The Surfeit of the Stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations

Michael Becker
Harvard University

Nihan Ketrez
İstanbul Bilgi University

Andrew Nevins
University College London

July 31, 2010

Abstract

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 nonce 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, 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.

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 et al. (2005) found that adult learners exposed to artificial grammars were much better at extracting transitional probability regularites over consonants than equally matched transitional probabilites over vowels, suggesting that learners preferentially pay more attention to statistics within consonantal frames. In a study of infant learning of phonotactic patterns, Saffran & 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 class, again suggesting that statistical learning may be less efficient when the regularities are inconsistent with natural language structure.

In this paper, 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 judgement 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 voiceless aspirated stop in their bare form, such as the pre-palatal stop [t͡ʃ^h], can either retain that [t͡ʃ^h] in the possessive (1a-b), or the [t͡ʃ^h] of the bare stem may alternate with the voiced [d͡ʒ] in the possessive (1c-d).[1]

(1)

	bare stem	possessive	
a.	at͡ʃ^h	at͡ʃ^h-ɨ	'hunger'
b.	anat͡ʃ^h	anat͡ʃ^h-ɨ	'female cub'
c.	tat͡ʃ^h	tad͡ʒ-ɨ	'crown'
d.	amat͡ʃ^h	amad͡ʒ-ɨ	'target'

Turkish exhibits a contrast between the voiced stops [b, d, d͡ʒ, ɡ] and the voiceless aspirated stops [p^h, t^h, t͡ʃ^h, k^h] in onset position, e.g. *t*^{h}*er* '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, e.g. *rop*^{h} 'dress' $<$ French *robe*. Voiced coda stops are allowed in the initial syllable of the word, e.g. *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, e.g. *d͡ʒo.p*^{h} 'club' vs. the possessed form *d͡ʒo.**b**-u* 'club.3SG'; however, when followed by a consonant-initial suffix, the final stop remains in coda position and thus stays voiceless: *d͡ʒop*^{h}*.-lar* 'club.PL'. This alternation occurs in 54% of the nouns of the language (Inkelas et al. 2000), and applies productively to loanwords, e.g. *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, e.g. *sop*^{h} $\sim$ *so.p*^{h}*-u* 'clan.3SG' $\sim$ *sop*^{h}*.-lar* 'clan.PL'.

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

The velar stops [k^h, ɡ] contrast in onset position, e.g. *so.**k**^{h}ak^h* 'street' vs. *ga.***g***a* 'beak'. In word-final position, they neutralize to the voiceless stop [k^h]. While post-consonantal dorals, as in *renk^h* ∼ *ren***g***-i* 'color', display the general process of laryngeal alternation, intervocalic velar stops undergo deletion rather than voicing, i.e. when nouns ending in *postvocalic* velar stop are followed by a vowel-initial suffix, the velar stop deletes e.g. *ete**k**^{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 will be shown below, whether a noun stem shows the k/∅ alternation or not is correlated 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 non-alternating stops is traditionally captured within generative phonology as the difference between an underlying voiced stem-final stop in the case of *d͡ʒop^h* ∼ *d͡ʒob-u* and an underlying voiceless stem-final stop in the case of *sop^h* ∼ *sop^h-u*, with the underlying contrast being neutralized in word-final coda position (Lees 1961). While the difference between alternating and non-alternating nouns may be captured in a variety of alternate theoretical frameworks which do not incorporate underlying representations (e.g. via reference to identity-relations vs. lack thereof 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 doesn't.

Whether the final stop of a given noun will or will not alternate is unpredictable. However, the noun's size strongly correlates with its status: 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 [t͡ʃ^h], for instance, correlates with more alternations, but Turkish speakers seem to ignore this correlation. These language-specific patterns can be understood given a cross-linguistic 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 cross-linguistic 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

[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^{h}ak^h* and *gaga* are protected in their non-derived environment.

[3] Our account contrasts with the one in Wedel (2002); Ussishkin & Wedel (to appear), who suggest that the source of the size effect is in neighborhood density. This argument is effectively refuted by Pycha et al. (2007) and by Becker & Nevins (2009), who show that similarity-based measurements such as neighborhood density do not correlate with alternation rates.

features [voice] and [s.g.] in initial syllables is not only seen in laryngeal alternations. Generally in the language, a coda stop followed by an onset stop will surface with the laryngeal features of the onset stop (e.g. *is.t*^{h}***ib**.dat* 'despotism', **is.t*^{h}***ip***$^{\boldsymbol{h}}$*.dat*), but a coda stop in the initial syllable may surface with its independent laryngeal specification (e.g. *a**p***$^{\boldsymbol{h}}$*.t*^{h}*al* 'fool' vs. *e**b**.k*^{h}*em* 'mute').

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.

In Optimality Theory (Prince & Smolensky 1993/2004), 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 [±back] next to some value of [±voice] or [±s.g.], e.g. *[+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 1993/2004, and in the context of learning by 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 Hayes & Wilson 2008; Flack 2007; 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 paper, 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. Pater 2006, 2009; Anttila 2002; Inkelas et al. 1997; Itô & Mester 1995; Coetzee 2008). A noun with a non-alternating final stop, like *anat͡ʃ*^{h} ∼ *anat͡ʃ*^{h}*-ɨ*, is associated with the ranking IDENT(lar) ≫ *Vt͡ʃV, meaning that faithfulness to laryngeal features outweighs the markedness pressure against voiceless intervocalic palatal stops. A noun with a final alternating stop, like *amat͡ʃ*^{h} ∼ *amad͡ʒ-ɨ*, is associated with the opposite ranking, i.e. *Vt͡ʃV ≫ IDENT(lar). This assumes that the final stop in *amat͡ʃ*^{h} is underlyingly voiceless and aspirated, and that it surfaces unfaithfully in *amad͡ʒ-ɨ*, 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 *at͡ʃʰ* ∼ *ad͡ʒ-ɨ*, can be recorded separately from the pattern of poly-syllabic nouns, by using a faithfulness constraint that protects the laryngeal features of stops in the base's initial syllable, IDENT(lar)$_{\sigma 1}$. 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. On the other hand, 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, does find support in previous work, which we discuss in §5.

This paper offers empirical data that bears on the relation between the projection of language-specific lexical trends and cross-linguistic patterns of phonological interactions, 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 mono-syllabic 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 non-alternators (2c), which we call "vacillators". Our statistical analysis treats them as intermediate between alternators and non-alternators, although in reality it is possible that their actual rate of alternation is different from 50%.

(2)

	Bare stem	Possessive	
a.	at͡ʃʰ	at͡ʃʰ-ɨ	'hunger'
b.	amat͡ʃʰ	amad͡ʒ-ɨ	'target'
c.	ɡyvet͡ʃʰ	ɡyvet͡ʃʰ-i, ɡyved͡ʒʰ-i	'cooking pot'

Several phonological properties of Turkish nouns will be discussed, showing that four of them correlate with stem-final alternations: (a) the noun's size (mono-syllabic vs. poly-syllabic), (b) the place of articulation of the stem-final stop, (c) the height of the vowel that precedes the stem-final stop, and (d) the backness of that vowel.

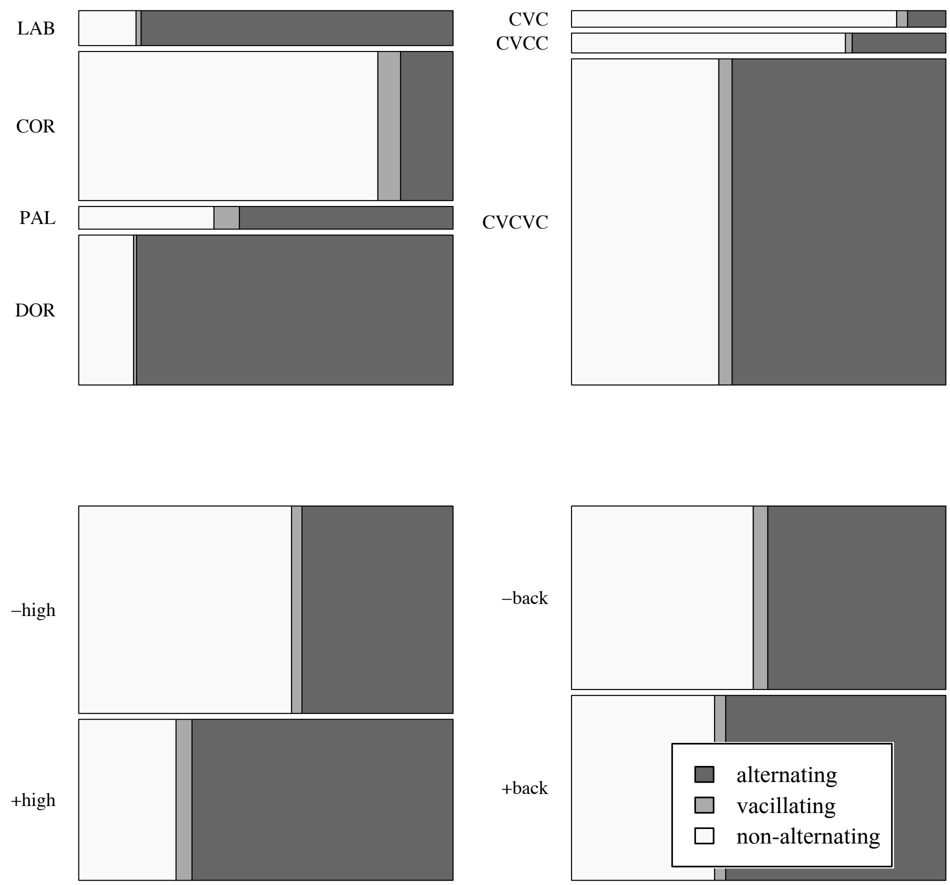
TELL lists a total of about 30,000 nouns, verbs, and adjectives that were collected from a variety of extant dictionaries. Of these, 18,000 were produced and transcribed by a native speaker in various inflected forms. Nouns are listed with their bare citation forms and with four suffixed forms (1.SG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

2.1. EXPLORATION OF LARYNGEAL ALTERNATION PREDICTORS IN TELL

We offer in (3) a closer look at the 3002 nouns in TELL whose bare stem ends in a voiceless stop. The first panel 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 3002 stop-final nouns of TELL.

