

# Asymmetries in Generalizing Alternations to and from Initial Syllables\*

Michael Becker  
becker@fas.harvard.edu

Andrew Nevins  
a.nevins@ucl.ac.uk

Jonathan Levine  
jplevine@fas.harvard.edu

Version of April 17, 2011. Prose subject to change, especially with your feedback!

## Abstract

We show that in the English lexicon, voicing alternations in the plural impact monosyllables more than finally-stressed polysyllables. This is the opposite of what happens typologically, and would thereby run contrary to the predictions of *initial syllable faithfulness*. However, in a wug-test we found that monosyllables are impacted just as much as finally-stressed polysyllables—a “surfeit of the stimulus” effect, in which speakers fail to learn a statistical generalization present in the lexicon. We then present two artificial grammar experiments in which English speakers in fact demonstrate the universal bias for protecting monosyllables and initial syllables more generally. The conclusion, therefore, is that speakers can exhibit spontaneous learning that goes directly against the evidence offered by the ambient language, a result attributed to Universal Grammar, which provides a set of formal and substantive biases in phonological acquisition.

## 1 Introduction

It is now quite generally known that ‘underlearning’ exists for unnatural linguistic patterns, and that language learners do not demonstrate equal willingness to generalize every statistical robust pattern in their lexicon. For example, focusing on patterns created by morphophonological alternations, Becker et al. (2011) offer a set of results in Turkish in which speakers learned natural phonotactic patterns and did not generalize unnatural patterns, even when robustly present in their lexicon. At the same time, Hayes et al. (2009) found a result with

---

\*For their valuable comments and discussion, we thank Adam Albright, Lauren Eby Clemens, Peter Graff, Peter Jurgec, John Kingston, Joe Pater, John McCarthy, Anne Pycha, Anne-Michelle Tessier, Marijn van ’t Veer, Ruben van de Vijver, Arun Viswanath, Matt Wolf, and the audiences at NELS 40, the LSA 85<sup>th</sup> annual meeting, and OCP8. Thanks to Parker Crane, Jacob Hoover, and Paul VanKoughnett for their work on different parts of this project. All errors remain the other two authors’.

Hungarian allomorph selection in which unnatural patterns are underlearned, but not completely ignored. In the current paper, we focus on a particular unnatural pattern found within the lexicon of English, in which monosyllables are impacted by voicing alternations (e.g. *knife* ~ *knives*) more than polysyllables are. This pattern would seem to contravene the widespread crosslinguistic generalization that initial syllables are more protected from alternations than non-initial syllables. We show that the pattern in the English lexicon goes unlearned in a nonce word task, and then move on to show that in two distinct artificial language experiments, English speakers in fact prefer the opposite pattern, and asymmetrically protect initial syllables.

The larger contribution of this paper is to demonstrate that phonological learning operates differently when generalizing *from* initial syllables to other syllables than from when generalizing *to* initial syllables from other syllables. We demonstrate that generalization in the former direction is much more robust, even when the two patterns are given equal statistical representation in the learning sequence (and in fact, even when the balance is decidedly tipped in the opposite direction, as in the lexicon of English).

We start in §2 with a survey of initial syllable protection effects in a variety of languages, and establish the cross-linguistic preference for protecting initial syllables, and monosyllabic stems – a particular subclass of initial syllables – from alternations. Then, §3 offers results from a survey of the English lexicon, showing that English has a preponderance of impacted monosyllabic items. This anomaly of the English lexicon is contrasted in §4 with a nonce word task (“wug test”, Berko 1958), where we see that monosyllables are impacted just as much as finally-stressed polysyllables. In other words, English speakers ignore the anomalous distribution in the lexicon (“surfeit of the stimulus”, Becker et al. 2011), and adopt a grammar that impacts monosyllables and polysyllables equally.

In §5, we present results from an artificial language experiment (a “poverty of the stimulus” experiment) that aims to bring out the protection of monosyllables in English speakers. We show that English speakers choose to protect monosyllables, confirming that they are subject to the same bias that shapes the typology we surveyed. In §6, we go beyond monosyllables to show that English speakers protect the initial syllables in polysyllables as well. The results are interpreted in §7 in terms of a universal approach to faithfulness in Optimality Theory (Prince & Smolensky 1993/2004), attributing the protection of monosyllables to initial

syllable faithfulness (Trubetzkoy 1939; Steriade 1994; Beckman 1997, 1998; Casali 1998; Barnes 2006; Jesney 2009; Becker 2009; Becker et al. 2011). Concluding remarks are offered in §8.

## 2 Typology: Initial syllables are protected from alternations

In this section, we survey the known typology of size-dependent local morphophonological alternations. We show that when alternations are sensitive to phonological size, they are skewed towards protection of monosyllables. This means that there is an implicational relationship: Alterations in monosyllables imply alternations in polysyllables, but not vice versa. In terms of Berwick (1985); Manzini & Wexler (1987), a language that allows alternations only in polysyllables is a *subset language*, and serves as the learner’s default assumption; in Optimality Theory, subset relationships are expressed in terms of ranking biases, such as the preference for specific faithfulness over general faithfulness (Smith 2000; Hayes 2004; Tessier 2007). Exposure to alternations in monosyllables will force the learner to adopt a more permissive grammar that allows alternations in both monosyllables and polysyllables. Crucially, there is no grammar that allows alternations only in monosyllables.

We start with an example of what we mean by a local morphophonological alternation, using the Slovenian adjectival suffix [-ən]. As the examples in (1) show, this affix palatalizes the stem-final consonant. Our consultant, Peter Jurgec, reports that this alternation applies without exception to all nouns that end in *k*, as well as to several other consonants (see also Toporišič 2000). But the alternation is morphophonological in that it is specific to this affix, as seen from the lack of palatalization in the monomorphemic [ˈpəkəw] ‘hell’ or the diminutive [ˈtʃúk-ətʂ] ‘owl.DIM’.

- (1) A regular local morphophonological alternation in Slovenian

base noun	adjective	
ˈbók	ˈbóʃ-ən	‘hip’/‘lateral’
ˈznák	ˈznáʃ-ən	‘sign’/‘marked’
baˈrók	baˈróʃ-ən	‘Baroque.N/ADJ’
oˈtrók	oˈtróʃ-ən	‘child’/‘childish’

Tamil seems to offer a case that is like Slovenian, except that monosyllables are protected from alternations, as illustrated in (2). Here, [n] becomes [ɲ] before the

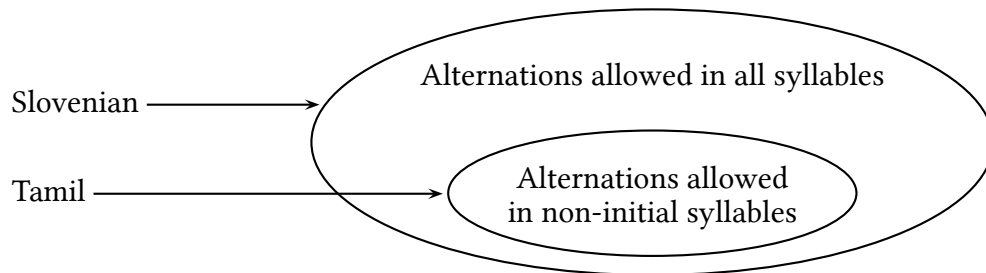
plural suffix [-gə], but [n] is protected from change in the initial syllable (Christdas 1988; Beckman 1997, 1998).

(2) Monosyllables protected in Tamil

singular	plural	
mi:n	mi:n-gə	'fish'
ma:n	ma:n-gə	'deer'
makən	makəŋ-gə	'son'
paj:ən	paj:əŋ-gə	'boy'

What's missing from the picture is the mirror-image of Tamil, i.e. a local morphophonological alternation that impacts stem-final segments in monosyllables but not in polysyllables. We show the range of known languages with the Venn diagram in (3).

(3) Alternations in initial syllables imply alternations in non-initial syllables:



A wider range of examples becomes available if we look at alternations that also have lexical exceptions. The best-known example is that of the stem-final laryngeal alternations in Turkish, where the process impacts the majority of polysyllables but only a minority of the monosyllables (Lees 1961; Inkelas & Orgun 1995; Inkelas et al. 1997; Hayes 1995; Kallestinova 2004; Petrova et al. 2006; Pycha et al. 2007; Becker 2009; Becker et al. 2011). As a representative example, the final stop of the polysyllabic *k<sup>h</sup>a'nat<sup>h</sup>* 'wing' comes out voiced in the possessive *k<sup>h</sup>ana'd-i*, but the stop is unchanged in the monosyllabic *k<sup>h</sup>at<sup>h</sup>* 'floor', where we find *k<sup>h</sup>a't<sup>h</sup>-i*. As Becker et al. (2011) have shown, this is not just a fact about the Turkish lexicon, but rather a fact about the grammar of Turkish speakers, as evidenced by the speakers' projection of the pattern from the lexicon onto nonce words.

Two interesting examples of the same type come from French and Brazilian Portuguese, as shown by Becker et al. (in preparation). In Brazilian Portuguese,

the plural induces a change of word-final [w] to [j] only in some lexical items (Gomes & Manoel 2010), as exemplified in (4). Becker et al. (in preparation) report that the alternation impacts a minority of the monosyllables and a majority of the polysyllables, and that in the nonce word task they ran, speakers extended this tendency from the lexicon to their treatment of nonce forms, exactly like in Turkish.

(4) Monosyllables protected more than polysyllables in Portuguese

singular	plural	
'saw	'sajs	'salt'
'paw	'paws	'stick'
de'daw	de'dajs	'thimble'
ka'kaw	ka'kaws	'cocoa'

In both Turkish and Brazilian Portuguese, the alternation is extended from native nouns to loan words, and in both languages we see the alternation apply more readily to polysyllables. As representative examples, Turkish has the non-alternating monosyllable  $t^hy^p^h \sim t^hy^p^h-y$  'tube' vs. the alternating polysyllable  $gu'rup^h \sim guru'b-u$  'group'. In Portuguese borrowings from English we find the non-alternating monosyllable  $'gow \sim 'gow-s$  'goal' vs. the alternating polysyllable  $koke'tew \sim koke'tej-s$  'cocktail'. In other words, the size effect is seen in these languages not only as a trend in the lexicon and as an effect of an experimental setting, but also in the natural treatment of new lexical items.

In French, the plural induces a change of word-final [al] and [aj] to [o], as in  $mal \sim mo$  'evil SG/PL' and  $travaj \sim travo$  'work SG/PL'. Becker et al. (in preparation) report that the alternation impacts only about 30% of the monosyllables and about 40% of the polysyllables, but even this very small trend is enough for the native speakers. In a nonce word task, speakers rated the alternation in polysyllables significantly more acceptable than in monosyllables. In this case, the French speakers are only given rather modest indication about the distribution of the alternation, yet they behave exactly like Turkish and Brazilian Portuguese speakers in preferring to protect nonce monosyllables from the alternation.

We propose that in cases in which speakers are known to protect monosyllables, they attribute this protection to the activity of *initial syllable faithfulness* constraints (Steriade 1994; Beckman 1997, 1998; Casali 1998; Barnes 2006; Jesney 2009; Becker 2009; Becker et al. 2011). In a monosyllable, effecting a change in the word-final segment impacts the word-initial syllable, as in the Portuguese  $'saw \sim$

'sajs, while the initial syllable stays unchanged in a polysyllable such as *de'daw* ~ *de'dajs*.<sup>1</sup>

Certain authors have found an atomic notion of initial syllable faithfulness to be unappealing, and have suggested that size effects are derived from phonetic or psycholinguistic sources. Barnes (2006) suggests that the source is phonetic duration: Initial syllables are phonetically longer in Turkish, and longer segments enjoy greater faithfulness. While we are sympathetic to this line of inquiry, we note that lengthening of initial syllables isn't universal. As Barnes himself reports, initial syllables in English are shortened, not lengthened. Furthermore, phonetic compression of segments in longer words as the force behind alternations makes sense in Turkish, where the alternation turns voiceless consonants to voiced ones. Making a consonant voiced makes it shorter, and thus a better fit to the shorter duration of segments in a long word. It's far from clear that the same notion extends to Brazilian Portuguese, where the alternation is one of backness, which tends to keep duration constant. Changing a coronal nasal to a velar, as in Tamil, is another case that isn't so clearly related to duration.

Others, such as Ussishkin & Wedel (2009) and Stausland Johnsen (to appear), suggest that short stems are protected from alternations by psycholinguistic rather than phonetic factors. Specifically, they claim that short stems are protected because shorter stems are in denser lexical neighborhoods, on average, than longer stems. Yet Pycha et al. (2007) show that in Turkish, such lexicon-based measures have very weak effects, and Becker et al. (2011) report that these weak effects are orthogonal to the simple monosyllabicity criterion.

