July 2014. To appear in *Phonology*

# Domain generalization in artificial language learning

Scott Myers (UT Austin) & Jaye Padgett (UC Santa Cruz)

## Abstract

Many languages have restrictions on word-final segments, such as a requirement that any word-final obstruent be voiceless. There is a phonetic basis for such restrictions at the ends of *utterances*, but not the ends of *words*. Historical linguists have long noted this mismatch, and have attributed it to an analogical generalization of such restrictions from utterance-final to word-final position. To test whether language learners actually generalize in this way, two artificial language learning experiments were conducted. Participants heard nonsense sentences in which there was a restriction on utterance-final obstruents, but in which no information was available about word-final, utterance-medial obstruents. They were then tested on utterances that included obstruents in both positions. They learned the pattern and generalized it to word-final utterance-medial position, confirming that learners are biased toward word-based distributional patterns.

## 1 Introduction: word-final phonology and utterance-final phonetics

Phonological patterns often make reference to word edges. For example, both voiced and voiceless obstruents occur in Russian, but at the end of a word only the voiceless ones occur, as illustrated by the alternations in (1), where the relevant consonants are bolded (Padgett 2012 and references therein).

(1) Russian word-final devoicing

| | | | | |
|---|---|---|---|---|
| slʲe**t** | ‘track (nom.sg.)’ | cf. | slʲiˈ**d**a | ‘track (gen. sg.)’ |
| knʲi**k** | ‘book (gen.pl.)’ | cf. | ˈknʲi**g**ə | ‘book (nom. sg.)’ |

Final devoicing is a wide-spread phonological pattern, variants of which occur in German, Dutch, Russian, Sanskrit, Walloon, Turkish, Hausa, and other languages (Blevins 2006; Myers 2012).[1]

Scholars have long related this restriction on voicing categories to the breakdown in voicing that occurs at the end of an utterance (Sievers 1901: 289-290; Jespersen 1926: 101; Bloomfield 1933: 373; Lightner 1972: 332-333; Lindblom 1983: 237; Ingram 1989: 35). The vocal folds in non-speech breathing are spread wide apart to facilitate air passage, and speakers begin spreading the vocal folds in anticipation of this posture as they approach the end of the utterance (Sweet 1877: 65; Lisker et al. 1969: 1545; Klatt & Klatt 1990; Shadle 1997: 42; Jessen 1998; Slifka 2006). At

[1] The arguments in this paper would apply equally to ‘final fortition’ (Jessen & Ringen 2002; Iverson & Salmons 2007), in which voiceless aspirated and unaspirated stops contrast, but only the aspirated ones occur in final position. For accounts of incomplete neutralization of laryngeal contrasts in final position (Port & O'Dell 1985; Warner et al. 2004), see Ernestus and Baayen (2007) and Gafos (2006).

the same time, the subglottal pressure that is necessary to drive vocal fold vibration declines over the course of the utterance and hits its low point at the end (Westbury & Keating 1986: 156). The result is a deterioration in voicing as the speaker approaches pause, often passing through nonmodal voicing before a final voiceless interval. Such utterance-final devoicing has been found in instrumental acoustic studies in English (Haggard 1978; Docherty 1992; Smith 1997), French (Smith 1999, 2003), Finnish (Lehtonen 1970: 45; Myers & Hansen 2007), Kinyarwanda (Myers 2005), and has been noted as well in many transcription-based studies (Hansen & Hansen 1969: 162-163).

Utterance-final devoicing affects listeners' identification of voicing categories. Myers (2012) had English speakers identify words contrasting in the voicing of a final fricative (e.g. *leaf/leave*), which had been excised from final or medial position in recorded utterances. The participants were significantly more likely to misidentify the voiced-final (e.g. *leave*) as the voiceless-final word (e.g. *leaf*) when that word had been drawn from utterance-final position than when it had been excised from utterance-medial position. This suggests that the coarticulatory effects of utterance-final position are great enough to influence voicing identification even for experienced speakers.

Such a tendency to misidentify final voiced obstruents as voiceless could lead a language learner to conclude that utterance-final obstruents are all voiceless. In making such a generalization, the learner would be adopting a pattern of phonological final devoicing, reanalyzing a gradient pattern in phonetic realization as a phonological restriction on speech categories (Hyman 1976; Hyman 2013). This occurs if the language learner fails to learn to compensate in perception for the effects of coarticulation, interpreting the partially devoiced final variants of voiced obstruents as intended voiceless obstruents (Ohala 1993).