[4]Dorsals delete post-vocalically, and voice post-consonantally; see §1 above for discussion.

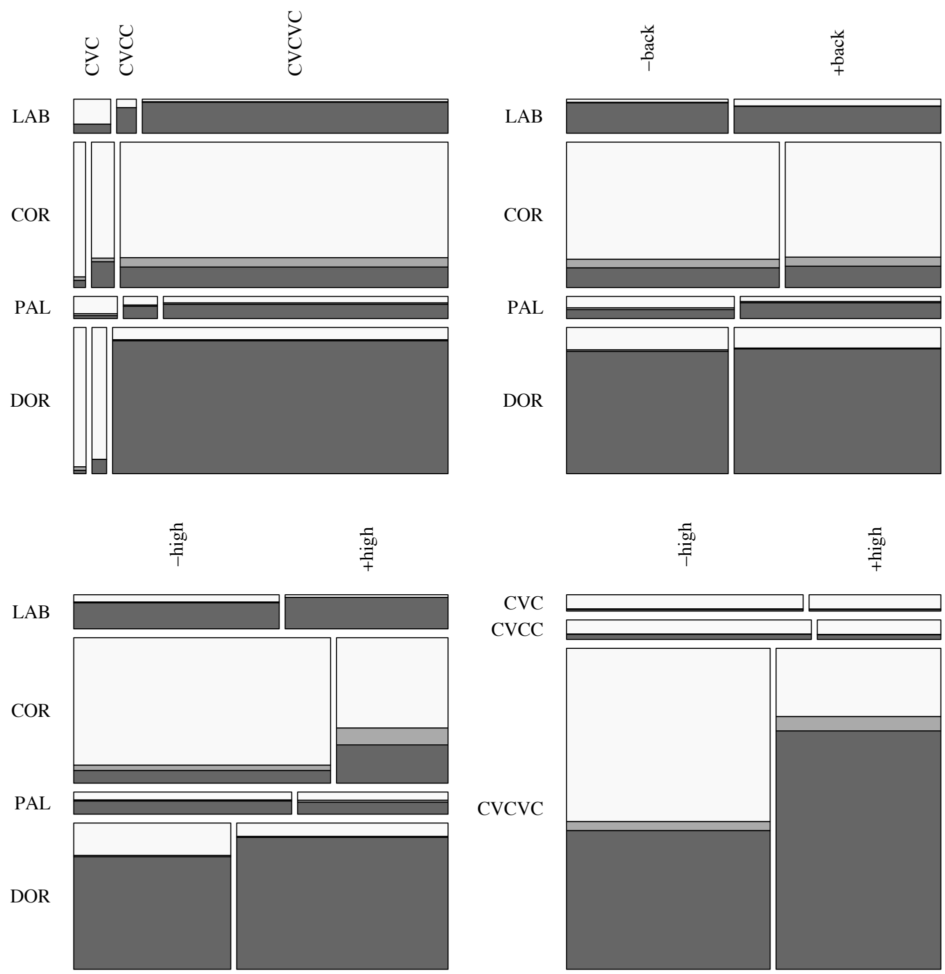
(3) Alternation rates in the lexicon, by single features



The second panel in (3) 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). The final two panels 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 cross-linguistically, vowel quality is not known to influence the laryngeal features of a neighboring obstruent.

The interaction of *size* and *place* is plotted in the first panel of (4). In all places, 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 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 will be seen that indeed speakers treat each place/size combination separately.

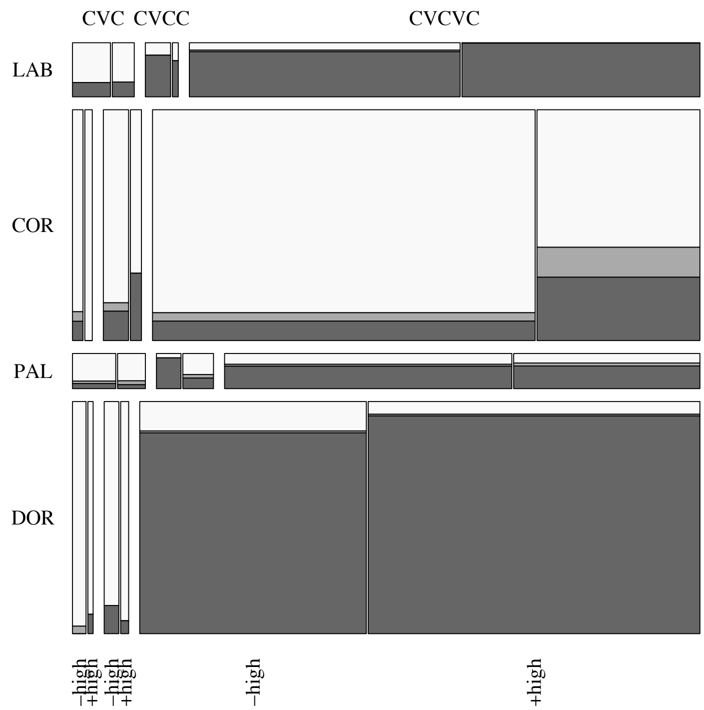
(4) Alternation rates in the lexicon, by pairs of features



The second panel in (4) 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. The last two panels 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 (5) a mosaic plot for the interaction of *place*, *size*, and *high*. It can be seen that the strongest height effect is within the *t*-final polysyllables.

(5) Alternation rates in the lexicon: place, size, high



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 non-round vowel.

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 non-high 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 deeper insight when considered in pairs: Size and place have a non-uniform interaction, with CVCC words behaving like CVC words when dorsal-final and like CVCVC words when labial- or palatal-final. Height and backness interact with place non-uniformly: 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 don't alternate, nouns that do, and "vacillators", i.e. nouns that allow either alternation or non-alternation), 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 non-alternation 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 mono-syllables with a simplex coda (CVC), mono-syllables with a complex codas (CVCC), and poly-syllables (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.
- High, back and round: each of the three features of the stem-final vowel was encoded as two-level unordered factor. The base levels chosen were non-high, 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 (6). This power is measured by Somers' D_{xy} and by the model's likelihood ratio (Model L.R.), which comes with a number of degrees of freedom and a p-value (Baayen 2008; p. 203ff). It turns out that *place*, *high*, *size*,[6] and *back* are highly predictive of alternations, in

[5]An anonymous reviewer asks whether we could have performed 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 this 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.

[6]We have 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 excluded it from the following presentation. 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 ($p = .03$). 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.

that order, and *round* is not.[7]

(6) Strength of individual predictors in TELL

	D_{xy}	Model L.R.	df	p
place	.66	1469	3	$<.001$
high	.29	284	1	$<.001$
size	.14	193	2	$<.001$
back	.11	37	1	$<.001$
round	0	0	1	$>.1$

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 interaction of *place* with each of *size*, *high*, and *back* were tested in separate models, summarized in (7).

(7) Strength of two-way interactions in TELL

	D_{xy}	Model L.R.	df	p
place*size	.74	1920	11	$<.001$
place*high	.73	1621	7	$<.001$
place*back	.67	1496	7	$<.001$

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 (7) as predictors reaches a D_{xy} of .81, with a model L.R. of 2078 for 19 degrees of freedom. This final model is given in (8).

The final model in (8) was validated with the fast backwards 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 5 of the 7 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 (8), with slight improvements of the p-values of the vowel-place interactions at the fourth decimal place.

The model in (8) 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

[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 (6), assigning the five predictors the following information gain values, respectively: .367, .071, .047, .009 and .0004.

alternate relative to the arbitrarily chosen baseline levels of the predictors, namely CVC size, dorsal place, non-high 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 doesn't reach significance. We will see in §3 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.

(8) Final regression model for laryngeal alternations in TELL.
Significant predictors are in boldface.

	Coef β	SE(β)	Wald z	p
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**

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 socio-economic background. The *validate()* function that we applied to the model in (8) 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 highly 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 the first two of these, from a crosslinguistic perspective are more likely than the other two to have a causal relationship with a stop's laryngeal features.