We suggest that *initial syllable faithfulness* has been abstracted away from its functional motivation and has become an atomic unit of phonology, or phonologized (Hyman 1976, see recent review in Hyman 2008). We leave open the possibility, however, that new functional factors may be identified in the future. We note that the motivation for *initial syllable faithfulness* goes beyond the protection of stem-final consonants from alternations; phonotactic considerations and patterns of allomorph selection are relevant as well. The analysis of Shona in Beckman (1997, 1998) shows increased faithfulness in vowel quality in initial

---

<sup>1</sup>We note a technical question with the application of this concept in Classical Optimality Theory (Prince & Smolensky 1993/2004; Moreton 2004). In Turkish, the alternating segment does not surface in the initial syllable of the word when a vowel-initial suffix is added, as in *k<sup>h</sup>at<sup>h</sup>* ~ *k<sup>h</sup>a.t<sup>h</sup>-i*. This question is successfully addressed in Jesney (2009), who notes that for a variety of reasons, the initial syllable should refer to the fully faithful candidate, not to the winner.

syllables, with evidence from phonotactic restrictions on roots and allomorphy selection. Similarly, in Turkish native words, non-high round vowels are only allowed in initial syllables. In Spanish, faithfulness to initial syllables protects high vowels from becoming glides (Cabr  & Prieto 2006).

We focused in this section on local alternations, where the suffix's impact is on the immediately adjacent segments of the stem. While we have limited information on long-distance size-dependent alternations, we suspect that the picture may be the opposite. Umlaut, for instance, is known to be attracted to stressed syllables in both German (Wiese 1996; Fanselow & F ry 2002; van de Vijver & Baer-Henney 2010) and Chamorro (Topping 1968; Chung 1983). In the German plural, umlaut impacts only initial syllables—the exact opposite of Turkish. In Chamorro, umlaut seems to impact initial syllables with all umlauting suffixes. We leave the question of umlaut in these languages, and the broader question of size-dependent non-local processes, to future work.

Other cases may be harder to categorize. In French liaison, a consonant appears at the end of monosyllabic stems more frequently than at the end of polysyllabic stems (Mallet 2008), yet it is far from clear what the correct analysis in terms of markedness and faithfulness might be (C t  2011). Finnish assibilation (Anttila 2003, 2006; Giavazzi 2010) is sensitive to length, but all of the affected stems are polysyllabic; we note that this kind of effect seems rather rare.

Coming back to the clearer cases we started with, we conclude that the known typology supports an asymmetrical view of size-based restrictions on local alternations: Monosyllables are protected from alternations, and alternations in monosyllables imply alternations in polysyllables. As for the ultimate source of this effect, we cast doubt on the viability of phonetic explanations, and while we leave open the possibility of an eventual psycholinguistic explanation, we adopt here *initial syllable faithfulness* as an atomic element of Universal Grammar that regulates these size effects. Ultimately, however, this paper is not about the source of size effects, it is about their asymmetrical application. Whatever force is at play, it protects monosyllables more than polysyllables and not the other way around.

Next, we turn our attention to English, which at first blush seems to provide a counterexample, as a language with alternations that impact monosyllables more strongly than polysyllables.

### 3 Experiment 1: Monosyllables impacted in the English lexicon

In English, final [f] and [θ] alternate with the voiced [v/ð] in some nouns (5), but not others (6):

- |     |                      |         |
|-----|----------------------|---------|
| (5) | naɪf ~ naɪvz, *naɪfs | ‘knife’ |
|     | pæθ ~ pæðz, *pæθs    | ‘path’  |
| (6) | stɪf ~ stɪfs, *stɪvz | ‘stiff’ |
|     | dεθ ~ dεθs, *dεðz    | ‘death’ |

The phenomenon goes back several centuries, and has been known for a while (Jespersen 1909; Berko 1958; Hayes 2009), though not widely discussed in the generative literature. And yet, the process carries several hallmarks of a (semi-)productive morphophonological process, as evidenced by its application to a growing set of lexical items, and its partial independence from spelling conventions.

Firstly, we note that the voicing alternations are being extended to items that did not used to alternate, especially finally-stressed polysyllables. College-age American speakers regularly offer the plurals *gira[v]es*, *photogra[v]es*, and *psychopa[ð]s*, whereas older speakers find these forms utterly ungrammatical. Note the orthographic ⟨ff⟩ in ‘giraffes’ and the ⟨ph⟩ in ‘photographs’, which indicate that these nouns didn’t alternate historically. Voicing alternations aren’t always reliably recorded in the spelling even for monosyllables. The plural of ‘roof’ is *roo[v]s* for most Americans, yet the spelling ⟨rooves⟩ is quite uncommon. Finally, we note that the alternations aren’t indicated at all for θ-final nouns, yet speakers have clear intuitions about such nouns, as in (5) and (6) above.

Secondly, we note that the set of alternating lexical items has both expanded and contracted over the last few centuries. As Jespersen notes, post-[ɪ] voicing is innovative, as in the plurals of [dwoɪf] ‘dwarf’, [woɪf] ‘wharf’, and [skaɪf] ‘scarf’. Many nouns that had concomitant voicing and vowel-length alternations, as in [stæf] ~ [stervz] ‘staff’, switched to leveled paradigms. Finally, some of our language consultants note massive loss of alternations in upstate New York and parts of the Midwest. All these changes suggest that grammatical forces are at work, shaping the distribution of the voicing alternations in each speaker’s grammar. While the alternations primarily affect Germanic words, we note the alternations in non-Germanic words such as ‘scarf’, ‘giraffe’, and ‘psychopath’.

The English voicing alternations are dependent on morphophonological con-

text. While they affect many nouns in the presence of the plural [z], the identically sounding genitive suffix never triggers alternations, as in *knife's*, *roof's*, *path's*, etc. Some of our consultants report the blocking of the plural alternations in compounds, as in the simplex plural [buðz] ‘booths’ vs. the compound [tol-buθs] ‘toll-booths’ (cf. Kiparsky 1982; Selkirk 1982; et seq.). Voicing alternations may be triggered in denominal verbs, without a clear connection to the voicing alternations: for some items, only the noun voices (*knives/to knife*, *leaves/to leaf*, *dwarves/to dwarf*); for others, only the verb voices (*beliefs/to believe*, *proofs/to prove*); and others still have the alternation in both (*halves/to halve*, *lives/to live*, *shelves/to shelve*). Adam Albright (p.c.) suggests that for some speakers, verbs may alternate upon the addition of the past ⟨-ed⟩, e.g. [naɪf] ~ [naɪvd]. Beyond the plural and the denominal verbs, voicing affects a smattering of other forms, as in *thief* ~ *thievery*. There may be a bias to use the voiced fricative in all derived forms for a given noun, perhaps a lexical conservatism effect (Steriade 1999).

To summarize, then, we see that the voicing alternations are limited in their morphological and lexical scope, and yet they are subject to strong fluctuations, and are only partly constrained by history or spelling. We take these fluctuations to reflect a grammatical organization of the alternations, an organization that is largely phonological. Specifically, we want to explore the role of prosodic shape (monosyllabicity and stress) in shaping these patterns. To get a realistic sense of the range of acceptable plurals, we surveyed a large number of American English speakers from a range of locations and ages, as explained in the rest of this section.

### 3.1 Materials

We extracted all of the f-final and θ-final words from the CMU dictionary, and chose a set of 126 nouns from them, aiming to represent a wide range of prosodic shapes and rhymes. The full list of items, with results by item, is in Appendix A. The list contains 82 monosyllables (52 f-final, 30 θ-final), 21 polysyllables with a stressed final syllable, or “iamb” (13 f-final, 8 θ-final), and 23 polysyllables with an unstressed final syllable, or “trochee” (12 f-final, 11 θ-final).

In addition, we selected 50 filler items, of which 17 were monosyllables, 15 “iamb” and 18 “trochee”. These included items whose plural is formed by umlaut (e.g. *goose*, *woman*), infrequent plural suffixes (e.g. *ox*, *appendix*), and items similar to those (e.g. *box*, *human*, *status*).

The items were recorded by a phonetically trained male native speaker of

English in his twenties from Maine. The list included each noun in the singular and two plurals, and it was presented to the speaker in random order three times. The recording was made in a sound-attenuated booth into a Macintosh computer. Then, the best token was chosen and converted to mp3 format using the LAME converter. The recordings were not manipulated in any other way, other than normalizing the intensity with Praat (Boersma & Weenink 2008), and sounded quite natural.

The speaker regularly lengthened vowels before voiced final fricatives, and pronounced many final fricatives voiceless, as is common in general American English. Thus, the voicing contrast was often mainly heard as a difference in allophonic vowel length.

### 3.2 Methods

The experiment was presented to the participants over the internet, using the web browser of the participant's choice. The web server started by making a random selection of items for each participant, choosing a total of 36 items: 20 target items and 16 fillers.

The item selection was done as follows: Six monosyllables (3 f-final and 3  $\theta$ -final), four trochees (2 f, 2  $\theta$ ), and a total of ten iambs: six monomorphemic iambs (4 f, 2  $\theta$ ), and four poly-morphemic iambs (2 f, 2  $\theta$ ). The fillers were 6 monosyllables, 5 trochees, and 5 iambs, in addition to the example item, *cactus*.

Before the experiment began, participants were reminded that some English plurals are exceptional, giving *ox*, *knife*, and *fish* as examples, and asking the participants for their preferred plural.

The items were presented orthographically on the screen, with buttons to play two plural forms in random order. Once both plurals were played, a seven-button scale appeared between them. After the participant pressed one of the seven buttons of the scale, an additional question appeared, asking about the *second* plural the participant heard. These questions were used to ensure that the participants were listening to the audio material and hearing the voicing distinctions. The questions asked about vowel quality or voicing, e.g. "did the second plural you heard have the 'f' sound of 'fan' or the 'v' sound of 'van'?"

After the participant responded to all 36 target items and fillers, they were asked to answer a few demographic questions (year of birth, sex, country of origin, kind of English spoken, other languages spoken, suggestions).

### 3.3 Participants

The participants were recruited online using Amazon’s Mechanical Turk, and were paid \$.30 for their time. Amazon’s Mechanical Turk is a web application that provides access to an arbitrarily large number of potential participants for survey-based experiments; see Sprouse (2010) about the use of the Mechanical Turk in linguistics.

We gathered completed surveys from 200 people. The server logs indicate that these participants took on average 13 minutes to complete the experiment (range 7–44 minutes, median 12). At the end of the experiment, participants volunteered the following demographic information. Year of birth was provided by 152 participants, who reported an average age of 34 (range 18–70, median 29). Sex: 127 females, 65 males, 8 didn’t say. Country of origin: 180 from the United States, 4 from another English-speaking country, 2 from another non-English speaking country, 14 didn’t say. As for the kind of English spoken, almost everybody named a state or a major city in the US, or said they speak general American English. Other languages spoken: 115 participants indicated that they are monolingual, 37 indicated some knowledge of Spanish, 26 indicated some knowledge of French, and others a smattering of other languages.

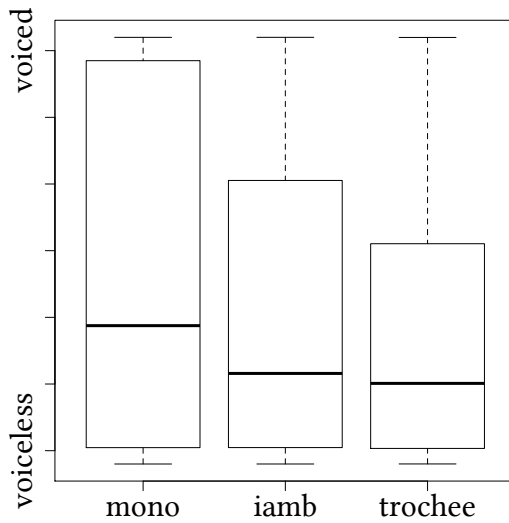
### 3.4 Results

Figure (7) shows the participants’ responses by prosodic shape, with monosyllabic items receiving the highest voiced ratings (3.57 on the 1–7 scale), followed by iambs (3.18) and trochees (2.88).<sup>2</sup>

---

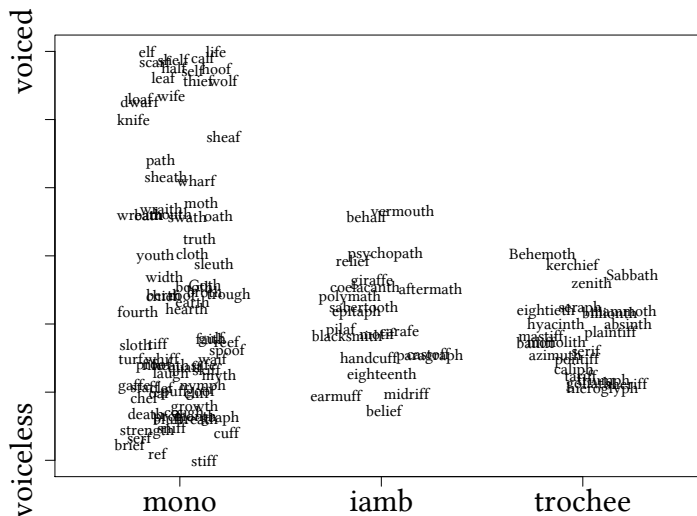
<sup>2</sup>In our discussion, “7” always refers the highest acceptability for the voiced plural. For the participants, the ends of the scale were randomly assigned to voiced or voiceless; the raw response was subtracted from 8 for the trials where the voiceless response was on the “7” side. Note also that in boxplots, the thick horizontal line in each box represents the median, not the mean.

(7) Lexicon study: Voicing alternations preferred with monosyllables



Looking at the average responses to each item (plotted with horizontal jitter in 8), we see that only monosyllables such as *elf* and *life* are unanimously voiced (7 on the 1–7 scale), while iambs reach a maximum of 4.67 with *vermouth*, and trochees are even lower with 4.03 for *Behemoth*.

