical neighborhood density and markedness in processing. Handout from talk given at LabPhon 8, Yale University.
- WETZELS, LEO. 1997. The lexical representation of nasality in Brazilian Portuguese. *Probus* 9.2.203–32.
- WILSON, STEPHEN. 2003. A phonetic study of voiced, voiceless and alternating stops in Turkish. *CRL Newsletter* 15.3–13. Online: <http://stephenw.bol.ucla.edu/papers/turkishphon.pdf>.
- ZHANG, JIE; YUWEN LAI; and CRAIG TURNBULL-SAILOR. 2006. Wug-testing the 'tone circle' in Taiwanese. *West Coast Conference on Formal Linguistics (WCCFL)* 25.453–61.
- ZIMMER, KARL. 1969. Psychological correlates of some Turkish morpheme structure conditions. *Language* 45.309–21.
- ZIMMER, KARL, and BARBARA ABBOTT. 1978. The k/∅ alternation in Turkish: Some experimental evidence for its productivity. *Journal of Psycholinguistic Research* 7.35–46.

- ZOLL, CHERYL. 1996. *Parsing below the segment in a constraint-based framework*. Berkeley: University of California, Berkeley dissertation.
- ZURAW, KIE. 2000. *Patterned exceptions in phonology*. Los Angeles: University of California, Los Angeles dissertation.

Becker
Harvard University
Department of Linguistics
Boylston Hall, Third Floor
Cambridge, MA 02138
[becker@fas.harvard.edu]

[Received 21 November 2008;
revision invited 13 October 2009;
revision received 5 August 2010;
accepted 13 October 2010]

Ketrez
Istanbul Bilgi University
Department of Comparative Literature
Silahtaraga Mah.
Kazim Karabekir Cad. No. 1
34060 Eyüp, Istanbul, Turkey
[fnketrez@bilgi.edu.tr]

Nevins
University College London
Division of Psychology and Language Sciences
Chandler House 115a
2 Wakefield Street
London, WC1N 1PF, United Kingdom
[a.nevins@ucl.ac.uk]