2.3. Non-phonological 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 poly-morphemic 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 non-alternating nouns contain the native nominalizer *–It* or the Arabic feminine suffix *–At*.[8] Similarly, the suffixes *–lIk*, *–Ik*, and *–Ak* all yield alternating *k*-final nouns. It is possible, in principle, that the height effect is due to an abundance of non-alternating non-high *–At* and/or an abundance of alternating high *–lIk*. This is not a concern for the backness effect, as the known $\widehat{tʃ}$-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 extracted all 1635 of the stop-final words from it that were identifiable as mono-morphemic 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 (8) 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 highly significant, 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 mono-morphemic 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 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 (8) 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

[8] Here, "A" stands for a non-high vowel that receives its backness and rounding from the preceding syllables, while "I" stands for a high vowel that does the same.

alternators. However, all of the other trends in the lexicon remained unchanged, 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 paper. This is in line with findings in Bybee (1995); Albright & Hayes (2002); 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 non-alternators. 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 non-alternating neighbors correlating with less alternations. We also added the various combinations of the three predictors to the model in (8), with 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 the these models, all of our grammatical predictors remained undisturbed, with the exception of *height*, which was still highly significant as a main effect in both of the models we mentioned, but the interaction of *height* with coronal place lost its significance in the second one. In both models, the *back* effect remained as in (8).

To summarize, we note that the grammatical effects we are interested in are impressively 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, but not all had both. The backness effect and its interaction with palatal place remained significant in all 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 in this paper.

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, (6) 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^{h}em* ‘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 *so**p**^{h}-u* ‘clan’ (as opposed to *guru**b**-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. Cross-linguistically, 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 & 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, on the other hand, 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 (e.g. 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 of *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 non-alternating 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 will turn 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 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 non-alternating 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 *t͡ʃop*^{h} does:

its possessed form is *d͡ʒo**b**-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, as the distinction between alternating and non-alternating 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 non-alternating kind.

This *deneutralization* task shows a number of parallels with more general schema of *backwards blocking* inference, discussed in the literature on causal reasoning and inductive inference. In studies on backwards 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 backwards 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 result 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 et. al's Experiment 3, they showed that 4-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, they 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, 1 out of 10 of the objects that the participants were exposed to beforehand were blickets. In the "common blicket" condition, 9 out of 10 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 4-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 backwards-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$^{\text{h}}$] in final position) which 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%. However, when Turkish speakers are presented with a word like *zop*^{h} and asked whether to judge whether the deneutralized form should be *zo**p***^{h}*-u* or *zo**b**-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 ($n = 24$; 13 males, 11 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 a web 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-20s) 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 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 *k*'s, which were deleted.

Creating stimuli that exemplify all size, place and vowel quality combinations would have come up to 96 (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 12 words of each place, compared to 24 for the coronal- and dorsal-final words. The total number of stimuli, then, was 72 (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, non-high 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 (9). The non-final consonants were chosen such that the resulting nouns all sounded plausibly native, with neighborhood densities equalized among the stimuli as much as possible.

Finally, 36 fillers were included. All 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* ∼ *ruːh-u* 'spirit') and vowel-∅ alternations (e.g. *burun* ∼ *burn-u* 'nose'). Eighteen fillers displayed vowel-length alternations with a CVC base, and the other eighteen displayed vowel-∅ 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.

(9) Experiment items

			CVC		CVCC		CVCVC	
			−high	+high	−high	+high	−high	+high
p/t͡ʃ	−round	−back	gep^{h}	jit͡ʃ^{h}	t^{h}elp^{h}	gint͡ʃ^{h}	hevet͡ʃ^{h}	d͡ʒisip^{h}
		+back	dap^{h}	nɨt͡ʃ^{h}	p^{h}ant͡ʃ^{h}	dɨrp^{h}	jɨjap^{h}	ma.ɨt͡ʃ^{h}
	+round	−back	k^{h}øt͡ʃ^{h}	zyp^{h}	jønt͡ʃ^{h}	k^{h}yrp^{h}		bølyt͡ʃ^{h} t^{h}yryt͡ʃ^{h}
		+back	p^{h}ot͡ʃ^{h}	t^{h}up^{h}	solp^{h}	munt͡ʃ^{h}		k^{h}onup^{h} gujup^{h}
t	−round	−back	p^{h}et^{h}	hit^{h}	zelt^{h}	t͡ʃ^{h}int^{h}	nik^{h}et^{h}	gevit^{h}
		+back	fat^{h}	mɨt^{h}	hant^{h}	ʃɨrt^{h}	ja.at^{h}	p^{h}ɨsɨt^{h}
	+round	−back	søt^{h}	d͡ʒyt^{h}	gønt^{h}	nyrt^{h}		sølyt^{h} bynyt^{h}
		+back	jot^{h}	nut^{h}	d͡ʒolt^{h}	bunt^{h}		t͡ʃ^{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	d͡ʒak^{h}	p^{h}ɨk^{h}	vank^{h}	nɨrk^{h}	t^{h}at^{h}ak^{h}	banɨk^{h}
	+round	−back	høk^{h}	syk^{h}	sønk^{h}	p^{h}yrk^{h}		nønyk^{h} dyjyk^{h}
		+back	mok^{h}	nuk^{h}	bolk^{h}	dunk^{h}		zoruk^{h} juluk^{h}

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 non-finally in a few of the filler items, no problems were found with the stimuli.

PROCEDURE: Before the beginning of the experiment, speakers were reminded that laryngeal alternations are lexically-specific by presenting a familiar non-alternating paradigm (*t*^{h}*op*^{h} ∼ *t*^{h}*op*^{h}*-u* 'ball') next to a familiar alternating paradigm (*d͡ʒep*^{h} ∼ *d͡ʒeb-i* 'pocket'). Then, speakers were asked to choose the possessive form of two familiar alternating nouns (*dolap*^{h} 'cupboard' and *a.at͡ʃ*^{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. *fe**t***^{h}*-i* and *fe**d**-i*. Speakers pressed "F" or "J" to choose the first or the second possessive form they heard. Most speakers took 15-20 minutes to complete the experiment.

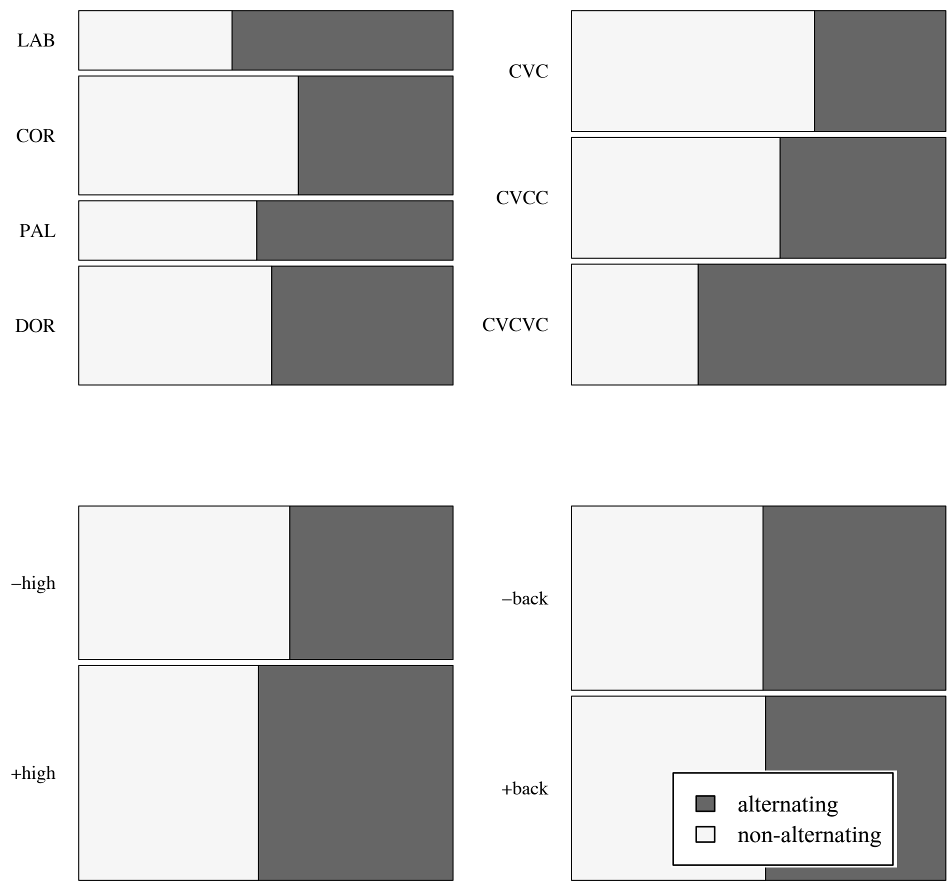
The order of the stimuli and the order of the choices were randomized. Additionally, the fillers were randomly distributed among the first three quarters of the stimuli.

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 panels in (10), which parallel the panels in (3), 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 the third panel, this is primarily due to the larger number of high vowels in the CVCVC stimuli, due to the phonotactic absence of {o, ø} in non-initial syllables. There is no difference to speak of between front and back vowels.