(8) Lexicon study: Average responses by item



In other words, among the real words of English, all of the items that are strongly impacted by the voicing alternation are monosyllabic. None of the polysyllables we tested reached a voicing score of 5, leaving the entire third top of the range to

the monosyllables.

As we will see in §4, the iambs and trochees alternate less than monosyllables for two separate reasons. The difference between monosyllables and trochees is a simple stress effect: In the trochees, the voicing alternation causes the appearance of a voiced fricative in an unstressed syllable, where the contrast between voiced and voiceless fricatives is poorly cued (see Giavazzi 2010 for a recent survey of such effects). This is especially true in English, where the contrast is mostly cued by vowel length, which in turn is diminished in an unstressed syllable.

Our focus here is on the difference between monosyllables and iambs, where stress is kept consistently on the alternating syllable. This is a positional effect, with alternations in iambs occurring in a non-initial syllable.

Voiced responses were preferred following the long vowels [æ,<sup>3</sup> a/ɔ,<sup>4</sup> i, u, eɪ, oʊ, aɪ, aʊ] relative to the short vowels [ɛ, ø, ɪ, ʊ, ʌ] (3.41 vs. 2.66 on the 1–7 scale). This vowel length difference was strong for the monosyllables (3.94 vs. 2.83) and iambs (3.34 vs. 2.29), but not for the trochees (2.88 vs. 2.89). The preference for voiced consonants following a long vowel is well-attested cross-linguistically, and is also mirrored a similar experiment on voicing alternations in Dutch (Ernestus & Baayen 2003, see a discussion of their vowel effect in Becker et al. 2011).

Morphological complexity had the expected effect, with morphologically complex forms less acceptable with voicing alternations, but the effect was rather small (3.33 vs. 2.94).

Since the items in this experiment are real words, we checked for the effect of token frequency, plotted in (9) with a *lowess()* trend line. We used the frequencies supplied by the Microsoft Web N-gram Services (based on the Bing search engine). Token frequency had a weak, non-significant positive correlation with the responses (Spearman's rank correlation,  $\rho = .06$ ,  $p > .1$ ). This is in line with findings in Bybee (1995); Albright & Hayes (2002); Hay et al. (2004); Becker et al. (2011), 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.<sup>5</sup>

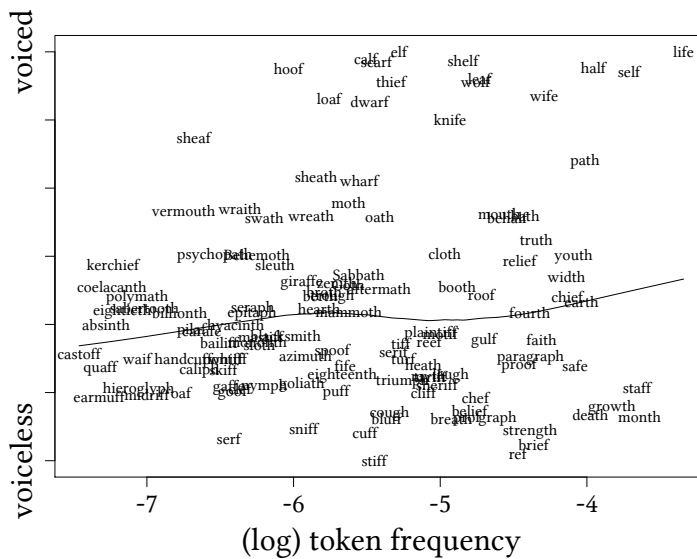
---

<sup>3</sup>While [æ] is historically short, it is phonetically one of the longest vowels of American English, as noted e.g. in Lisker (1974). The remaining vowels in the list are uncontroversially long.

<sup>4</sup>The speaker who recorded the stimuli did not have the *cot/caught* distinction.

<sup>5</sup>Looking at the chart in (9) from a 90-degree angle, we see that the most frequent items tend to be closer to the extremes of the voicing continuum, while less frequent items are closer to the middle. This makes sense, as more frequent items are uttered more often in the plural, and can get entrenched as either clearly voiceless or clearly voiced, while the behavior of the infrequent

(9) Lexicon study: Weak, non-significant token frequency effect



To assess the statistical strength of these effects, we used a mixed-effects regression model using the *lmer()* function from the *lme4* package (Bates & Maechler 2009) in R (R Development Core Team 2007). We used the following predictors: *long vowel*, a binary predictor that encoded vowel length as explained above, *shape*, a three-level unordered factor that distinguished monosyllables, iambs, and trochees, *morphological complexity*, *place* (f vs. θ), and *token frequency*. The effects that reached statistic significance are reported in (10).<sup>6</sup>

items is less fully determined. This effect is statistically significant, with token frequency being predictable from the square of the voicing score. This, however, is irrelevant to the question we ask here, which is how the speaker generalizes from known items to unknown items.

<sup>6</sup> We started with a base model that had *item* and *participant* as random effects, and no fixed effects. We then added *long vowel*, an addition that made a significant improvement to the model (ANOVA model comparison,  $\chi^2(1) = 7.0, p < .01$ ). Next, we added *shape*, which made another significant improvement to the model that contained *long vowel* (ANOVA model comparison,  $\chi^2(2) = 6.86, p < .05$ ). The interaction of *long vowel* and *shape* did not improve the model significantly. Other predictors, such as morphological complexity, place (f vs. θ), and token frequency, made no significant improvement to the model. To reduce the correlations between the predictors in the model, we first normalized *long vowel* using R's *scale()* function. Then, we helmert-coded *shape* as two binary predictors, one that contrasts monosyllables with iambs, and one that contrasts monosyllables and iambs with trochees, and normalized each predictor. Finally, we added these fixed effects (*long vowel* and the two *shape* predictors) to the model's random effects. The fully crossed model, which had *long vowel* and *shape* added to both *participant* and *item* did not converge, but between the models that only had *long vowel* and *shape* added to either, the model with them added to *item* made a significant improvement over the simpler model (ANOVA model comparison,  $\chi^2(9) = 35.7, p < .0001$ ). This is the model we report in (10). The model has low collinearity measures (VIF  $\leq 1.02, \kappa = 1.27$ ), calculated using *mer-utils*, by Austin Frank, available at <https://h1plab.wordpress.com/2011/02/24/diagnosing-collinearity-in-lme4/>.

(10) Lexicon model

	$\beta$	SE( $\beta$ )	$t$	$p$ -value <sup>7</sup>
(Intercept)	3.19	.12	25.54	
<i>mono vs. iamb</i>	-0.28	.12	-2.37	<.05
<i>mono &amp; iamb vs. trochee</i>	-0.20	.07	-2.70	<.01
<i>long vowel</i>	.26	.09	2.70	<.05

The model in (10) says that having a long vowel is conducive to more voicing (positive  $\beta$ ), while being iambic or trochaic is less conducive to voicing (negative  $\beta$  for both).

To assess the effect of age, we ran a model just on the data from the 152 participants who supplied their year of birth. Adding *age* as a predictor did not improve the model. A simple by-participant analysis showed that the correlation between average voicing scores and age was strongest for iambs, but did not reach significance (Spearman’s rank correlation,  $\rho = -.14$ ,  $p > .1$ ). We conjecture that the age effect we noted among our consultants in a Massachusetts university may be limited by class or level of education, and thus did not come out sufficiently strongly in our more diverse participant pool.

To summarize, then, we found two factors that affect the voicing of existing f-final and  $\theta$ -final nouns: The length of the noun’s final vowel, with long vowels conducive to significantly more voicing, and the shape of the noun, with monosyllables conducive to significantly more voicing than iambs, and monosyllables and iambs conducive to significantly more voicing than trochees. The effects of morphological complexity, place (f vs.  $\theta$ ), token frequency, and age were not strong enough to reach statistical significance.

### 3.5 Discussion

Our survey supplies a solid starting point for a study of English voicing alternations. We asked a large number of speakers about most of the f/ $\theta$ -final nouns that English has, and established that only monosyllables score highly on the voicing scale. Out of the 82 monosyllabic nouns we tested, 19 nouns scored over 5, while

<sup>7</sup>Exact p-values could not be obtained with *lme4* or *pvals.fnc()*, due to disagreement in the community about the best way to calculate them. We obtained p-values by taking the three predictors out of the model, one at a time, and testing their improvement to the superset model; this method is widely accepted as reliable and impervious to collinearity.

none of the polysyllabic nouns did.

This results confirms what we heard from our consultants: Many consultants allow the alternations only in monosyllables. Others, especially young ones, like the alternations in iambs (or more precisely, in polysyllables with a stressed final syllable).

## 4 Experiment 2: No monosyllabicity effect in *wugs*

We have seen in §3 above that in the English lexicon, monosyllables are impacted by voicing alternations more strongly than polysyllables. This finding is surprising given the typological background we discussed in §2: in most of the languages we have data for, monosyllables enjoy greater protection from alternations. To gain insight into the grammar that speakers use to regulate alternations, we need to go beyond the existing words of the language. To our knowledge, ours may be the first study to wug-test the English fricative voicing alternations since Berko (1958), who included the nonce item *heaf* in her classical experiment.

### 4.1 Materials

To test the applicability of vowel length and prosodic shape on novel nouns, we created 132 target nonce words (“wugs”). We made these by crossing 2 places of articulation (f, θ) \* 3 prosodic shapes (monosyllable, iamb, trochee) \* 11 vowel categories (æ, aɪ, a/ɔ,<sup>8</sup> eɪ, ε, ø, i, ɪ, oʊ, u/ʊ,<sup>9</sup> ʌ), and then made two words in each of these combinations, manually selecting consonants to match general English frequencies, allow reasonable wellformedness, and keep sufficient distance from real fricative-final nouns. The full list of nonce items, with results by item, is in Appendix B. In addition, we created 30 filler nonce items (13 monosyllables, 5 iambs, 12 trochees), which were modeled on the fillers in experiment 1.

The items were recorded by a phonetically trained male native speaker of English in his thirties from California, following the same procedure as in §3.

---

<sup>8</sup>The speaker who recorded the stimuli did not have the *cot/caught* distinction.

<sup>9</sup>We made half as many items for [ʊ] and [u] as we did for the other vowels, to better represent the rarity and marginal status of [ʊ]. We considered [u] long and [ʊ] short.

## 4.2 Methods

The experiment was presented to the participants over the internet, closely matching the procedure described in §3.2. The web server started by making a random selection of items for each participant, choosing a total of 36 items: 18 target items and 18 fillers. The item selection was done as follows: Six monosyllables (3 f, 3 θ), six trochees (3 f, 3 θ), and six iambs (3 f, 3 θ). The 18 fillers were chosen randomly.

Before the experiment began, the participant was reminded that some English plurals are exceptional, giving *ox*, *knife*, and *phenomenon* as examples. The participant was asked to give their opinion about made-up words of English, using the nonce *pharon* as the example word. The rest of the experiment proceeded as described in §3.2, asking for ratings on a 1–7 scale.

## 4.3 Participants

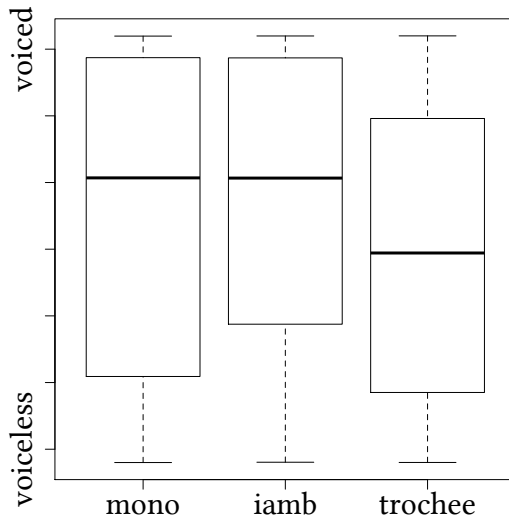
The participants were 200 users of Amazon’s Mechanical Turk who did not participate in the previous experiment. Participants were paid \$.30 for their time. The server logs indicate that these participants took on average 14 minutes to complete the experiment (range 6–54 minutes, median 12).

The participants volunteered the following demographic information. Year of birth was provided by 140 participants, who reported an average age of 31 (range 18–62, median 27). Sex: 126 females, 62 males, 12 didn’t say. Country of origin: 173 from the United States, 1 from Canada, 26 didn’t say. As for the kind of English spoken, almost everybody named a state or a major city in the US, or said they speak general American English. Other languages spoken: 118 participants indicated that they are monolingual, 41 indicated some knowledge of Spanish, 27 indicated some knowledge of French, and others a smattering of other languages.