This account of the diachronic origins of phonological final devoicing provides an explanation of some basic properties of the pattern. Voiced obstruents are subject to change in pre-pausal position because they are less voiced in that position than in phrase-medial position, due to anticipation of the laryngeal posture during pause. Voiced obstruents change to voiceless obstruents there, because that is what a partially devoiced obstruent tends to be mistaken for. The pattern is restricted to obstruents, because a devoiced sonorant is so low in intensity that it tends to be mistaken for silence rather than for a voiceless segment (Myers & Hansen 2007).

There is, however, a basic challenge for such a diachronic scenario. The phonetic devoicing effect is limited to utterance-final pre-pausal position, while the phonological pattern is most often a restriction on word-final position. Consider the contexts exemplified in (2).

(2) Phrasal contexts of Russian word-final devoicing[2]

| | | |
|---|---|---|
| a. | dub-a | 'oak tree (gen.)' |
| b. | dup | 'oak tree (nom.)' |
| c. | dup kustarnʲikovij | 'coastal scrub oak' |
| d. | dup bʲelij | 'white oak' |
| e. | dup mʲullʲera | 'muller oak' |
| f. | dup arʲizonskʲij | 'Arizona white oak' |

The word-final /b/ is realized as voiceless [p] in phrase-medial position as well as phrase-final position, though only the latter is phonetically a context for devoicing.

A standard historical account of this mismatch has been to posit a process of analogical generalization from utterance-final to word-final position (Ewert 1943: 75; Wackernagel 1957: 301-309; Vennemann 1974; Hyman 1978b; Westbury & Keating 1986: 161; Hock 1991: 239). Any obstruent in utterance-final position is necessarily also at the end of a word, so extension of the pattern to word-final position could be understood as a generalization from one subset of words (the utterance-final ones) to all words. (See discussion in conclusion.) The process can be seen underway in Polish, where some dialects maintain utterance-final devoicing and others have innovated word-final devoicing (Jassem & Richter 1989: 317). We refer to this diachronic process as DOMAIN GENERALIZATION: a shift in the domain of a distribution pattern from a longer, more inclusive prosodic domain (e.g. the utterance) to a shorter, less inclusive one (e.g. the word). Hyman (1978b: 452-3) refers to this process as BOUNDARY NARROWING, since the boundaries referred to in word-final devoicing are closer together than the utterance boundaries that the phonetic pattern refers to.

Domain generalization plays a role not only in final devoicing, but also in the evolution of other restrictions on word-final elements that have their phonetic origin in utterance-final position. For example, high tone is in many languages avoided in word-final position, either through lowering of final high tones or retraction of such tones to nonfinal position (Hyman 1978a: 265; Myers 1999: 216). This seems clearly related to phrase-final f0 lowering, which is the tendency for high tone targets at the end of a phrase to have lower f0 values than comparable tones in non-final position (Liberman & Pierrehumbert 1984; Pierrehumbert & Beckman 1988; Herman 1996). The avoidance of final high f0 peaks is also reflected in the tendency for a high tone in phrase-final position to occur earlier in the syllable than in phrase-medial position (Silverman & Pierrehumbert 1990; Prieto et al. 1995; Myers 1999). Becker (1977; 1979) argues for domain generalization as a historical process in the history of final tone retraction in Serbo-Croatian, with different dialects representing different steps in the progression of domain generalization.

Myers and Hansen (2007) discuss a widely distributed phonological pattern in which a word-final vowel must be short (e.g. Kinyarwanda). They relate this to utterance-final phonetic devoicing, since devoicing of the final portion of a vowel greatly reduces the intensity of that portion. They provide evidence that this can affect vowel length perception in two experiments

[2] Some sources describe voicing assimilation in obstruent sequences as in (2)b, i.e. **dub bʲelij**. See Padgett (2012) for discussion.

in which Finnish speakers identified partially-devoiced long vowels as short, with the likelihood of a long vowel identification depending on the duration of the voiced portion of the vowel. This tendency could explain why utterance-final long vowels might be reanalyzed diachronically as short, since it is those vowels that are subject to partial devoicing, but the extension of the pattern to word-final position requires a process of domain generalization.