(10) Alternating choices for nonce words, by single features

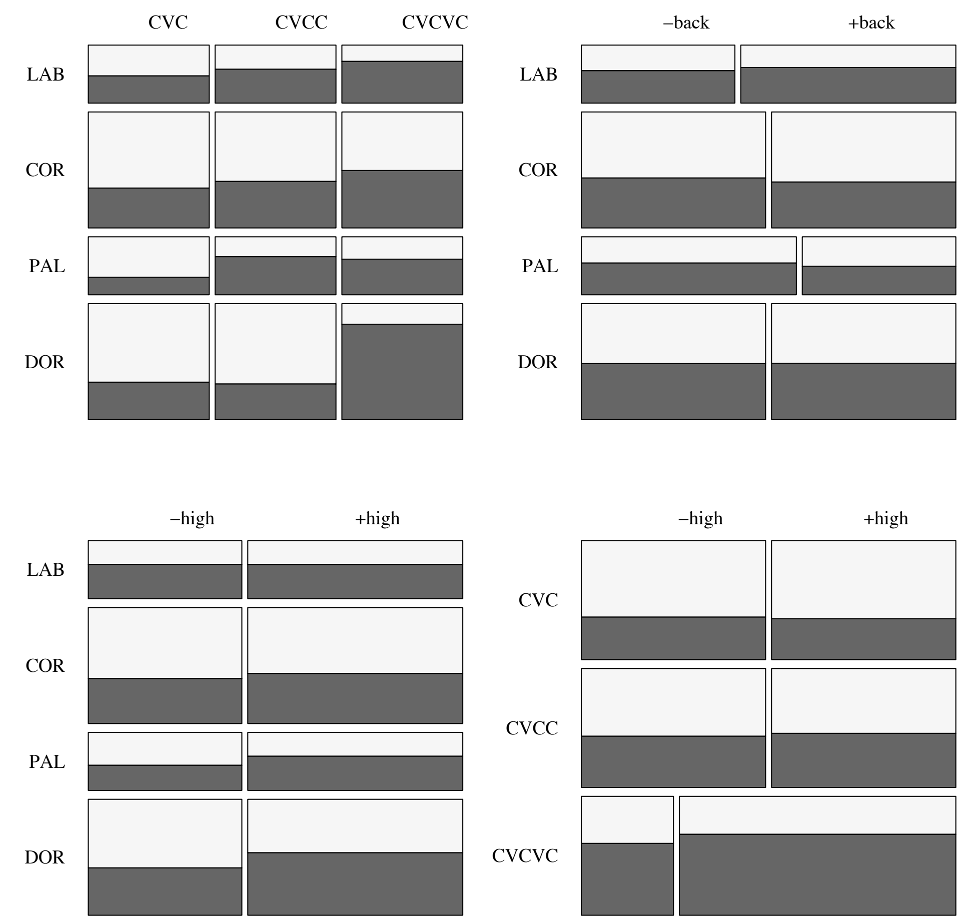


The interaction in the first panel in (11), which parallels the one in (4), shows a clear gradation for size for labials and coronals. For palatals, we see CVCC nouns patterning with CVCVC nouns, while for dorsals, the CVCC nouns pattern with the CVC nouns.

The backness effect, or rather its absence, can be seen in the second panel. In the lexicon the backness effect was concentrated in the palatals, whereas in the experimental results the effect is not simply absent in the palatals, it is reversed.

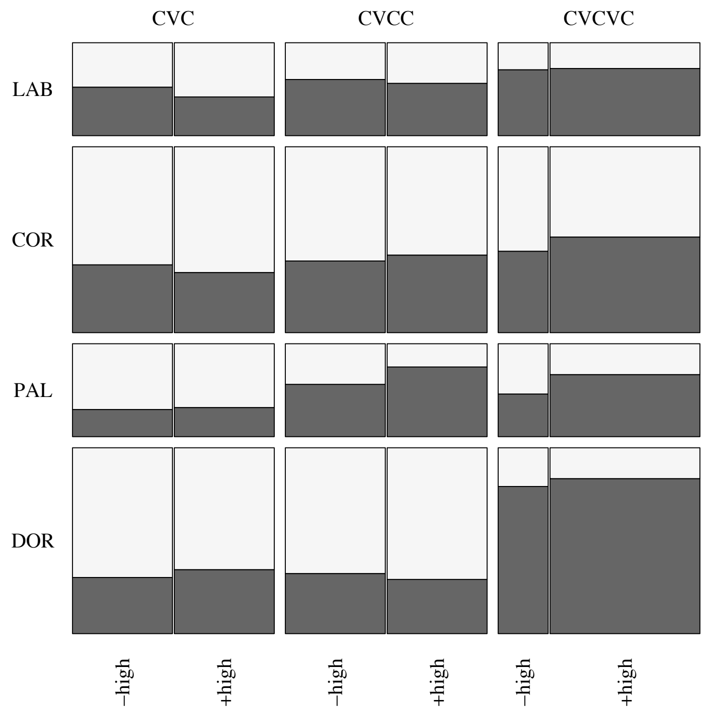
The third panel shows a small preference for alternations in high vowels in the palatals and dorsals, contrary to the lexicon, where the largest height effect was in the coronals. The fourth panel shows a slight preference for alternations after high vowels in CVCVC nouns.

(11) Alternating choices for nonce words, by pairs of features



Finally, the chart in (12), which parallels the one in (5), 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.

(12) Alternating choices for nonce words: place, size, high



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 *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, $\chi^2(6) = 50.58$, $p < .001$). The improved model with the interaction term is given in (13). 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 (8). 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 (8). 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 (8).

(13) Final regression model for alternating choices of nonce words

	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**

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 (13) yields a new model where the interaction of coronal place and *high* is almost exactly at chance level ($p =$ .981). Adding *place**back* the to baseline model gives an interaction of palatal place and *back* that is non-significant ($p = .661$) and its coefficient is negative, i.e. going in the opposite direction from the lexicon, where palatal place and backness are positively correlated.

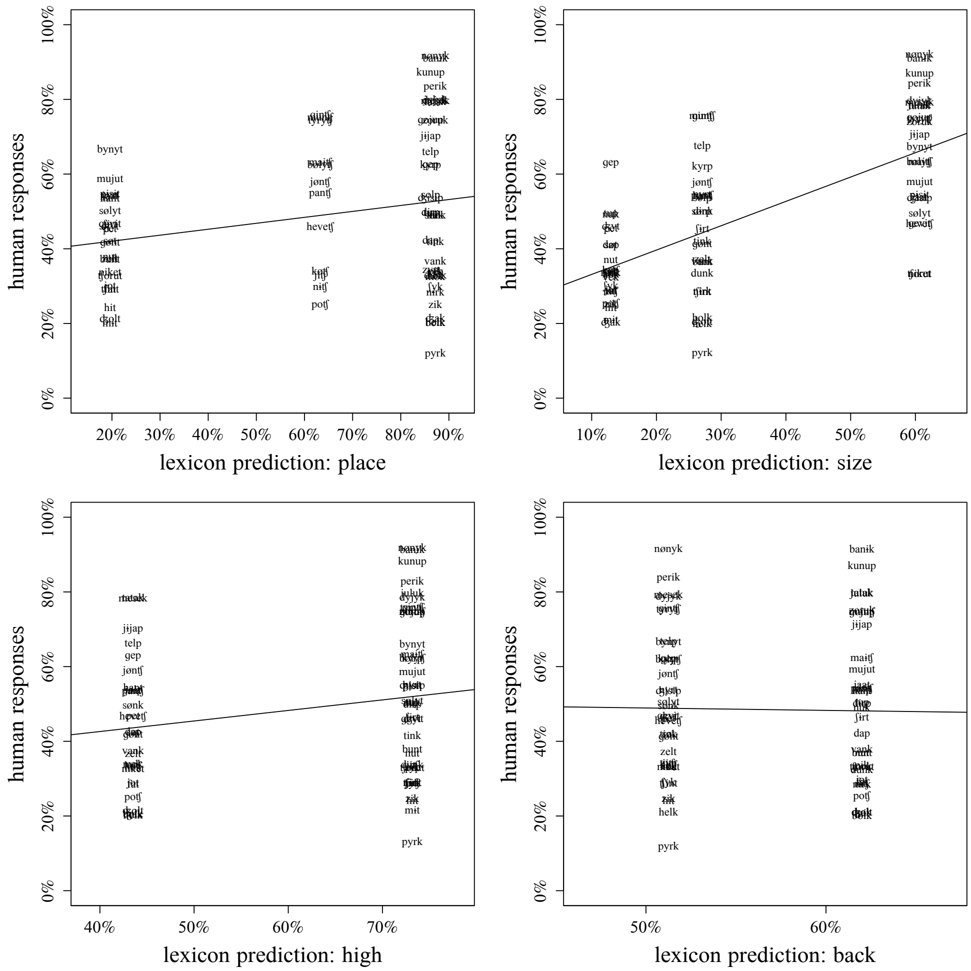
In other words, *size* and *place* had statistically significant power in predicting the participants' choice of alternation vs. non-alternation of stem-final stops. Crucially, however, none of the vowel features had an effect on the participants' choices, not even as a trend.

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.