## 4.4 Results

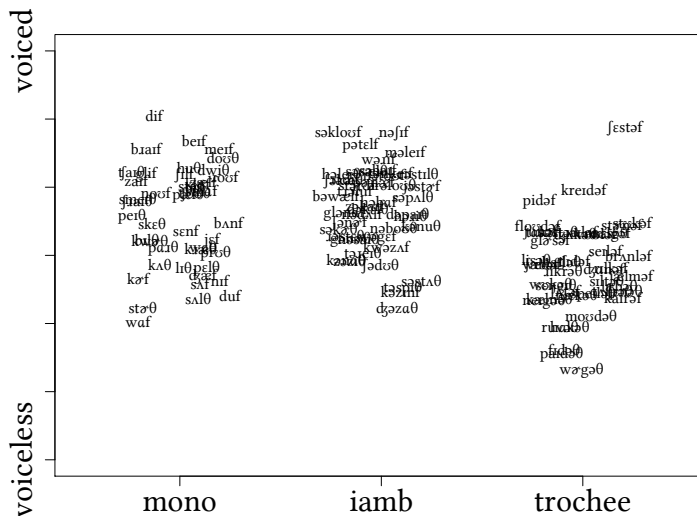
Figure (11) shows the participants’ responses by size, with iambs receiving the highest voicing ratings (4.87 on the 1–7 scale), followed closely by monosyllables (4.38). This is the opposite of what we found in the lexicon study, where monosyllables receive the highest ratings for voiced plurals. Trochees came in lowest (4.06), as they did in the lexicon study.

(11) Novel iambs and monosyllables alternate more than trochees



The same pattern is seen when looking at the average responses by item in (12). Note that unlike real items, most of which have relatively established plural forms, nonce words don't usually elicit extreme responses. We take this to be a validation of the statistical model in (10), which predicts highly reliable but numerically small differences between items based on prosodic shape and vowel length.

(12) Voicing alternations for novel nouns, by item



Here again we see the difference between monosyllables and trochees, i.e. the same stress effect that was seen in the lexicon. Alternations are judged worse

in unstressed syllables, both in the lexicon and in nonce words. But looking at monosyllables and iambs, where stress is held constant, we see no difference. The positional effect that was seen in the lexicon is gone, and nonce monosyllables are not more strongly impacted by the alternation than nonce iambs.

For the statistic analysis, we used a mixed-effects regression model with the *lmer()* function, as in §3. As predictors, we used *long vowel*, *shape*, and *place*, defined as in §3. The effects that reached significance are reported in (13).<sup>10</sup>

The model in (13) says that having a long vowel is conducive to more voicing (positive  $\beta$ ), and being trochaic is conducive to less voicing (negative  $\beta$ ) – both effects also seen in the lexicon study. The effect of vowel length is strongest (.16) for monosyllables, and weaker for iambs (.16–.09=.07) and for trochees (.16–.11=.05). Being  $\theta$ -final is also conducive to less voicing, even though this effect wasn't so strong in the lexicon. Most importantly, however, the iambs are slightly *more* conducive to voicing than monosyllables, unlike what we have seen in the lexicon.

(13) Nonce words model

	$\beta$	SE( $\beta$ )	<i>t</i>	<i>p</i> -value
(Intercept)	4.34	.08	54.75	
<i>mono vs. iamb</i>	.05	.05	.95	>.1
<i>mono&amp;iamb vs. trochee</i>	–.36	.05	–6.86	<.001
<i>long vowel</i>	.16	.05	3.08	<.005
<i>place=θ</i>	–.26	.06	–4.58	<.001
<i>mono vs. iamb:long</i>	–.09	.05	–1.67	<.1
<i>mono&amp;iamb vs. trochee:long</i>	–.11	.05	–2.12	<.05

To assess the effect of age, we ran a model just on the data from the 140 participants

<sup>10</sup> We started with a base model that had *item* and *participant* as random effects, and no fixed effects. We then added *shape*, an addition that made a significant improvement to the model (ANOVA model comparison,  $\chi^2(2) = 35.3$ ,  $p < .0001$ ). Next, we added *long*, which made another significant improvement to the model that contained *shape* (ANOVA model comparison,  $\chi^2(1) = 8.04$ ,  $p < .01$ ). We then added *place*, making another improvement to the model (ANOVA model comparison,  $\chi^2(1) = 23.7$ ,  $p < .0001$ ). The interaction of *shape* and *long* made another significant improvement to the model (ANOVA model comparison,  $\chi^2(2) = 6.87$ ,  $p < .05$ ), but no other interactions did. We reduced the correlations between the predictors in the model as explained in §3. Finally, we added the fixed effects (the two *shape* predictors, *long*, the interaction, and *place*) to the model's random effects. The fully crossed model did not converge, nor did the model that had *shape*, *long*, and *place* added to *item*. The model with *shape*, *long*, and *place* added to *participant* made a significant improvement over the simpler model (ANOVA model comparison,  $\chi^2(27) = 53.3$ ,  $p < .005$ ), and this is the model we report in (13). This model has low collinearity measures (VIF<=1.007,  $\kappa=1.09$ ). We calculated *p*-values as explained in §3.

who supplied their year of birth. Adding *age* as a predictor did not improve the model. A simple by-participant analysis showed a significant and negative correlation between age and voicing ratings, but this correlation was comparable for all three shapes (Spearman's rank correlation, for monosyllables:  $\rho = -.20$ ,  $p < .05$ , iambs:  $\rho = -.21$ ,  $p < .05$ , and trochees:  $\rho = -.19$ ,  $p < .05$ ). This isn't the same age effect we've seen in the lexicon, then, which was concentrated in the iambs.

To summarize, we found three factors that affect the voicing of novel f-final and  $\theta$ -final nouns: The length of the noun's final vowel, with long vowels conducive to significantly more voicing, the place of the final consonant, with  $[\theta]$  being conducive to significantly less voicing than  $[f]$ , and the shape of the noun, with monosyllables and iambs conducive to significantly more voicing than trochees. The difference between monosyllables and iambs trended in favor of the iambs, but did not reach significance; this contrasts with the effect we found in the real words of English, where iambs were significantly *less* conducive to voicing than monosyllables.

## 4.5 Discussion

The main point of interest for this experiment was the propensity of monosyllables to alternate relative to iambs. In the real words of English, monosyllables are impacted significantly more than iambs; in the nonce word experiment, monosyllables are trending in the other way, getting slightly more protection than iambs. We take this to mean that the generalization in the English lexicon is unlearned; speakers do not learn that monosyllables can be impacted more than iambs. Speakers are biased to protect monosyllables, and this bias shapes the generalizations that they can extract from the real words they know. This inability to learn unnatural generalization is what Becker, Ketrez & Nevins (2011) call the *surfeit of the stimulus*. In using over 100 stimuli and 200 participants, we responded to Hayes et al.'s (2009) call for larger data sets.

We note that speakers do learn natural generalizations from their lexicon. Long vowels are conducive to more alternations in the lexicon, and this natural effect is extended productively to nonce words.

## 5 Experiment 3: English speakers protect monosyllables in an artificial grammar

We have seen that in the English lexicon, monosyllables are impacted by alternations more than polysyllables, and that this anomaly is absent from the treatment of nonce words. Here, we go a step further to show that English speakers are biased to protect monosyllables in the setting of an artificial language learning experiment.

In other words, we want to show that alternations in monosyllables imply alternations in polysyllables, but not the other way around. Artificial language experiments have been successfully used to show implicational relationships in phonology, most notably in Wilson (2006) and by several others since; see Moreton & Pater (2011) for a recent review of this literature.

In their real language, English speakers regulate voicing alternations in the plural on [f, θ]. In this artificial language, we kept plural as the morphological category, but changed the plural suffix to [ni], and the set of alternating consonants to [p, t, k].

### 5.1 Materials

We constructed 104 nouns, 60 stop-final target items and 44 sonorant-final fillers. The 60 target items, listed in appendix C, were made by crossing 3 final consonants (p, t, k) \* 2 shapes (monosyllable, iamb), and making 10 items in each category by manually filling in stressed vowels from the set [a, e, i, o, u, ai] and all the consonants of English except for the interdental and glides. In the iambs, the initial unstressed vowel was consistently schwa. The fillers were 22 monosyllables and 22 iambs, with final consonants drawn from the set [l, r, m, ŋ], with non-final segments using the same vowels and consonants that were used for the target items.

The items were recorded by a phonetically trained male native speaker of English in his twenties from Pennsylvania. The list included each noun in the “singular” (i.e. the bare form) and two “plurals” (one made by simple suffixation of [ni], and one by voicing of the stem-final stop and suffixation of [ni]). The recording was done using the same procedure as in §3 and §4.

To assign the experimental items plausible meanings as concrete nouns, we collected 100 pairs of pictures of easily recognizable objects (e.g. *airplane*, *anteater*,

*apple*) using Google’s image search with the “line drawing” option. The “singular” was represented by one of these pictures, and the “plural” was represented by showing five copies of the other picture in two rows.

## 5.2 Methods

The experiment was presented over the internet as an artificial grammar learning game. In the beginning of the experiment, the web server randomly assigned each participant to one of two training groups: monosyllabic or iambic. The experimental setup is summarized in (14), and detailed below.

The training phase consisted of 15 trials, with 10 randomly chosen stop-final items (4 p-final, 3 t-final, 3 k-final) and five randomly chosen sonorant-final items for each participant. Participants in the monosyllabic training group received monosyllabic stop-final nouns, 2 monosyllabic sonorant-final nouns, and 3 iambic sonorant-final nouns. Participants in the iambic training group received the exact opposite: iambic stop-final nouns, 2 iambic sonorant-final nouns, and 3 monosyllabic sonorant-final nouns. Each item was randomly assigned to a pair of pictures as its meaning.

### (14) Experiment 3: Artificial grammar setup

	monosyllabic training	iambic training
Training	<i>10 stop-final monos</i>	<i>10 stop-final iambs</i>
	'mip      'mibni	tə'gep    tə'gebni
	'stut      'studni	gə'ʃut    gə'ʃudni
	<i>5 sonorant-finals</i>	<i>5 sonorant-finals</i>
	'muŋ      'muŋni	'muŋ      'muŋni
	nə'dʒol    nə'dʒolni	nə'dʒol    nə'dʒolni
Testing	<i>10 stop-final monos</i>	<i>10 stop-final monos</i>
	'gairp      _____	'gairp      _____
	'klet        _____	'klet        _____
	<i>10 stop-final iambs</i>	<i>10 stop-final iambs</i>
	fə'ʃop      _____	fə'ʃop      _____
	bə'git       _____	bə'git       _____
	<i>10 sonorant-finals</i>	<i>10 sonorant-finals</i>
	'pler        _____	'pler        _____
	ʒə'taim    _____	ʒə'taim    _____

The training phase started with participants hearing the singular and seeing it

written with English-like spelling, and then hearing and seeing the plural with the same English-like spelling. After 7 such trials, the next 8 trials showed and played the singular, but the plural was only played and not written, and the participant was asked to type it in. Once typed in, the participant was shown the intended spelling, and an indication of whether their spelling matched the intended one. In this phase, plurals always contained a voicing alternation, e.g. *'mip* ~ *'mibni*, which was reflected in the proposed spelling. As is normal in American English, the voicing alternation was accompanied by lengthening of the preceding vowel.

Then, the testing phase consisted of 30 trials. The server randomly chose 20 stop-final items: 10 stop-final items of the shape that was seen before (monosyllabic for the monosyllabic training group and iambic for the iambic training group) and 10 stop-final items of the shape that was not seen before. Of the 10 stop-final items of the shape that was seen before, 4 were actually seen before in the training phase, and 6 were completely new. The fillers were 10 sonorant-final nouns, 5 of each shape.

In the testing phase, participants started with hearing the singular and seeing it written with English-like spelling, and they were asked to supply the plural. No feedback was given at this phase. After 15 such trials, the next 15 trials only played the singular, but it was not shown written. Participants were asked to supply the plural, then, based on their ability to hear the singular.

Throughout the task, the vowels [a e i o u ai ə] were spelled ⟨a e i o u ai -⟩, and the consonants [k ʒ] were spelled ⟨k zh⟩. The spelling of schwa as ‘-’ was introduced with the example ‘puh-lease’, and our participants learned that well.

### 5.3 Participants

The participants were 100 users of Amazon’s Mechanical Turk who did not participate in any of the previous experiments, 50 assigned to the monosyllabic training group and 50 assigned to the iambic training group. They were paid \$.50 for their time. The server logs indicate that these participants took on average 10 minutes to complete the experiment (range 5–25 minutes, median 10).

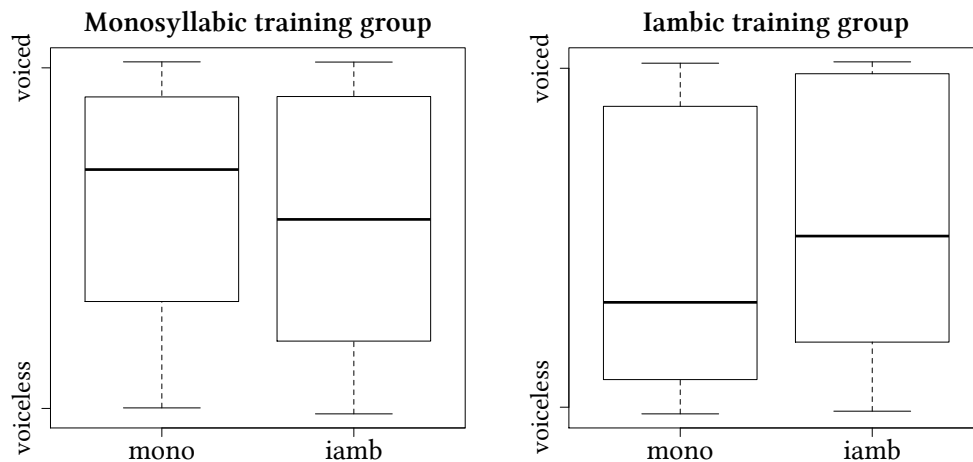
The participants volunteered the following demographic information. Year of birth was provided by all participants, who reported an average age of 31 (range 18–62, median 28). There was no significant age difference between the groups (difference in mean <1 year, two-sample t-test,  $t(97.95) = .13$ ,  $p > .1$ ). Sex: 53 females, 47 males. There was no significant sex difference between the groups

(monosyllabic training: 22 females, iambic training: 31 females, Fisher exact test,  $p > .1$ ). Ninety-five participants said they grew up speaking English, 3 said they didn't, 2 didn't say. As for the kind of English spoken, almost everybody named a state or a major city in the US, or said they speak general American English. Other languages spoken: 44 participants indicated that they are monolingual, 27 indicated some knowledge of Spanish, 9 indicated some knowledge of French, and others a smattering of other languages.