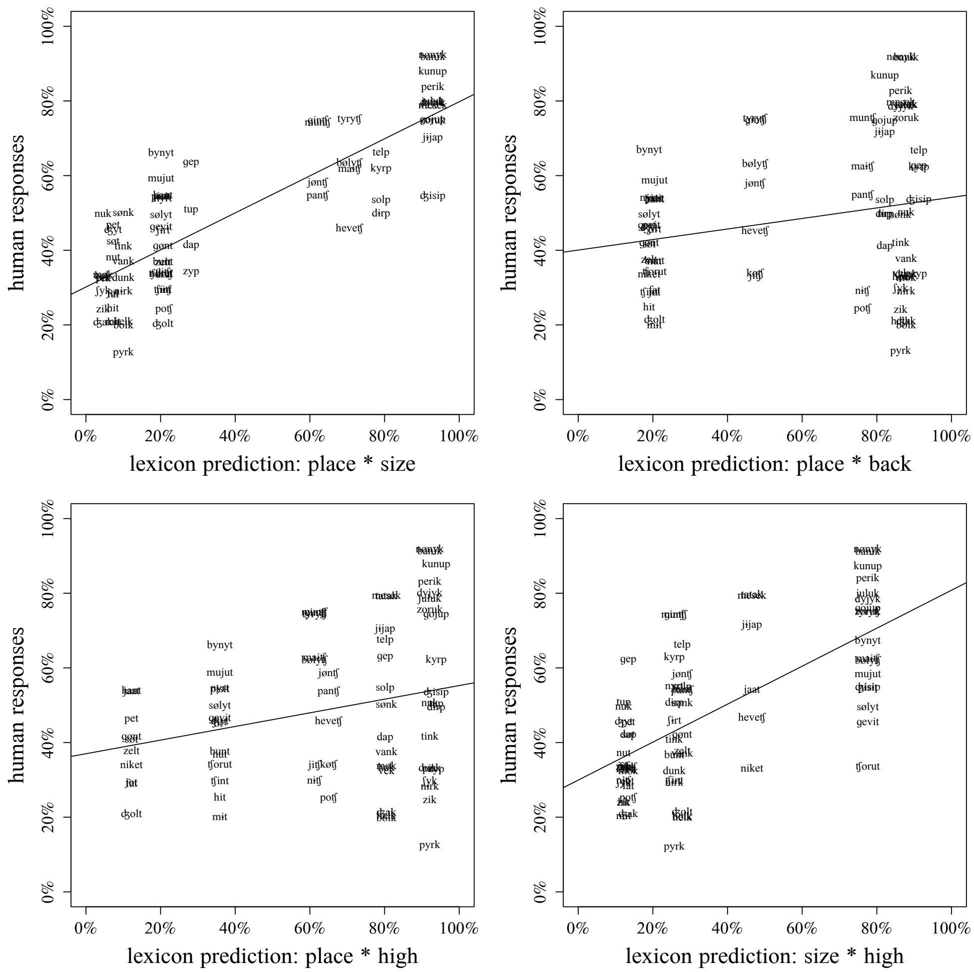
The effect of individual factors is shown in (14), where the panels parallel those in (3) and (10). Each panel was made by fitting an ordinal logistic regression model to TELL, as described in §2, and using the *predict()* function to generate the expected alternation rate. In such a simple model, this is just the average for the relevant items, e.g. 20% for *t*-final nouns and 63% for *t͡ʃ*-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 non-high vowels.

(14) Lexicon vs. nonce words, by single features



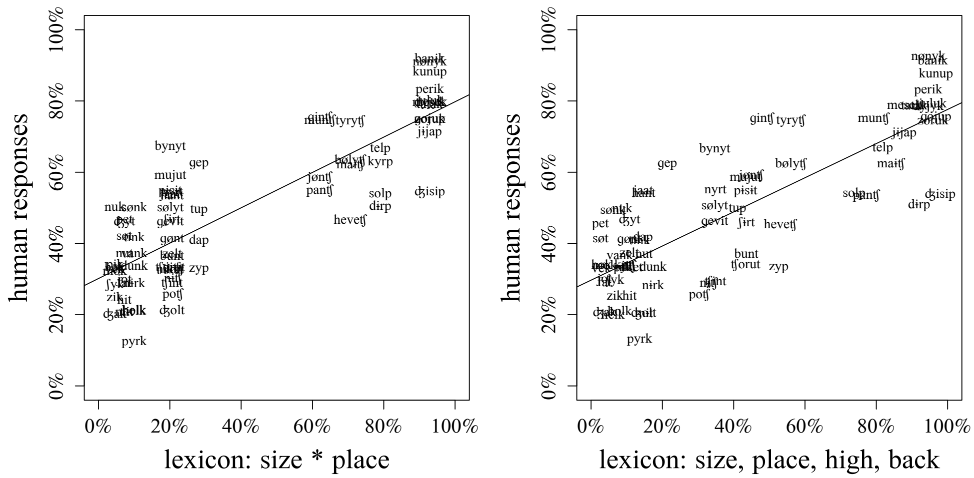
The panels in (15) show the predictions of lexicon models that use pairs of predictors with their interactions, paralleling those in (4) and (11). Here it becomes clearer that *place***size* is a powerful combination that offers an excellent prediction of the experimental results. The correlation is much more modest when *place* is augmented by *back* or *high*. Comparing the bottom two panels, we see that *high* has a rather a modest effect on the correlation, and the difference between the panels is attributed to the stronger effect of *size* relative to *place*.

(15) Lexicon vs. nonce words, by pairs of features



Finally, the two panels in (16) compare the a model of the lexicon that only uses *size***place* on the left, with a model of the lexicon that uses *size***place* + *high***place* + *back***place* on the right. This right model is identical to the one in (8). While the data points are more evenly distributed on the x-axis in the right panel, we see no improvement in correlation relative to the left panel.

(16) Lexicon vs. nonce words, without and with vowel features



Moving beyond the visualization in (16), we employed model comparison tests to show that the left panel offers a superior match to the experimental results. To test the effect of *high* separately from *back*, we define four models, shown in (17). 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 (8).

(17)

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

As mentioned above, each model was fitted to TELL using *lrm()*, the *predict()* function was used to generate a single predictor from each model, and the four predictors were then merged with the experimental results. The table in (18) shows on the left that in the lexicon itself, the predictive power of the models, as measured by the models' L.R. and D_{xy} is lowest for BASE and highest for FULL. On the right, the opposite is true when these models are applied to the experimental results: FULL is lowest, while BASE is highest. In other words, the model that *lacks* any information about vowel quality is the one that offers the better fit to the experimental results.

(18)

	fit to the lexicon		fit to the experimental results	
	Model L.R.	D_{xy}	Model L.R.	D_{xy}
BASE	1920	.744	196	.360
BACK	1948	.757	187	.349
HIGH	2049	.800	181	.345
FULL	2078	.810	175	.349

The BASE model is a significantly better predictor of the experimental results, as determined by model comparison tests. All the comparisons were made using nested models, e.g. the BASE predictor and the HIGH predictor were compared to a to superset model that had both of these predictors in it. The addition of BASE to an *lrm()* model that contains HIGH makes a significant improvement (χ^2(1)=13.78, p < .001), and the same is true for BACK (χ^2(1)=8.00, p < .005) and for FULL (χ^2(1)=19.79, p < .001). The opposite comparisons were all non-significant (p > .1). The same model comparisons were also done with *lmer()* models that were fitted to the experimental results, and the same results obtained: HIGH (χ^2(1)=11.86, p < .001), BACK (χ^2(1)=6.26, p < .05), and FULL (χ^2(1)=15.14, p < .001).

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 non-alternating 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 file was presented with an orthographic representation of it, 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 that's very similar to the one we used for the main study. Overall, the speakers found the materials to be of high quality, with 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 isn't masking the grammatical effects, we calculted 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 (13) made no noticeable change, as confirmed by an ANOVA model comparison ($\chi^2(1) = .260$, $p > .1$).

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 non-alternating status of single nouns, speakers 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 *t͡ʃ*-final CVC and CVCC nouns was successfully reproduced in the experiment results, even though the evidence comes from 23 and 18 nouns, respectively, and it is a non-significant 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 (19), 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 (20a-b), or they alternate (20c). It is also known that final voiced stops, as in (20b), are rare in the language.

(19) Two-way surface distinction in roots

	initially		inter-vocalically	
a.	t^hin	'soul'	at^ha	'ancestor'
b.	din	'religion'	ada	'island'

(20) Three-way contrast finally

	bare stem	possessive	
a.	at^h	at^h-ɨ	'horse'
b.	ad	ad-ɨ	'name'
c.	$\text{t}^\text{h}\text{at}^\text{h}$	t^had-ɨ	'taste'

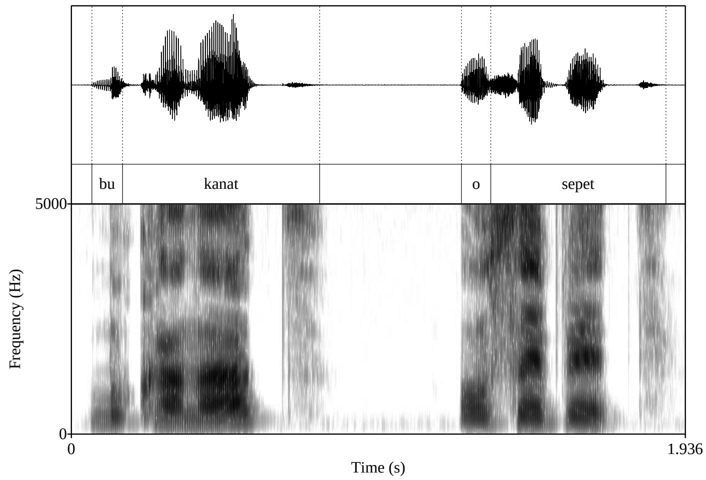
In Turkish orthography, the surface distinction is represented by the letters ⟨p, t, ç, k⟩ and ⟨b, d, c, g⟩, and the distinction was taken to be one of voicing by much of the literature on Turkish (Lees 1961; Inkelas & Orgun 1995; Inkelas et al. 1997, and many others).