## 5.4 Results

To interpret the plurals that participants typed in as responses for stop-final singulars, we looked at responses that ended in the correct plural suffix (i.e., the spelling ⟨ni⟩, which was given in 97% of the responses), and then picked the grapheme that preceded this suffix. The graphemes ⟨p⟩, ⟨b⟩, ⟨t⟩, ⟨d⟩, ⟨k⟩, ⟨g⟩ were given in 94% of all responses, and those were considered valid responses to stop-final items; other responses were considered invalid. The graphemes ⟨b⟩, ⟨d⟩, ⟨g⟩ were considered voiced responses, and ⟨p⟩, ⟨t⟩, ⟨k⟩ voiceless. Furthermore, each response was categorized as to whether it matched the place of articulation of the singular's final stop, e.g. either ⟨p⟩ or ⟨b⟩ in the plural were considered to match a ⟨p⟩ in the singular (84% of all responses).

### (15) Experiment 3: Artificial grammar experiment results



In the monosyllabic training group, voiced responses were given to monosyllabic items more than to iambs (61% and 57%, respectively, a small but significant difference). In the iambic training group, voiced responses were given significantly

more often to iambic items than to monosyllabic items (55% vs. 45%, respectively, a larger significant difference), as seen in figure (15). In both groups, the shape that was seen in the training (monosyllables with monosyllabic training and iambs with iambic training) received voiced responses at essentially the same rate (61% and 55%, respectively, an insignificant difference), suggesting that both groups noticed the alternation equally well.

The statistical analysis was done with an *lmer* mixed-effects logistic regression model. We used the following two predictors and their interaction: *group*, a binary factor that distinguished monosyllabic training from polysyllabic training, and *untrained*, a binary factor that coded the prosodic shape that participants have not seen in the training, i.e. iambs for monosyllabic training and monosyllables for iambic training. The resulting model is reported in (16).<sup>11</sup>

(16) Experiment 3: Artificial Grammar experiment model

	$\beta$	SE( $\beta$ )	$z$	$p$ -value
(Intercept)	.49	.32	1.59	>.1
<i>group</i>	-.28	.31	-.92	>.1
<i>untrained</i>	-.30	.07	-4.16	<.0001
<i>group:untrained</i>	-.21	.08	-2.79	<.01

We see that there is no overall difference between the groups in their propensity to offer voiced responses for the shape they were trained on, meaning that there was no significant difference between the groups in their ability to learn the alternation and apply it to novel forms of a familiar prosodic shape. Furthermore, participants in both groups apply the voicing alternation significantly less on the shape they were not trained on, i.e. they extend the alternation cautiously. Of most interest here is the significant interaction: Participants were significantly more reluctant to extend the alternation in the iambic training group than they were in the monosyllabic training group. This is exactly what we expected: with monosyllables being protected by default, participants are reluctant to impact them if they only see the alternation impacting iambs.

<sup>11</sup> We started with a base model that had *group*, *untrained*, and their interaction as fixed effects. To reduce the correlations in the model, both fixed effects were normalized using R's *scale()* function. As random effects, we used *item* and *group\*untrained* given *participant* (the fully crossed model didn't converge). The model we report in (16) uses all 2000 data points (20 stop-final stimuli \* 100 participants); the models that use only valid responses or only place-matching responses are essentially identical. We note that the model has low collinearity measures (VIF<=1.12,  $\kappa$ =1.0002).

Finally, we note that a smaller group of 22 Harvard students participated in a pilot for this experiment, with the same materials and similar methods as we describe above. The Harvard students were asked to speak their responses rather than type them (which then required transcription, and agreement among transcribers about the intended responses). This much smaller group gave responses that are qualitatively identical to the results we obtained from the Mechanical Turk participants. We prefer the Mechanical Turk's bigger participant pool for its greater coverage in geography, age, and other sociolinguistic factors.

## 5.5 Discussion

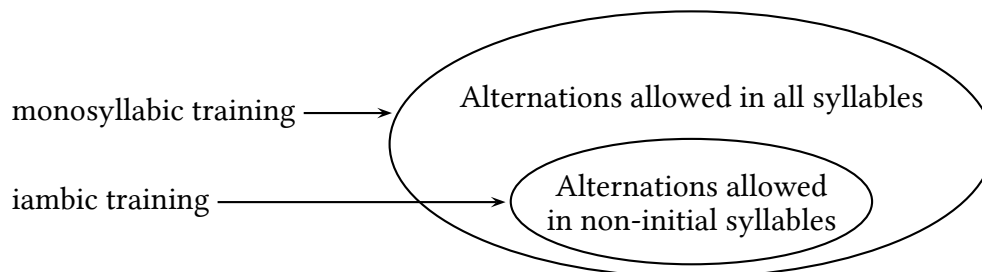
In the training phase, participants were exposed to a fragment of a 'language', only seeing alternations impacting monosyllables or only seeing alternations impacting iambs. We then tested the expectations they formed by testing them on the shape they haven't seen.

The participants in the monosyllabic training group were given a language fragment that resembles real English, where the monosyllables are the best alternators. Recall that for many speakers of English, voicing alternations only impact monosyllables. The participants in the iambic training group were given a language fragment that resembles Turkish, French, or Brazilian Portuguese (see §2), where voicing alternations mostly impact polysyllables.

Participants in both groups were cautious in applying the alternation to the shape they haven't seen — presumably, a simple task effect with little linguistic meaning. Interestingly, however, participants in the iambic training group showed significantly more caution, protecting monosyllables. In other words, the participants followed the typologically common pattern, not the English-specific pattern. This is a clear case of the poverty of the stimulus: Without information about the ability to impact monosyllables, speakers deployed an internal bias to protect initial syllables. They did not assume that they can freely extend the alternation from polysyllables to monosyllables.

This bias, which we would like to attribute to Universal Grammar, is illustrated with the Venn diagram in (17). Since alternations in initial syllables include alternations in non-initial syllables, monosyllabic training licenses alternations in all nouns, but iambic training does not license alternations in monosyllables.

(17) Alternating monosyllables are part of a superset language



Note that all the stimuli had final stress, which always put the alternating stop in the coda of the stressed syllable, effectively keeping stress constant. The difference between the monosyllables and the iambs is the prosodic position: Only in monosyllables does the alternation impact the word-initial syllable.

We conclude that given a chance, English speakers ignore the license that the English lexicon gives to impact initial syllables, and prefer the typologically attested protection of initial syllables. The participants went directly against the evidence from their ambient language, showing the emergence of their innate biases.

## 6 Experiment 4: Initial syllables also protected in polysyllables

The three experiments we described in §3, §4, and §5 contrasted monosyllables with polysyllables, where we saw an asymmetrical preference to protect monosyllables from alternation. We claimed that the asymmetry is best understood in terms of *initial syllable faithfulness*, which can limit alternations in monosyllables but not in polysyllables. To strengthen the connection between the monosyllabicity effect and initial syllable faithfulness, we go on here to show that initial syllables are protected in polysyllables as well.

Loosely inspired by German umlaut,<sup>12</sup> we created a language where all stems are disyllabic, either trochaic or iambic. For the plural, non-low vowels switch their backness, e.g. *'zuməp ~ 'ziməp* and *sə'fup ~ sə'fɪp*.<sup>13</sup> The alternation always happens

<sup>12</sup>English umlaut, which always impacts the initial syllable, is limited to the seven nouns *man*, *woman*, *goose*, *tooth*, *foot*, *mouse*, *louse*. German umlaut is more productive (Wiese 1996; Fanselow & Féry 2002; van de Vijver & Baer-Henney 2010). In the plural, it always impacts the initial stressed syllable, as in [*flus ~ 'fly:sə*] 'river' and [*bru:dər ~ 'bry:dər*] 'brother'. The diminutive [-çən] can impact non-initial syllables so long as they are stressed, e.g. [*admi'ra:l ~ admi'rɛ:l-çən*] 'admiral'.

<sup>13</sup>We limited ourselves to changes that keep vowel height constant, as these have a limited effect

in the stressed syllable, so stress is controlled for, and what changes is the position of the alternation: initial or non-initial. Note that we had to switch from a local process (where the affix causes a change on immediately adjacent segments) to a non-local process, where stem-internal segments are impacted; there is no other way for a process to impact either syllable in a set of disyllables.

## 6.1 Materials

We constructed 83 nouns: 48 target items with non-low vowels, and 35 fillers with [a]. The 48 target items, listed in appendix D, were made by crossing 2 non-low vowels (e, u) \* 2 shapes (trochee, iamb), and making 12 items in each category by manually filling in consonants. The unstressed vowel in each stem was schwa. The fillers consisted of 17 trochees and 18 iambs, with [a] in their stressed syllable and schwa in their unstressed syllables, filling in consonants manually.

The “plurals” for the targets were created by flipping vowel backness, i.e. singular [e] was switched with [o], and singular [u] was switched with [i], as shown in (18). For the fillers, the suffix [-ni] was added to make the plural, leaving the stem vowels unchanged, e.g. [tə'kaʃ ~ təkəʃni], ['sagətʃ ~ 'sagətʃni].

The vowel inventory of the language, then, was [a e i o u ə], which the participants saw spelled ⟨a e i o u –⟩, e.g. [tɹə'mel] was spelled ⟨tr–mel⟩.

The items were recorded by a phonetically trained male native speaker of English in his twenties from Wisconsin, using the same procedure described in §5. To assign the experimental items plausible meanings, the same pictures were used as in §5.

## 6.2 Methods

The experiment was delivered over the internet, and presented as an artificial grammar learning game. In the beginning of the experiment, the web server randomly assigned each participant to one of two training groups: trochaic or iambic. The experimental setup is summarized in (18) and detailed below.

The training stage consisted of 10 trochaic target items with non-low vowels (5 [e], 5 [u]) in the trochaic training group, and 10 such iambs for the iambic training group. Both groups were presented with 6 fillers with [a], 3 iambs and 3 trochees. In the testing stage, all participants were presented with 10 trochees (5 [e], 5 [u]),

---

on duration. We did not use the pair [a ~ æ], since for many speakers, [æ] isn't as low as [a].

10 iambs (5 [e], 5 [u]), and 10 fillers with [a] (5 iambs, 5 trochees).

(18) Experiment 4: Artificial grammar setup

	Trochaic training	Iambic training
Training	<i>10 trochees</i>	<i>10 iambs</i>
	'zuməp    'ziməp 'bɪɛzəl    'bɪɔzəl	sə'fup    sə'fip tɪə'mel    tɪə'mol
	<i>6 fillers</i>	<i>6 fillers</i>
	mə'fat    mə'fatni 'gastəf    'gastəfni	mə'fat    mə'fatni 'gastəf    'gastəfni
Testing	<i>10 trochees</i>	<i>10 trochees</i>
	'funəl    _____ 'febəf    _____	'funəl    _____ 'febəf    _____
	<i>10 iambs</i>	<i>10 iambs</i>
	pə'dul    _____ kə'zem    _____	pə'dul    _____ kə'zem    _____
	<i>10 fillers</i>	<i>10 fillers</i>
	bɪə'gam    _____ 'baləd    _____	bɪə'gam    _____ 'baləd    _____

The training and testing were done following the same procedure as in §5, ending with 15 trials where participants have to provide plurals for stems that they hear but do not see written.

### 6.3 Participants

The participants were 100 users of Amazon's Mechanical Turk who did not participate in any of the previous experiments, 50 assigned to the monosyllabic training group and 50 assigned to the iambic training group. They were paid \$.50 for their time. The server logs indicate that these participants took on average 10 minutes to complete the experiment (range 5–29 minutes, median 9).

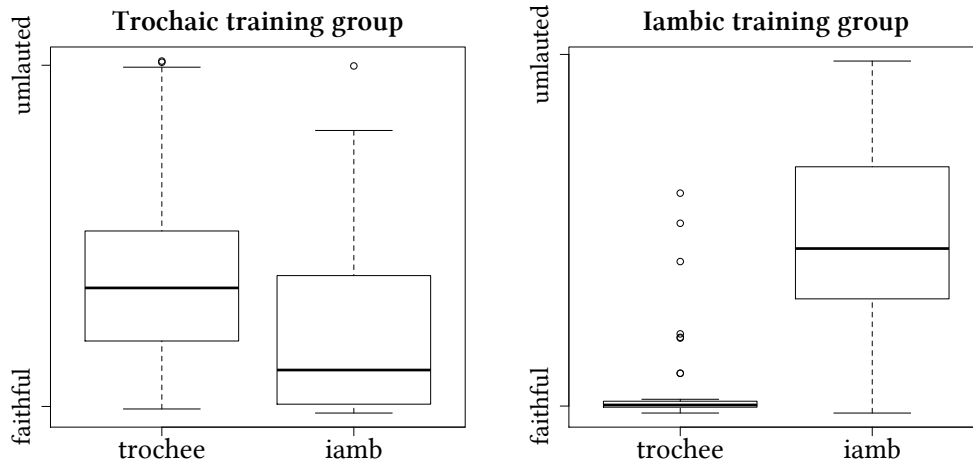
The participants volunteered the following demographic information. Year of birth was provided by all participants, who reported an average age of 29 (range 18–60, median 27). There was no significant age difference between the groups (difference in mean  $\sim 2$  years, two-sample t-test,  $t(85.93) = 1.13$ ,  $p > .1$ ). Sex: 67 females, 33 males. There was no significant sex difference between the groups (monosyllabic training: 33 females, iambic training: 34 females, Fisher exact test,  $p > .1$ ). Ninety-six participants said they grew up speaking English, 3 said they

didn't, one didn't say. As for the kind of English spoken, almost everybody named a state or a major city in the US, or said they speak general American English. Other languages spoken: 61 participants indicated that they are monolingual, 14 indicated some knowledge of Spanish, 7 indicated some knowledge of French, and others a smattering of other languages.

#### 6.4 Results

To interpret the plurals that participants typed in, we removed any ⟨ni⟩ suffixes, and removed all consonants, only leaving the vowel graphemes ⟨a e i u o⟩ and ⟨-⟩. Responses that had one full vowel grapheme and one dash/schwa in the same order as the singular were considered to match the singular (69% of all responses). Plurals that had ⟨i⟩ in response to a singular [u] or ⟨o⟩ in response to a singular [e] were considered successful vowel mapping (32% of all responses). Finally, fully successful (“umlauted”) responses were those that matched the prosodic shape of the singular and also had a successful vowel mapping (28% of all responses). All other responses were considered unsuccessful.

##### (19) Experiment 4: Artificial grammar experiment results



In the trochaic training group, umlauted (fully successful) responses were given to trochaic items more than to iambs (39% and 22%, respectively, a significant difference). In the iambic training group, umlauted responses were given significantly more often to iambic items than to trochaic items (47% vs. 5%, respectively, a much larger significant difference), as seen in figure (19). We note that umlaut was applied to iambs in the iambic group more often than umlaut was applied

to trochees in the trochaic group (47% vs. 39%), suggesting a small yet significant preference to apply umlaut word-finally.

The statistical analysis was done with an *lmer* mixed-effects logistic regression model, again with the predictors *group* and *untrained* and their interaction, as in §5. The resulting model is reported in (20).<sup>14</sup>

(20) Experiment 4: Artificial grammar experiment model

	$\beta$	SE( $\beta$ )	$z$	$p$ -value
(Intercept)	-1.37	.15	-9.21	<.0001
<i>group</i>	-.41	.15	-2.66	<.01
<i>untrained</i>	-1.16	.12	-9.90	<.0001
<i>group:untrained</i>	-.55	.11	-4.95	<.0001

We see that participants in both groups apply the voicing alternation significantly less on the shape they were not trained on, i.e. they extend the alternation cautiously. Of interest here is the significant interaction: Participant were significantly more reluctant to extend the alternation in the iambic training group than they were in the trochaic training group. This is exactly what we expected: with initial syllables being protected by default, participants are reluctant to impact the initial syllables of trochees if they only see the alternation impacting non-initial syllables in iambs. We note additionally that there was an overall preference to impact iambs, suggesting that English speakers prefer umlaut in word-final position; the protection of trochees we see here goes beyond this baseline preference for word-final umlaut.

## 6.5 Discussion

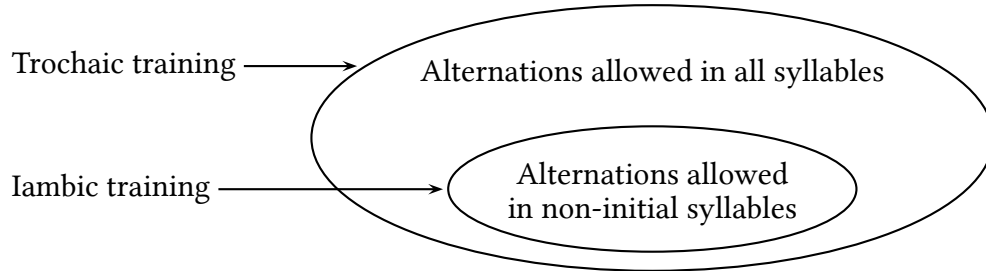
In this experiment, participants learned the vowel changes [e → o] and [u → i] in stressed syllables. Unstressed syllables were kept as [ə] throughout. Covarying the stress and the position of the alternation allowed us to keep stress constant while choosing to impact the initial or non-initial syllable.

Participants who were trained on iambs, where the alternation impacts the non-initial syllable, were strongly reluctant to impact initial syllables in the

<sup>14</sup>To reduce the correlations in the model, both fixed effects were normalized using R's *scale()* function. As random effects, we used *item* and *participant* (the fully crossed model did not converge). The model we report in (20) uses all 2000 data points (20 non-low vowel stimuli \* 100 participants); the model that uses the 1384 data points that match the prosodic shape of the singular is essentially identical. The model has low collinearity measures (VIF<=1.47,  $\kappa$ =1.0002).

testing phase; this contrasts with those trained on trochees, where the alternation impacted the initial syllable, and who were more willing to extend the alternation to the non-initial syllable. This results confirms our hypothesis that alternations in initial syllables imply alternations in later syllables, as shown in the Venn diagram in (21). We see that the protection of monosyllables in §5 was a case of the more general protection of initial syllables.

(21)



We note one difference between the results here and the results in §5. The baseline application of the alternation to the shape that participants have seen in training was the same in both groups in §5, while here we see a small yet significant overall bias towards applying the alternation in final syllables. It is possible that the irregular pasts in English are relevant here, since vowel changes always impact the final syllable of the verb, as in *become* ~ *became*. In fact, the [e → o] mapping is true of one English verb, *break* ~ *broke*. This may bias English speakers to notice vowel changes more when they occur in the final syllable. At any rate, this effect is orthogonal to the effect of interest here, which is the significantly stronger reluctance of participants in the iambic training group to extend the vowel alternations to initial syllables.

## 7 Discussion: Initial syllable faithfulness protects monosyllables

As discussed in §2, most known cases of local morphophonological alternations impact stem-final elements in polysyllables more strongly than they impact the same stem-final elements in monosyllables. In this section, we connect this result to the literature on *initial syllable faithfulness* (Steriade 1994; Beckman 1997, 1998; Casali 1998; Barnes 2006; Jesney 2009; Becker 2009; Becker et al. 2011).

Given a set of constraints that includes both general faithfulness (IDENT) and a faithfulness constraint that only penalizes unfaithful mappings in the initial

syllable (IDENT- $\sigma_1$ ), alongside a relevant markedness constraint (MARK), there are only three languages that are possible, as schematized in (22). With MARK outranking both faithfulness constraints, alternations will be observed in all stems, regardless of size. With general IDENT outranking MARK, no alternations will be observed, regardless of the ranking of IDENT- $\sigma_1$ . When MARK is outranked by IDENT- $\sigma_1$  but not general IDENT, monosyllables are protected from alternations, but polysyllables aren't. Crucially, this set of constraints cannot describe the English lexicon, where monosyllables are more strongly impacted by markedness than polysyllables.

- (22) MARK  $\gg$  IDENT, IDENT- $\sigma_1$       Alternations observed everywhere  
IDENT, IDENT- $\sigma_1$   $\gg$  MARK      Alternations observed nowhere  
IDENT- $\sigma_1$   $\gg$  MARK  $\gg$  IDENT      Alternations observed in polysyllables

The results of the ranking possibilities in (22) are important, because they demonstrate that once the learner brings to the task an inventory of constraints with specialized reference to initial syllables but without any specialized reference to non-initial syllables, it will not be possible to encode (and therefore generalize) a pattern of greater alternation in initial syllables as opposed to non-initial syllables.

In what follows, we sketch out an implementation in which the predictions made in (22) – stated in categorical, exceptionless terms – are not disrupted by the presence of the lexical exceptions. We demonstrate that the basic pattern holds when supplemented with the USELISTED approach to exceptionality (Zuraw 2000; Hayes & Londe 2006), noting that many other approaches to exceptionality may be compatible as well.

USELISTED is a constraint that penalizes the productive combination of stems and affixes when the learner already has the combination listed in their lexicon. More familiar forms are more strongly listed, and thus respond more strongly to USELISTED and less strongly to other faithfulness constraints and to markedness constraints. This means that for an adult English speaker, a common plural like 'knives' is protected by USELISTED, as shown in (23). Candidates (a) and (b) were generated directly from the listed form [narvz], whereas candidates (c) and (d) are made by combining the root [naf] with the plural suffix. The constraints IDENT and IDENT- $\sigma_1$ , which here only penalize unfaithful mappings in the root, prefer the faithful fricative (voiceless in the UR, voiced in the listed plural). Markedness prefers the voiced fricative (see below), but both faithfulness and markedness are outranked by USELISTED.

(23) ‘knives’ protected by USELISTED

/naɪf + z/, listed: [naɪvz]	USELISTED	IDENT	IDENT-σ <sub>1</sub>	MARK
a. $\text{naɪvz} \rightarrow \text{naɪvz}$				
b. $\text{naɪvz} \rightarrow \text{naɪfs}$		*	*	*
c. $\text{naɪf+z} \rightarrow \text{naɪvz}$	*!	*	*	
d. $\text{naɪf+z} \rightarrow \text{naɪfs}$	*!			*

With a nonce word such as [bɪaɪf], the speaker has no plural form listed, so combining the base and the plural suffix is the only option, as shown in (24). Both candidates violate USELISTED equally, and the decision is handed down to the stochastically ranked or weighted markedness and faithfulness constraints.

(24) ‘brife’ controlled by markedness and faithfulness

/bɪaɪf + z/, listed: []	USELISTED	IDENT	IDENT-σ <sub>1</sub>	MARK
a. $\text{bɪaɪf+z} \rightarrow \text{bɪaɪvz}$	*	*	*	
b. $\text{bɪaɪf+z} \rightarrow \text{bɪaɪfs}$	*			*

USELISTED is similarly irrelevant for a novel polysyllable such as [sə'baɪf], in which a voicing alternation only incurs a violation of IDENT, not IDENT-σ<sub>1</sub>. This means that regardless of the ranking/weighting of IDENT-σ<sub>1</sub>, [sə'baɪvz] has to be at least as acceptable as [bɪaɪvz]; the grammar cannot replicate the lexical tendency to voice more in monosyllables.

(25) ‘sebife’ controlled by markedness and general faithfulness

/sə'baɪf + z/, listed: []	USELISTED	IDENT	IDENT-σ <sub>1</sub>	MARK
a. $\text{sə'baɪf+z} \rightarrow \text{sə'baɪvz}$	*	*		
b. $\text{sə'baɪf+z} \rightarrow \text{sə'baɪfs}$	*			*

We see, then, that USELISTED can enforce faithfulness to any arbitrary list of plurals in the memorized lexicon. Importantly, however, for nonce words (as in a wug task, or any generalization to novel data), the decision is handed down to markedness and faithfulness, and therefore the range of possible generalizations is the one schematized in (22). This effect is precisely what was observed with English speakers in Experiments 1 and 2: the lexicon contains an anti-universal

arrangement of existing nouns, with more alternations in monosyllables, but the nonce words exhibit a pattern that can be derived with the constraints in (22).

We reiterate that this result fundamentally depends on the limited ranking possibilities offered by the set of constraints in (22), and that the treatment of exceptions alongside that grammar need not be limited to the USELISTED approach. The analysis of lexical exceptions in terms of constraint cloning (Pater 2006, 2009; Coetzee 2008; Becker 2009; Becker et al. 2011), provides the same result. Existing items are listed with clones of the constraints in (22), with nonce forms again giving rise to one of the universally available generalizations.

We note that our analysis assumes an underlyingly voiceless fricative in the root /naf/. This is an aspect of what Hayes (1995, 1999) calls an “inside-out” analysis, where the UR is assumed to be identical to the base of the paradigm. This approach is also argued for in Becker (2009); Becker et al. (2011), and the same approach is used successfully outside Optimality Theory in Albright (2002, 2008). Alternatively, however, one can think of the faithfulness constraints in (23) and (24) as OO-Faith constraints (Benua 1997 et seq.), in which case it doesn’t matter what the laryngeal specification of the fricative is in the UR, since OO-Faith protects the unarguably voiceless fricative that surfaces in the singular form.