More recently, Kallestinova (2004) and Petrova et al. (2006) have shown that the voiceless stops of Turkish are in fact aspirated in onset position.[9] 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 (21) exemplifies the finding in Kopkallı (1993), showing a clear, voiceless burst at the end of both the alternating k^hanat^h and the non-alternating sep^het^h. In fact, this token, spoken by a 30 year old male speaker from Istanbul, happens to have an even stronger burst for k^hanat^h, although Kopkallı (1993) shows that there is no significant difference in the duration of the final burst between alternating and non-alternating nouns.

[9] 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.

(21) [bu k^hanat^h o sep^{h}et^h] "This is a wing; that is a basket" (lit. this wing; that basket)



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 will show, the inconsistent ranking arguments allow the speaker to build lexical information into their grammar, and thus learn the distribution of the laryngeal alternations in grammatical terms. In this paper, 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 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).

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 (22). 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 larnygeal 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.

(22) a. The UR's 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-ɨ] **requires** IDENT(lar) ≫ *VtV

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

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

t^haD + I	IDENT(lar)	*VtV
a. t^hat^h-ɨ		*!
b. ☞ t^had-ɨ		

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 this analysis is that both rankings in (22) are consistent. Alternating nouns like t^hat^h and non-alternating 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).

We propose that the status of a word as alternating or non-alternating must be represented by lexically-specific grammatical rankings, instead of (or in addition to) an underlying difference. In essence, our argument is that only by including the alternating or non-alternating status of a word as a *grammatical* rather than lexically memorized phenomenon can one make sense of the grammatical biases against extending all lexical statistics. Underlying representations may very well be computed, but they are too powerful as a tool for predicting which lexical trends will be generalized.

The analysis we offer in §4.3, summarized in (23) below, posits the bare forms of nouns as their underlying representations. As (23c) 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.

(23) 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-ɨ] **requires** IDENT(LAR) ≫ *VtV
/t^hat^h + I/ → [t^had-ɨ] **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, 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, however, goes beyond the scope of this paper.

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). However, they do not perform an unconstrained comparison to their lexicon, 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 will be outlined here, with further details available in [reference redacted].

As explained in §4.2, we assume the bare stems of Turkish nouns as their underlying forms. Nouns with non-alternating 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. non-alternating 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 (24-25) exemplify the conflicting grammars required by the non-alternator *anat͡ʃʰ* 'cub' and the alternator *amat͡ʃʰ* 'goal'.

(24)

/ anat͡ʃʰ + ɨ /	IDENT(lar)	*Vt͡ʃV
a. ☞ anat͡ʃʰ-ɨ		*
b. anad͡ʒ-ɨ	*!	

(25)

/ amat͡ʃʰ + ɨ /	*Vt͡ʃV	IDENT(lar)
a. ☞ amad͡ʒ-ɨ		*
b. amat͡ʃʰ-ɨ	*!	

When the learner encounters a conflict as in (24-25), they 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 some lexical items. If IDENT(lar) is chosen, the resulting grammar is the one in (26).

(26) $\text{IDENT(lar)}_{\text{anat͡ʃ}} \gg$ *Vt͡ʃV $\gg \text{IDENT(lar)}_{\text{amat͡ʃ}}$

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, *Vt͡ʃV, *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.[10]

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 *ta͡tʃ* ∼ *ta͡dʒɨ* 'crown', the alternation impacts the initial syllable of the base, and thus *V͡tʃV must outrank both IDENT(lar)$_{\sigma 1}$ and the general IDENT(lar). In a polysyllabic alternator like *ama͡tʃ* ∼ *ama͡dʒɨ* 'goal', the initial syllable is not disturbed, and *V͡tʃV 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 MAX$_{\sigma 1}$ in addition to MAX allows the deletion of stem-final *k* to be learned separately for monosyllable and polysyllables.

Adding *a͡tʃ* ∼ *a͡tʃɨ* 'hunger' and *ta͡tʃ* ∼ *ta͡dʒɨ* 'crown' to the grammar-fragment in (26) yields the grammar-fragment in (27), which separates monosyllables from polysyllables.

(27) IDENT(lar)$_{\sigma 1_{a\widehat{t\int}}}$, IDENT(lar)$_{ana\widehat{t\int}}$ ≫ *V͡tʃV ≫ IDENT(lar)$_{\sigma 1_{ta\widehat{t\int}}}$, IDENT(lar)$_{ama\widehat{t\int}}$

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, *V͡tʃV, *VkV) and faithfulness (IDENT(lar), IDENT(lar)$_{\sigma 1}$, MAX, MAX$_{\sigma 1}$), 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.

Now suppose a speaker encounters a novel noun in its bare form, such as *heve͡tʃʰ*, and they are 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 *V͡tʃV, and one that has the reverse ranking. The clone of IDENT(lar) that ranks below *V͡tʃV has more items associated with it, and it therefore has a stronger effect on novel words, making the possessive of *heve͡tʃʰ* more likely to be *heve͡dʒ-i* than *heve͡tʃʰ-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. Extending this approach to phonotactics, and especially to exceptional phonotactics, is left for future work, yet an interesting attempt in this direction can be found in Coetzee (2008). Another interesting approach to 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

[10]The full analysis of dorsal deletion also involves a constraint against post-vocalic [ɡ] and appropriate ranking of IDENT(lar). These details are covered in [reference redacted].

inventory of extant and plausible UG constraints. Since UG doesn't contain constraints that relate vowel height and vowel backness to laryngeal features of a neighboring consonants, such relations cannot be encoded using Universal constraints. The technical details of the analysis are worked out in [reference redacted]; 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 paper 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, fall 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.

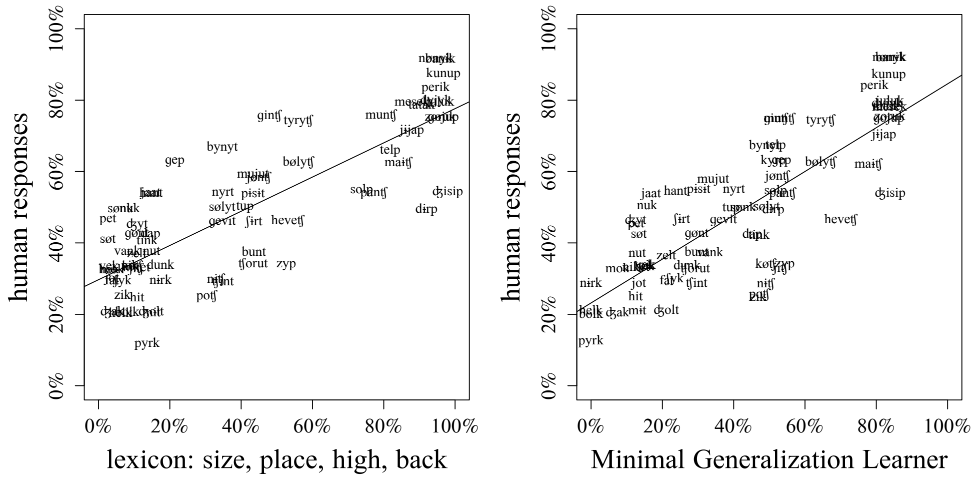
5.1. General-purpose learning with the MGL

In this subsection we compare a biased 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. 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.

We supplied the MGL with the lexical items in TELL, and assessed its predictions for the experimental items. The results are plotted in (28), where the left panel is repeated from (16), and the right panel shows the MGL predictions. We see similarly impressive fits in both.

(28) Predicting alternations in nonce words, regression model vs. MGL



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.[11]

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.

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 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 effect could be hard-wired by not supplying the model with information about vowel quality (although this would be a pointless exercise, 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.

[11]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 ($p < .05$).

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 or not, 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. Cross-linguistically, 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. Napikoğlu & Ketrez (2006) find that children quickly master suffixal allomorphy for the aorist, which is based on syllable-count. Ketrez (2007) finds that children's metathesis errors involving labials (e.g. $k^h it^h ap^h \rightarrow k^h ip^h at^h$ 'book') do not occur with monosyllables (e.g. yap^h) and attributes this to protection of 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 & 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 pre-existing 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 *t͡ʃi* 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 [t͡ʃ] 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 the consonant (i.e. a high tone on a vowel implies voiceless consonants, or vice versa), as found in Shanghainese or Jabem (Poser 1981). However, this vowel-consonant phonotactic interaction involves a common phonetic feature in both the trigger and target as well: 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. However, the putative interaction of vowel height with consonantal laryngeal features does not even remotely fit within this rubric, 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 hypotheses 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 with Moreton (2010), and before that in Pycha et al. (2003), Peperkamp et al. (2006), and others. Whether or not the same-feature constraint on vowel-consonant phonotactics is innate, or perhaps itself induced in parallel, e.g. 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 [d͡ʒ] than [ʒ], are ultimately traceable to lexical borrowing from French centuries ago, when Old French had [d͡ʒ] 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 *burd͡ʒ* 'sign' > Turkish *burt͡ʃʰ* ∼ *burd͡ʒ-u*), whereas source words that end in a voiceless stop are non-alternating across the paradigm. Arabic lacks the consonants [p] and [t͡ʃ] and has many nouns that end in [b] and [d͡ʒ], and as a consequence, the lexicon's overall alternation rates are boosted for those places of articulation. On the other hand, the existence of many Arabic nouns with feminine suffix *-at/-et* boosted the number of non-alternating, non-high 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 in Hayes (1999), who claims that when history creates non-Universal 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* > *guru**b**-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 non-adjacent 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 which 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 et al., 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 2003a, 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." (p.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 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 *aMɨ* in the lexicon. This is clearly a complex and somewhat unnatural constraint, 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 & 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 & 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 non-high. 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 & Pater (2010) explore Dutch speakers' phonotactic knowledge that long vowels are rarely followed by non-coronal consonants (e.g. mɛlk 'milk' vs. *meːlk). While the generalization is strong in the Dutch lexicon both in monosyllables and in polysyllables, Kager & 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 non-local interaction between vowel length and the non-coronality of a non-adjacent following consonant. A different notion of complexity arises from Zhang et al. (2006), who found very little extension of Taiwanese tone circle sandhi to wugs or to novel combinations of existing lexical items. Here, 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 (as opposed to other productive tone sandhi patterns, such as 24 $\rightarrow$ 33 in the language).

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 non-native 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 in which may not be recognized as part of a grammatical system. Results such as 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 — e.g., 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 fo the stimulus, and deploy linguistically unnatural mechanisms to help?

Casting the failure of incorporating phonotactic knowledge as we have above may also invite re-visiting 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 towards 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 & Lidz (2009; p.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 D. T. 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 all possibilities when going from lexicon to grammar.

REFERENCES

Albright, Adam (2008a). A Restricted Model of UR Discovery: Evidence from Lakhota. Ms. MIT.

Albright, Adam (2008b). Explaining universal tendencies and language particulars in analogical change. In Jeff Good (ed.) *Language Universals and Language Change*, Oxford University Press. 36 pp.

Albright, Adam, Argelia Andrade & Bruce Hayes (2001). Segmental environments of Spanish diphthongization. In Adam Albright & Taehong Cho (eds.) *UCLA Working Papers in Linguistics 7 (Papers in Phonology 5)*, UCLA. 117–151.

Albright, Adam & Bruce Hayes (2002). Modeling English past tense intuitions with minimal generalization. In Michael Maxwell (ed.) *Proceedings of the sixth meeting of the ACL special interest group in computational phonology*. Philadelphia: ACL, 58–69.

Albright, Adam & Bruce Hayes (2003). Rules vs. Analogy in English past tenses: a computationalexperimental study. *Cognition* **90**. 119–161.

Albright, Adam & Bruce Hayes (2006). Modeling productivity with the gradual learning algorithm: The problem of accidentally exceptionless generalizations. In Gisbert Fanselow, Caroline Féry,

Matthias Schlesewsky & Ralf Vogel (eds.) *Gradience in Grammar*, Oxford University Press. 185–204.
Anttila, Arto (2002). Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory* **20**. 1–42.
Archangeli, Diana & Douglas Pulleyblank (1994). *Grounded Phonology*. MIT Press.
Avery, Peter (1996). *The Representation of Voicing Contrasts*. Ph.D. dissertation, University of Toronto.
Baayen, R. Harald (2008). *Analyzing Linguistic Data: A practical introduction to statistics*. Cambridge University Press.
Bailey, Todd & Ulrike Hahn (2001). Determinants of wordlikeness: phonotactics or lexical neighborhoods. *Journal of Memory and Language* **44**. 568–591.
Barnes, Jonathan (2001). Domain-initial strengthening and the phonetics and phonology of positional neutralization. Paper presented at the Northeast Linguistics Society Meeting, CUNY.
Becker, Michael & Andrew Nevins (2009). Initial-syllable faithfulness as the best model of word-size effects in alternations. Talk given at NELS 40.
Beckman, Jill (1997). Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* **14**. 1–46.
Beckman, Jill (1998). *Positional Faithfulness*. Ph.D. dissertation, University of Massachusetts Amherst, Amherst, MA.
Beckman, Jill N. & Catherine O. Ringen (2004). Contrast and redundancy in OT. In Vineeta Chand, Ann Kelleher, Angelo J. Rodríguez & Benjamin Schmeiser (eds.) *Proceedings of WCCFL 23*, Cascadilla Press. 85–98.
Berko, Jean (1958). The child's learning of English morphology. *Word* **14**. 150–177.
Bhat, D.N.S. (1978). A general study of palatalization. In Joseph H. Greenberg (ed.) *Universals of human language, vol. 2: phonology,*, Stanford, CA: Stanford University Press. 47–92.
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. In Robert Kirchner, Wolf Wikeley & Joe Pater (eds.) *Papers in Experimental and Theoretical Linguistics*, University of Alberta, vol. 6. 24–35.
Boersma, Paul & David Weenink (2008). Praat: Doing phonetics by computer (version 5.0.24). [Computer program].
Bonatti, Luca, Marcela Peña, Marina Nespor & Jacques Mehler (2005). Linguistic constraints on Statistical Computations. *Psychological Science* **16.6**. 451–459.
Breiman, Leo, Jerome Friedman, R. A. Olshen & Charles J. Stone (1984). *Classification and regression trees*. Belmont, CA: Wadsworth International Group.
Buder, E. & C. Stoel-Gammon (2002). American and Swedish children's acquisition of vowel duration: Effects of vowel identity and final stop voicing. *Journal of the Acoustic Society of America* **111**. 1854–1864.
Burzio, Luigi (2002). Surface-to-Surface Morphology: when your representations turn into constraints. In P. Boucher (ed.) *Many Morphologies*, Cascadilla Press. 142–177.
Bybee, Joan (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* **10**. 425–455.
Casali, Roderic (1998). *Resolving Hiatus*. Garland, New York.
Chambers, J.K. (1973). Canadian Raising. *Canadian Journal of Linguistics* **18**. 113–135.
Chater, Nick & Christopher D. Manning (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences* **10.7**. 335–344.
Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* **22**. 125–159.
Cho, Young-mee Yu (1999). *Parameters of Consonantal Assimilation*. Munich: Lincom Europa.

Coetzee, Andries W. (2008). Grammaticality and ungrammaticality in phonology. *Language* **84**. 218–257.

Cohn, Abigail (1993). *Phonetic and Phonological Rules of Nasalization*. Ph.D. dissertation, UCLA.

Coleman, John & Janet Pierrehumbert (1997). Stochastic phonological grammars and acceptability. In *Computational Phonology. Third Meeting of the ACL Special Interest Group in Computational Phonology*, Somerset, NJ: Association for Computational Linguistics. 49–56.

Daelemans, Walter, Jakub Zavrel, Ko Van Der Sloot & 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**. 104–137.

Denes, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustic Society of America* **27**. 761–764.

Dresher, B. Elan & Jonathan Kaye (1990). A computational learning model for metrical phonology. *Cognition* **34.2**. 137–195.

Ernestus, Miriam & R. Harald Baayen (2003). Predicting the Unpredictable: Interpreting Neutralized Segments in Dutch. *Language* **79**. 5–38.

Flack, Kathryn (2007). *The sources of phonological markedness*. Ph.D. dissertation, University of Massachusetts Amherst.

Gildea, Daniel & Daniel Jurafsky (1996). Learning Bias and Phonological-Rule Induction. *Computational Linguistics* **22.4**. 497–530.

Good, I.J. (1980). Some history of the hierarchical Bayesian methodology. In J.M. Bernardo, M.H. DeGroot, D.V. Lindley & A.F.M. Smith (eds.) *Bayesian statistics,*, Valencia University Press. 489–519.

Goodman, Nelson (1955). *Fact, Fiction, and Forecast*. Harvard University Press.

Gussmann, Edmund (1980). *Studies in Abstract Phonology*. Cambridge, MA: MIT Press.

Hall, Tracy & Silke Hamann (2006). Towards a typology of stop assibilation. *Linguistics* **44.6**. 1195–1236.

Halle, Morris & Kenneth N. Stevens (1971). A note on laryngeal features. *MIT Quarterly Progress Report* **11**. 198–213.

Hay, Jennifer & R. Harald Baayen (2005). Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences* **9**. 342–348.

Hay, Jennifer, Janet Pierrehumbert & Mary Beckman (2004). Speech perception, well-formedness and the statistics of the lexicon. In J. Local, R. Ogden & R. Temple (eds.) *Phonetic Interpretation: Papers in Laboratory Phonology VI*, Cambridge University Press. 58–74.

Hayes, Bruce (1995). On what to teach the undergraduates: Some changing orthodoxies in phonological theory. In Ik-Hwan Lee (ed.) *Linguistics in the Morning Calm 3*, Seoul: Hanshin. 59–77.

Hayes, Bruce (1999). Phonological Restructuring in Yidiɲ and its Theoretical Consequences. In Ben Hermans & Marc van Oostendorp (eds.) *The derivational residue in phonology*, Amsterdam: Benjamins. 175–205.

Hayes, Bruce (2004). Phonological acquisition in optimality theory: the early stages. In René Kager, Joe Pater & Wim Zonneveld (eds.) *Constraints in Phonological Acquisition*, Cambridge: Cambridge University Press. 158–203.

Hayes, Bruce & Zsuzsa Londe (2006). Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* **23**. 59–104.