As for the exact nature of the markedness constraint that causes the alternation in English, we are not in a position to decide; we have studied the morpho-lexical distribution of the alternation, not its phonetic realization. Clearly, English does contrast two kinds of fricatives word-finally, as in [truθ] ‘truth’ vs. [smuð] ‘smooth’. Yet the transcription does not do justice to the realization of the contrast, which involves a short vowel and long fricative in ‘truth’, but a longer vowel and shorter, weaker, optionally voiceless fricative in ‘smooth’. Since it’s not clear that the process generates a genuinely voiced fricative, it is not clear that this relates to the debate about final voicing (Blevins 2004; Yu 2004; Kiparsky 2006; et seq.). We conjecture that the ‘voicing’ alternation is a case of morphologically restricted lenition, which historically was more widespread in English (see Honeybone & Spaargaren 2011 and references within). We note that connecting the alternation to the differences in duration heard on the fricative itself and on the preceding vowel, makes sense of the effect of vowel length: Leniting a fricative following a long vowel like [i] further lengthens that vowel and enhances its length. Leniting a fricative following a short vowel like [ɪ] lengthens the vowel, thus obscuring its identity as a short vowel.

To summarize, then, we have seen in this section how the *initial syllable faithfulness* approach accounts for the pattern of Experiments 1-3. The factorial typology generates languages with alternations that apply to monosyllables and polysyllables equally, or languages with alternations that apply to polysyllables only, but no languages with alternations applying to monosyllables only. While lexical exceptions can be found in a language like English, where the alternation applies to the existing listed monosyllables more than it does to existing listed polysyllables, this difference does not make it into a grammatical encoding that generalizes, as reflected in the speakers' treatment of nonce words.

## 8 Conclusions

We have shown that in the English lexicon, monosyllables are impacted by voicing alternations (as in *knife ~ knives*) more than polysyllables. In fact, the third top of the scale in our rating task was populated by monosyllables only. Iambs and trochees (defined here as polysyllables with a final stressed and unstressed syllable, respectively) were less acceptable with voicing alternations. In a nonce word task (wug-test), however, monosyllables and iambs were equally acceptable as voiced by English speakers. The means that a generalization that is true about the lexicon is not being projected onto nonce words; the generalization goes unlearned by the speakers ("the surfeit of the stimulus", Becker et al. 2011). Our results underline the necessity of complementing typological work with experimental work, since some languages may offer generalizations in the lexicon that don't form part of the speakers' grammar. Furthermore, we note that omnivorous statistical learners must be restrained by a set of built-in biases, or they will fail to learn the grammar that humans have. A pattern may be strongly present in the data that is available to the learner, and yet go unlearned.

We connected the English effect to a universal implicational relationship: Alternations in monosyllables imply alternations in polysyllables, but alternations on polysyllables allow monosyllables to remain faithful. This is seen in Turkish and in other languages, where alternations are preferred in polysyllables (§2). The grammar cannot prefer having alternations in monosyllables, and if there is such a pattern in the lexicon of a language, it cannot be learned. We derive this implicational relationship from *initial syllable faithfulness*, which states that the word-initial syllable is protected from alternations. Constraints that protect

*non*-initial syllables are excluded from CON, thus restricting the range of possible grammars. We note that while much recent literature explores the idea of inducing markedness constraints from the ambient language (e.g. Hayes & Wilson 2008; Boersma & Pater 2007; Moreton 2010), we know of no such proposals for learning faithfulness constraints. It would seem at the moment that faithfulness constraints, including their sensitivity to strong positions (onset, stressed syllable, initial syllable, etc.), are still a part of the what the learner brings to the table.

To show that even English speakers are biased to prefer the protection of monosyllables, we ran two artificial grammar experiments. In the first, we trained half of our participants on alternations in monosyllables, and half on alternations in polysyllables. While both groups applied the alternation cautiously to the shape they haven't seen, those trained on polysyllables applied the alternation significantly more cautiously to the monosyllables they haven't seen, relative to those trained on monosyllables. In other words, the English speakers deployed the universal implicational relationship, as seen in Turkish.

To strengthen the connection between the monosyllabicity effect and *initial syllable faithfulness*, we ran a second artificial grammar experiment, where all stems were disyllabic. We trained half of our participants on alternations in trochees, and half on iambs. Both groups of participants were cautious in applying the alternation to the position they haven't seen it in, but those trained in applying the alternation in iambs were significantly more cautious than those trained on trochees. It is clear, then, that the same kind of bias is seen in both experiments: Initial syllables are protected, and seeing alternations in non-initial syllables does not imply that the alternations will apply in initial syllables.

This paper's main focus is on local morphophonological alternations, which affect segments that are adjacent to the triggering affix, and which are expected to exhibit protection of initial syllables. In our artificial grammar experiments, we see the protection of initial syllables in both local and non-local alternations. Typologically, however, non-local alternations such as the German umlaut and the Chamorro umlaut may in fact show the opposite pattern, preferring to apply in initial syllables. If this is true, we expect to see attraction to initial syllable changes in irregular pasts in English, such as *feel* ~ *felt*. To our knowledge, however, wug testing of German umlaut, as in van de Vijver & Baer-Henney (2010), and wug testing of the English past, as in Prasada & Pinker (1993); Albright & Hayes (2003), have been limited to monosyllabic items only, leaving the prediction about the

treatment of polysyllables untested. There is more experimental work to be done, therefore, in exploring non-local alternations and their impact on initial vs. non-initial syllables.

Finally, we discussed the ultimate source of this protection of monosyllables, and generally of initial syllables. Our finding, that English speakers do not extend the pattern from their lexicon, and prefer the Turkish-like pattern in an artificial language, is inconsistent with Barnes's (2006) suggestion that initial syllable faithfulness is due to increased duration of initial syllables. As Barnes (2006) notes, initial syllables have longer duration in Turkish and shorter duration in English, and yet we find that speakers of both languages prefer the Turkish-like protection of initial syllables. It seems doubtful, then, that the protection of monosyllables is phonetically grounded. Deriving the protection of monosyllables from denser lexical neighborhood, as proposed in Ussishkin & Wedel (2009), has the advantage of predicting a uniformly cross-linguistic preference for less alternations in short words, since neighborhood density is always inversely correlated with length. However, the neighborhood density effect is shown to be too weak by Pycha et al. (2007); Becker & Nevins (2009). Furthermore, neighborhood density will not be able to distinguish local and non-local alternations, potentially leaving Chamorro umlaut unaccounted for. We conclude that at present, there is no known functional motivation for the protection of monosyllables. We attribute the protection of monosyllables to *initial syllable faithfulness* as an atomic element of the theory. We leave open the possibility that some functional motivation was phonologized (Hyman 1976, 2008) and incorporated into the grammatical apparatus.

The emerging picture is that *initial syllable faithfulness*, and perhaps positional faithfulness more generally, is a powerful organizing force in the regulation of morphophonological alternations on stems. It exerts an asymmetrical pressure to keep initial syllables intact, while allowing alternations in later syllables. Since we see its effect even in speakers who possess evidence to the contrary in their own language, we attribute this to an inherent bias, a part of Universal Grammar. We have shown that this bias causes a lexical generalization to go unlearned (a *surfeit* of the stimulus), and that it shapes the expectations that speakers have in artificial grammar experiments with partial training data (a *poverty* of the stimulus). Unlike those positional faithfulness constraints that are clearly phonetically based (faithfulness to onsets, stressed syllables), or positional faithfulness constraints that are sensitive to the morphology (faithfulness to roots,

nouns, Smith 2002, 2010), *initial syllable faithfulness* seems to be a purely formal and purely phonological element of the theory. While it may be grounded in phonetic and psycholinguistic pressures, its status is phonologized, as it exerts a formal guiding hand in hypothesis formation during acquisition, as illustrated in the Subset Principle schema; it asymmetrically biases learners against generalizing from non-initial to initial syllables, but not vice-versa.

## A Experiment 1 results by item

The items are listed with average responses (1=voiceless, 7=voiced).

monosyllables				iambs		trochees	
bath	4.6	nymph	2.1	aftermath	3.5	absinth	3.0
berth	3.4	oaf	2.0	behalf	4.6	azimuth	2.5
bluff	1.6	oath	4.6	belief	1.7	bailiff	2.7
booth	3.5	path	5.4	blacksmith	2.8	Behemoth	4.0
breath	1.6	prof	1.6	carafe	2.9	billionth	3.2
brief	1.2	proof	2.4	castoff	2.6	caliph	2.3
broth	3.5	puff	2.0	coelacanth	3.5	eightieth	3.2
calf	6.9	quaff	2.4	earmuff	1.9	goliath	2.1
chef	1.9	reef	2.8	eighteenth	2.3	hieroglyph	2.0
chief	3.4	ref	1.1	epitaph	3.2	hyacinth	3.0
clef	2.1	roof	3.4	giraffe	3.6	kerchief	3.9
cliff	2.0	safe	2.4	handcuff	2.5	mammoth	3.2
cloth	4.0	scarf	6.9	midriff	2.0	mastiff	2.8
cough	1.7	self	6.7	motif	2.9	monolith	2.7
cuff	1.4	serf	1.3	paragraph	2.5	plaintiff	2.9
death	1.7	sheaf	5.8	pilaf	2.9	pontiff	2.5
dwarf	6.3	sheath	5.2	polymath	3.4	Sabbath	3.7
earth	3.3	shelf	6.9	psychopath	4.0	seraph	3.2
elf	7.0	skiff	2.3	relief	3.9	serif	2.6
faith	2.8	sleuth	3.9	sabertooth	3.2	sheriff	2.1
fife	2.4	sloth	2.7	vermouth	4.7	tariff	2.2
fourth	3.2	sniff	1.5			triumph	2.2
gaffe	2.1	spooft	2.6			zenith	3.6
goof	2.0	staff	2.1				
Goth	3.6	stiff	1.0				
graph	1.6	strength	1.4				
growth	1.8	swath	4.6				
gulf	2.8	thief	6.6				
half	6.8	tiff	2.7				
hearth	3.2	trough	3.4				
heath	2.4	truth	4.2				
hoof	6.8	turf	2.5				
knife	6.0	waif	2.5				
laugh	2.3	wharf	5.1				
leaf	6.6	whiff	2.5				
life	7.0	width	3.7				
loaf	6.3	wife	6.4				
month	1.7	wolf	6.6				
moth	4.8	wraith	4.7				
mouth	4.6	wreath	4.6				
myth	2.2	youth	4.0				

## B Experiment 2 results by item

The items are listed with average responses (1=voiceless, 7=voiced).

monosyllables		iamb		trochees	
berf	5.7	bəwælf	4.9	brʌnləf	3.9
bɪɑf	5.5	glənəf	4.6	ɔʒənəf	3.9
bʌnf	4.5	həbaf	4.8	floʊdəf	4.5
dif	6.0	jəstəf	5.1	glɔsəf	4.3
duf	3.5	kəzɪf	3.3	junəf	4.3
ɔʒəf	3.8	kwəzʌf	4.1	jælməf	3.6
glif	5.2	mələif	5.5	karəf	3.4
jef	4.2	nədlɪf	4.6	kreɪdəf	5.0
kəf	3.5	nəʃɪf	5.8	nelgəf	4.3
kləf	5.1	nətuəf	5.2	pidəf	4.9
meɪf	5.6	pətelf	5.6	rɪlkəf	3.8
nɪf	3.7	pə'næf	5.1	ɪələf	4.3
noʊf	4.9	ɪədɪf	5.2	seɪləf	4.2
ɪoʊf	5.2	səbaf	5.3	stə'nəf	4.5
senf	4.4	sə'klouf	5.8	stəkəf	4.5
sməf	5.0	stə'vəf	5.1	ʃədəf	3.8
stəf	5.0	ʃənəf	4.5	ʃestəf	5.7
sʌf	3.4	təŋkɪf	5.2	ʃʌtəf	3.6
ʃɪlf	5.2	tɪəluf	5.1	təkəf	4.6
waf	3.0	wə'gef	4.3	ʃɪdəf	3.9
zælf	5.1	wə'ɪɪf	5.4	voʊləf	4.4
zɑf	5.1	zəkɑf	4.8	wɪbəf	4.0
biθ	5.0	brələʊθ	5.1	ɔʒəkəθ	4.2
bɪlθ	4.2	dəkɛlθ	4.6	fɪdəθ	2.6
doʊθ	5.5	dəpaɪθ	4.4	haɪləθ	3.8
dwiθ	5.3	ɔʒəzəθ	3.4	həkəθ	2.9
huθ	5.3	gənəθ	4.1	kælnəθ	3.2
kɪrəθ	4.1	hələɪθ	5.1	lə'bəθ	3.6
kʌθ	3.9	həriθ	4.5	lɪsəθ	3.9
kwəθ	4.0	həsɑɪθ	4.2	mouʊdəθ	3.2
lɪθ	3.7	kənəθ	3.8	nækəθ	3.5
pælθ	4.9	kənuθ	4.6	neɪgəθ	3.3
pɑɪθ	4.0	nəboʊθ	4.5	paɪdəθ	2.5
peɪθ	4.4	rəstɪlθ	5.3	roʊsəθ	4.3
pɛlθ	3.8	ɪəstɛθ	4.3	ruvəθ	3.0
prəθ	4.1	səkəθ	4.3	ɪkrəθ	3.9
skɛθ	4.5	səpʌlθ	4.8	səstəθ	4.2
smɛɪθ	4.8	səsælθ	5.2	sɛŋgrɪθ	3.4
stəθ	3.2	səstʌθ	3.6	sɪltəθ	3.6
sʌlθ	3.6	ʃədəθ	3.9	speɪləθ	3.4
ʃɪrɪθ	4.9	ʃəkəθ	5.0	ʃənəθ	3.6
ʃɑrɪθ	5.2	təɪɪθ	4.0	tʌlpəθ	3.6
ʃoʊθ	4.9	təspɪθ	3.5	wə'gəθ	2.3
wəθ	4.2	zəziθ	3.9	wəkəθ	3.6

## C Experiment 3 results by item

The items are listed with the average responses they received in each group (0=voiceless, 1=voiced).

	mono training	iamb training		mono training	iamb training
zap	0.60	0.46	dərap	0.60	0.57
flep	0.72	0.42	həsap	0.76	0.67
sep	0.47	0.44	təgep	0.52	0.65
mip	0.67	0.44	kənip	0.42	0.50
bop	0.74	0.43	rəlip	0.47	0.62
ʃop	0.58	0.63	fəʃop	0.64	0.67
nup	0.67	0.36	gətɔp	0.65	0.59
ʃup	0.70	0.76	ɔʒəlup	0.65	0.71
brap	0.64	0.42	sərup	0.61	0.62
gap	0.80	0.56	ʃənarp	0.69	0.53
grat	0.65	0.57	frədət	0.59	0.55
dret	0.71	0.47	nəmat	0.67	0.60
klet	0.36	0.43	zənət	0.50	0.47
vit	0.65	0.45	bəgit	0.35	0.60
zot	0.62	0.31	səplit	0.56	0.31
zot	0.75	0.55	ʃəlɔt	0.50	0.55
prut	0.50	0.38	ʃədɔt	0.64	0.50
stut	0.53	0.53	gəʃut	0.78	0.44
part	0.82	0.42	kləput	0.48	0.47
skart	0.46	0.21	fəhart	0.50	0.50
strak	0.64	0.47	gləzak	0.33	0.46
glek	0.65	0.46	ʃəpak	0.31	0.55
ʃek	0.44	0.47	pənek	0.92	0.57
nik	0.46	0.47	bəsik	0.44	0.56
dok	0.45	0.25	gəfrok	0.50	0.62
prok	0.53	0.41	ləʃok	0.44	0.67
fuk	0.48	0.40	bəʃuk	0.57	0.23
huk	0.65	0.27	zətuk	0.47	0.50
blak	0.55	0.46	rədərk	0.73	0.53
ʃark	0.45	0.38	səpark	0.69	0.62

## D Experiment 4 results by item

The items are listed with the average responses they received in each group (0=non-alternating vowel, 1=umlauted).

	trochee training	iamb training		trochee training	iamb training
bəchuk	0.17	0.21	blukəm	0.20	0.06
chəmum	0.20	0.52	bulək	0.38	0.05
drətup	0.39	0.67	funəl	0.46	0.15
frəkuf	0.23	0.28	gunək	0.46	0.00
gəsut	0.32	0.47	kushət	0.35	0.00
kənul	0.33	0.45	lupət	0.38	0.08
kləput	0.32	0.57	mustəp	0.53	0.11
pədul	0.50	0.48	prusəm	0.46	0.05
pləsüm	0.17	0.40	shundəf	0.56	0.14
səfup	0.11	0.67	suməl	0.53	0.12
shrətuk	0.32	0.62	sutrəf	0.47	0.10
təshuf	0.35	0.50	zuməp	0.52	0.06
bəndef	0.38	0.32	brezəl	0.44	0.20
kəshel	0.13	0.33	chedrəl	0.25	0.00
kəzem	0.65	0.36	deshəf	0.44	0.10
məpek	0.16	0.53	frenəp	0.42	0.08
nədəp	0.23	0.36	kefət	0.39	0.27
pəsnem	0.18	0.47	krefəm	0.59	0.04
pləzek	0.29	0.50	kreshək	0.48	0.05
skənet	0.14	0.68	lefəp	0.42	0.00
sməref	0.25	0.56	prekət	0.35	0.17
stəfet	0.07	0.62	sefək	0.32	0.15
təgep	0.16	0.73	shebəf	0.43	0.05
trəmel	0.23	0.44	zedrəm	0.43	0.00

## References

- Albright, Adam (2002). *The Identification of Bases in Morphological Paradigms*. Ph.D. dissertation, UCLA.
- Albright, Adam (2008). A Restricted Model of UR Discovery: Evidence from Lakhota. Ms. MIT.
- 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 computational experimental study. *Cognition* 90. 119–161.
- Anttila, Arto (2003). Finnish Assibilation. In Makoto Kadowaki & Shigetō Kawahara (eds.) *The Proceedings of NELS 33*, Amherst, MA: GLSA. 13–24.
- Anttila, Arto (2006). Variation and opacity in Finnish phonology. In Juha Fagerholm, Leila Puska, Jan Åström, Johanna Blomqvist, Pirjo-Leena Forsström, Leena Jukka, Sirpa Kotila, M. Minna Laine, Manne Miettinen, Jura Tarus & Satu Tissari (eds.) *CSC Report on Scientific Computing in Finland 2004-2005*. 171–174.
- Barnes, Jonathan (2006). *Strength and Weakness at the Interface: Positional Neutralization in Phonetics and Phonology*. Berlin/New York: Mouton de Gruyter.
- Bates, Douglas & Martin Maechler (2009). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-32.
- Becker, Michael (2009). *Phonological Trends in the Lexicon: The Role of Constraints*. Ph.D. dissertation, University of Massachusetts Amherst.
- Becker, Michael, Lauren Eby Clemens & Andrew Nevins (in preparation). The monosyllabicity effect in French and Brazilian Portuguese. Ms. University of Massachusetts Amherst.
- Becker, Michael, Nihan Ketrez & Andrew Nevins (2011). The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87. 84–125.
- 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.
- Benua, Laura (1997). *Transderivational Identity: Phonological Relations Between Words*. Ph.D. dissertation, University of Massachusetts Amherst.

- Berko, Jean (1958). The child's learning of English morphology. *Word* 14. 150–177.
- Berwick, Robert C. (1985). *The Acquisition of Syntactic Knowledge*. MIT Press.
- Blevins, Juliette (2004). *Evolutionary phonology: the emergence of sound patterns*. Cambridge: Cambridge University Press.
- Boersma, Paul & Joe Pater (2007). Constructing constraints from language data: The case of Canadian English diphthongs. Talk given at NELS38, University of Ottawa.
- Boersma, Paul & David Weenink (2008). Praat: Doing phonetics by computer (version 5.0.24). [Computer program].
- Bybee, Joan (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* 10. 425–455.
- Cabré, Teresa & Pilar Prieto (2006). Exceptional hiatuses in Spanish. In S. Colina & F. Martínez-Gil (eds.) *Optimality-Theoretic Advances in Spanish Phonology*, John Benjamins. 205–238.
- Casali, Roderic (1998). *Resolving Hiatus*. Garland, New York.
- Christdas, Prathima (1988). *The Phonology and Morphology of Tamil*. Ph.D. dissertation, Cornell University.
- Chung, Sandra (1983). Transderivational relationships in Chamorro phonology. *Language* 59. 35–66.
- Coetzee, Andries W. (2008). Grammaticality and ungrammaticality in phonology. *Language* 84. 218–257.
- Côté, Marie-Hélène (2011). French liaison. In Colin Ewen, Beth Hume, Marc van Oostendorp & Keren Rice (eds.) *The Blackwell Companion to Phonology*, Blackwell.
- Ernestus, Miriam & R. Harald Baayen (2003). Predicting the Unpredictable: Interpreting Neutralized Segments in Dutch. *Language* 79. 5–38.
- Fanselow, Gisbert & Caroline Féry (2002). Ineffability in grammar. In Gisbert Fanselow & Caroline Féry (eds.) *Resolving Conflicts in Grammars: Optimality Theory in Syntax, Morphology, and Phonology*, Helmut Buske Verlag, vol. 11 of *Linguistische Berichte*. 265–304.
- Giavazzi, Maria (2010). *The phonetics of metrical prominence and its consequences for segmental phonology*. Ph.D. dissertation, MIT.
- Gomes, Christina Abreu & Carolina Gonçalves Manoel (2010). Flexão de número na gramática de criança e na gramática do adulto. *Veredas*. 122–134.
- 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 Yidj 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 (2009). *Introductory Phonology*. Wiley-Blackwell.
- 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.
- Honeybone, Patrick & Marleen Spaargaren (2011). Not another case of final obstruent voicing? laryngeal realism, late middle english and impossible phonological change. Handout for talk at OCP 8, Marrakesh.
- Hyman, Larry (1976). Phonologization. In Alphonse Juilland (ed.) *Linguistic studies presented to Joseph H. Greenberg*, Saratoga, CA: Anna Libri. 407–418.
- Hyman, Larry (2008). Enlarging the scope of phonologization. In *UC Berkeley Phonology Lab Annual Report*, UC Berkeley. 382–409.
- 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.
- Jesney, Karen (2009). Positional faithfulness, non-locality, and the harmonic serialism solution. Proceedings of the 39th Meeting of the North East Linguistic Society.
- Jespersen, Otto (1909). *A Modern English Grammar on Historical Principles Part I: Sounds and Spellings*. Carl Winter, Heidelberg.
- 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.
- Kiparsky, Paul (1982). Word-formation and the lexicon. In F. Ingemann (ed.) *Proceedings of the Mid-America Linguistics Conference*.

- Kiparsky, Paul (2006). The amphichronic program vs. evolutionary phonology. *Theoretical Linguistics* 32. 217–236.
- Lees, Robert (1961). *The Phonology of Modern Standard Turkish*. Bloomington: Indiana University Press.
- Lisker, Leigh (1974). On “explaining” vowel duration variation. *Glossa* 8. 233–246.
- Mallet, Géraldine (2008). *La liaison en français: descriptions et analyses dans le corpus PFC*. Ph.D. dissertation, Université Paris Ouest, Nanterre.
- Manzini, Rita & Kenneth Wexler (1987). Parameters, Binding Theory, and Learnability. *Linguistic Inquiry* 18. 413–444.
- Moreton, Elliott (2004). Non-computable functions in optimality theory. In John J. McCarthy (ed.) *Optimality Theory in Phonology*, Blackwell. 141–163.
- Moreton, Elliott (2010). Connecting paradigmatic and syntagmatic simplicity bias in phonotactic learning. Talk given at MIT, April 9.
- Moreton, Elliott & Joe Pater (2011). Learning artificial phonology: A review. Ms. UNC Chapel Hill and UMass Amherst.
- 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.
- 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.
- Prasada, Sandeep & Steven Pinker (1993). Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes*. 1–56.
- 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.
- R Development Core Team (2007). R: A language and environment for statistical computing. Vienna, Austria. (<http://www.R-project.org>).
- Selkirk, Elisabeth O. (1982). *The Syntax of Words*. MIT Press.
- Smith, Jennifer L. (2000). Positional faithfulness and learnability in optimality theory. In Rebecca Daly & Anastasia Riehl (eds.) *Proceedings of ESCOL 99*, Ithaca: CLC Publications. 203–214.
- Smith, Jennifer L. (2002). *Phonological Augmentation in Prominent Positions*. Ph.D.

- dissertation, University of Massachusetts Amherst.
- Smith, Jennifer L. (2010). Category-specific effects. In Colin Ewen, Beth Hume, Marc van Oostendorp & Keren Rice (eds.) *The Blackwell Companion to Phonology*, Blackwell.
- Sprouse, Jon (2010). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* .
- Stausland Johnsen, Sverre (to appear). Neighborhood density in phonological alternations. In *Proceedings of the 36th annual meeting of the Berkeley Linguistics Society*.
- Steriade, Donca (1994). Positional neutralization. Ms., UCLA.
- Steriade, Donca (1999). Lexical conservatism in french adjectival liaison. In Jean-Marc Authier, Barbara E. Bullock & Lisa A. Reed (eds.) *Formal Perspectives in Romance Linguistics*, John Benjamins. 243–270.
- Tessier, Anne-Michelle (2007). *Biases and stages in phonological acquisition*. Ph.D. dissertation, University of Massachusetts Amherst, Amherst, MA.
- Toporišič, Jože (2000). *Slovenska slovnica*. Maribor: Obzorja.
- Topping, Donald (1968). Chamorro vowel harmony. *Oceanic Linguistics* 7. 67–79.
- Trubetzkoy, Nikolai (1939). *Grundzüge der Phonologie*. Göttingen: Vandenhoeck & Ruprecht.
- Ussishkin, Adam & Andrew Wedel (2009). Lexical access, effective contrast and patterns in the lexicon. In Paul Boersma & Silke Hamann (eds.) *Perception in Phonology*, Mouton de Gruyter. 267–292.
- van de Vijver, Ruben & Dinah Baer-Henney (2010). Acquisition of alternations: The role of substance, frequency and phonotactics in the acquisition of german voicing and vowel alternations. Ms. University of Potsdam.
- Wiese, Richard (1996). Phonological versus morphological rules: on German umlaut and ablaut. *Journal of Linguistics* 32. 113–135.
- Wilson, Colin (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science* 30. 945–982.
- Yu, Alan (2004). Explaining final obstruent voicing in Lezgian: Phonetics and history. *Language* 80.1. 73–97.
- Zuraw, Kie (2000). *Patterned Exceptions in Phonology*. Ph.D. dissertation, UCLA.