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

Hayes, Bruce, Kie Zuraw, Péter Siptár & Zsuzsa Londe (2009). Natural and unnatural constraints in hungarian vowel harmony. *Language* **85**. 822–863.

Hume, Elizabeth (1994). *Front Vowels, Coronal Consonants and their Interaction in Nonlinear Phonology*. New York: Garland.

Inkelas, Sharon, Aylin Kuntay, John Lowe, Orhan Orgun & Ronald Sprouse (2000). Turkish Electronic Living Lexicon (TELL). Website, http://socrates.berkeley.edu:7037/.

Inkelas, Sharon & Cemil Orhan Orgun (1995). Level ordering and economy in the lexical phonology of Turkish. *Language* **71**. 763–793.

Inkelas, Sharon, Cemil Orhan Orgun & Cheryl Zoll (1997). The implications of lexical exceptions for the nature of the grammar. In Iggy Roca (ed.) *Derivations and Constraints in Phonology*, Oxford: Clarendon. 393–418.

Itô, Junko & Armin Mester (1995). The core-periphery structure in the lexicon and constraints on reranking. In Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, GLSA. 181–210.

Iverson, Gregory & Joseph Salmons (2007). Domains and directionality in the evolution of german final fortition. *Phonology* **24**. 121–145.

Jannedy, Stefanie (1995). Ohio State University Working Papers in Linguistics. *Journal of Linguistics* **45**. 56–84.

Jarosz, Gaja (2006). *Rich Lexicons and Restrictive Grammars - Maximum Likelihood Learning in Optimality Theory*. Ph.D. dissertation, Johns Hopkins University.

Kager, René (1999). *Optimality Theory*. Cambridge University Press.

Kager, René & Joe Pater (2010). Phonotactics as Phonology: Knowledge of a Complex Constraint in Dutch. Ms., University of Utrecht and UMass Amherst.

Kallestinova, Elena (2004). Voice and aspiration of stops in Turkish. In Grzegorz Dogil (ed.) *Folia Linguistica 38: Special Issue on Voice*, Berlin: Mouton de Gruyter. 117–143.

Keating, Patricia (1985). Universal phonetics and the organization of grammars. In V. A. Fromkin (ed.) *Phonetic Linguistics*, Academic Press. 115–132.

Kemp, Charles, Amy Perfors & Joshua Tenenbaum (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science* **10.3**. 307–321.

Ketrez, Nihan (2007). Alignment versus linearity constraints in a Turkish child's speech. *Dilbilim Araştırmaları* .

Kingston, John (2002). Keeping and losing contrasts. In *Proceedings of the 28th Annual Meeting*, Berkeley Linguistics Society. 155–176.

Kirchner, Robert (1998). *An Effort-Based Approach to Consonant Lenition*. Ph.D. dissertation, UCLA.

Kopkallı, Handan (1993). *A phonetic and phonological analysis of final devoicing in Turkish*. Ph.D. dissertation, University of Michigan.

Lees, Robert (1961). *The Phonology of Modern Standard Turkish*. Bloomington: Indiana University Press.

Lewis, Geoffrey L. (1967). *Turkish Grammar*. Oxford: Clarendon.

Lisker, Leigh & Arthur Abramson (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word* **20**. 384–422.

Luce, Paul A. & D.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*. Ph.D. dissertation, Rutgers University.

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 & Erik Thomas (2007). Origins of Canadian Raising in voiceless-coda effects: a case study in phonologization. In Jennifer Cole & José I. Hualde (eds.) *Papers in Laboratory Phonology 9*, Mouton de Gruyter. 37–64.

Nakipoğlu, Mine & Aslı Üntak (2006). What does the acquistion of stems that undergo phonological alternation reveal about rule application. Paper presented at the International Conference on Turkish Linguistics, Uppsala, Sweden.

Napikoğlu, Mine & Nihan Ketrez (2006). Children's overregularizations and irregularizations of the Turkish aorist. In *BUCLD 30: Proceedings of the Boston University Conference on Language Development, Volume 2*, Somerville, MA: Cascadilla Press. 399–410.

Nicolae, Andreea & Andrew Nevins (2010). Underlying Laryngeal Specifications, Fricative Alternations, and Word–Size Effects. Ms., Harvard University.

Ohala, John (1983). The origin of sound patterns in vocal tract constraints. In *The production of speech*, New York: Springer-Verlag. 189–216.

Onishi, Kristine, Kyle Chambers & 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. In Leah Bateman & Adam Werle (eds.) *UMOP: Papers in Optimality Theory III*, Amherst, MA: GLSA. 1–36.

Pater, Joe (2009). Morpheme-specific phonology: Constraint indexation and inconsistency resolution. In Steve Parker (ed.) *Phonological Argumentation: Essays on Evidence and Motivation*, Equinox. 1–33.

Pearl, Lisa & Jeffrey Lidz (2009). When domain-general learning fails and when it succeeds: Identifying the contribution of domain specificity. *Language Learning and Development* **5**. 235–265.

Peperkamp, Sharon, Katrin Skoruppa & Emmanuel Dupoux (2006). The role of phonetic naturalness in phonological rule acquisition. In D. Bamman, T. Magnitskaia & C Zaller (eds.) *Proceedings of the 30 Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 464–475.

Peterson, G. E. & I. Lehiste (1960). Duration of syllable nuclei in english. *Journal of the Acoustic Society of America* **32**. 693–703.

Petrova, Olga, Rosemary Plapp, Catherine Ringen & 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**. 483–488.

Prince, Alan (2002). Arguing optimality. ROA 562-1102.

Prince, Alan & Paul Smolensky (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell. [ROA-537].

Pycha, Anne, Sharon Inkelas & Ronald Sprouse (2007). Morphophonemics and the Lexicon: A Case Study from Turkish. In M. J. Solé, P. Beddor & M. Ohala (eds.) *Experimental Approaches to Phonology*, Oxford University Press. 369–385.

Pycha, Anne, Pawel Nowak, Eurie Shin & Ryan Shosted (2003). Phonological Rule-Learning and Its Implications for a Theory of Vowel Harmony. In Gina Garding & Mimu Tsujimura (eds.) *WCCFL 22*, Cascadilla Press. 423–435.

Quinlan, J. R. (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, Austria. (http://www.R-project.org).

Saffran, Jenny R. (2003). Statistical language learning: Mechanisms and constraints. *Current*

Directions in Psychological Science **12.4**. 110–114.
Saffran, Jenny R. & Erik D. Thiessen (2003). Pattern induction by infant language learners. *Developmental Psychology* **39.3**. 484–494.
Sezer, Engin (1981). The k/∅ alternation in Turkish. In *Harvard Studies in Phonology*, Bloomington: Indiana University Linguistics Club. 354–382.
Sobel, David, Joshua Tenenbaum & Alison Gopnik (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science* **28**. 303–333.
Stevens, Alan M. (1968). *Madurese Phonology and Morphology. American Oriental Series, #52*. New Haven: American Oriental Society.
Tang, Katrina (2008). *The Phonology and Phonetics of Consonant-Tone Interaction*. Ph.D. dissertation, UCLA.
Tesar, Bruce (1998). Using the mutual inconsistency of structural descriptions to overcome ambiguity in language learning. In P. Tamanji & K. Kusumoto (eds.) *Proceedings of NELS 28*. Amherst, MA: GLSA, 469–483.
Tesar, Bruce (2006). Learning from paradigmatic information. In *Proceedings of NELS 36*.
Tesar, Bruce & Alan Prince (2006). Using phonotactics to learn phonological alternations. In *CLS 39*. Available as ROA-620.
Tesar, Bruce & Paul Smolensky (1998). Learnability in Optimality Theory. *Linguistic Inquiry* **29**. 229–268.
Tesar, Bruce & Paul Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.
Ussishkin, Adam & Andrew Wedel (to appear). Lexical access, effective contrast and patterns in the lexicon. In Paul Boersma & Silke Hamann (eds.) *Perception in Phonology*, Mouton de Gruyter.
Vaux, Bert & Bridget Samuels (2005). Laryngeal markedness and aspiration. *Phonology* **22**. 395–436.
Volatis, Lydia & Joanne Miller (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *J. Acoustic Soc. Am.* **92.2**. 723–735.
Warker, Jill & Gary Dell (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory and Cognition* **32.2**. 387–398.
Wedel, Andrew (2002). Phonological alternation, lexical neighborhood density and markedness in processing. Handout from presentation at LabPhon 8, Yale University.
Wetzels, Leo (1997). The Lexical Representation of Nasality in Brazilian Portuguese. *Probus* **9.2**. 203–232.
Wilson, Stephen (2003). A phonetic study of voiced, voiceless and alternating stops in Turkish. *CRL Newsletter* **15**. 3–13. URL http://stephenw.bol.ucla.edu/papers/turkishphon.pdf.
Zhang, Jie, Yuwen Lai & Craig Turnbull-Sailor (2006). Wug-testing the "tone circle" in taiwanese. In David Montero Donald Baumer & Michael Scanlon (eds.) *Proceedings of the 25th West Coast Conference on Formal Linguistics*. 453–461.
Zimmer, Karl (1969). Psychological correlates of some Turkish Morpheme Structure Conditions. *Language* **45.2**. 309–321.
Zimmer, Karl & 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*. Ph.D. dissertation, UC Berkeley.
Zuraw, Kie (2000). *Patterned Exceptions in Phonology*. Ph.D. dissertation, UCLA.