Hualde (2013) shows that domain generalization can also have the effect of *restricting* application of a process. He argues that in Western Romance intervocalic lenition (e.g. spirantization of voiced stops and voicing of voiceless obstruents) was an optional gradient phonetic process that applied freely across word boundaries, as in V#CV. When the process was phonologized, however, as in Istanbul Judeo-Spanish, with the lenited form becoming the phonological target, the domain of the phonological process became the word. Where **b* was between two vowels within a word, it changed to *v*, as in *saber* > *saver* 'to know', while word-initial **b* was preserved even when intervocalic, as in *la bóka* 'the mouth', due to the influence of the word in non-leniting contexts (e.g. utterance-initial *bóka* 'mouth').

The mismatch between utterance-final devoicing (the phonetically motivated pattern) and word-final devoicing (the most commonly observed phonological pattern) also presents a challenge to phonological accounts of devoicing that appeal to phonetic bases. According to Steriade's (1997) licensing by cue hypothesis, for example, a contrast is favored where the cues to that contrast are richly present. The ideal position for a voicing contrast in obstruents is before a sonorant, because the cues to the contrast are better in that position, and final devoicing results when the constraints that enforce that restriction outrank general laryngeal faithfulness.[3]

The challenge for such an account is the same as for the diachronic account. In forms such as (2e-f) there is devoicing of word-final obstruents in Russian although the following sound is a sonorant. Steriade (1997:55-8) suggests a solution via 'paradigm uniformity'. According to this approach, /dub/ is realized as [dup] in [dup ar$^{j}$izonsk$^{j}$ij] (and in other phrasal contexts) because it should be identical to [dup] in 'citation form' or utterance-final position. Output-output faithfulness (Benua 1995) would be one way to formalize this idea (see also Steriade 2000; McCarthy 2005). However, such an appeal to the influence of related utterance-final or citation forms cannot be the whole story behind domain generalization, since speakers of languages like Russian apply word-final devoicing to forms they have never encountered. Our approach shares with Steriade's the assumption that word-final devoicing is phonetically motivated in utterance-final position, not word-final position, and that it extends to all words by a process of generalization. In the conclusion we consider how this domain generalization might work.

Domain generalization is a plausible way of relating restrictions on corresponding positions in different domains (cf. Wackernagel 1957: 301-309; Chafe 1959: 486; Flack 2009). But while it seems plausible that language learners might generalize in this way, there is no direct empirical evidence that they actually do so. In contrast to the extensive experimental literature on the phonetic conditions in utterance-final position, and the extensive documentation of word-final phonological patterns in the phonological literature, the discussion of an analogical

[3] The important distinction between pre-sonorant position and elsewhere was anticipated by Lombardi (1995), but without the emphasis on the cue-based underpinnings.

generalization connecting the two patterns remains purely speculative. The present study is designed to provide evidence on this point.

## 2. Experiments

### 2.1 Artificial language learning experiments

Since our goal is to investigate how language learners generalize about distributional restrictions from a speech sample, we would like an experimental set-up that allows us to control the structure of the language material that the participant is exposed to and to gauge learning that occurs. In an *artificial language learning* (also known as artificial grammar learning or miniature grammar learning) paradigm, the participant is exposed to expressions of a constructed language during a familiarization or learning phase. Then the knowledge she has acquired through familiarization is measured in a test phase (e.g. by having her judge whether a novel item belongs to the language she has just been exposed to).

Such studies have been used to investigate what kinds of syntactic patterns learners can acquire (Miller 1958; Braine 1963; Reber 1967; Morgan et al. 1987; Pothos & Bailey 2000), as well as what kinds of phonological patterns learners acquire most easily (Schane et al. 1975; Gómez 2002; Onishi et al. 2002; Pycha et al. 2003; Wilson 2003; Seidl & Buckley 2005; Peperkamp et al. 2006; Pycha et al. 2006; Tessier 2006; Wilson 2006; Moreton 2008; Coetzee 2009; Kapatsinski 2009; Carpenter 2010). An insightful review and synthesis of artificial language learning work in phonology is provided by Moreton and Pater (2012a, b).

The particular question at issue in our study is whether language learners generalize from the more restricted case of utterance-final consonants to the more general case of word-final consonants. The *poverty of stimulus* method in artificial language learning experiments provides a way to investigate how learners generalize from a limited speech sample (Wilson 2003, 2006; Finley & Badecker 2009; White 2013). In a poverty of stimulus learning experiment, there are classes of items included in the test phase of the experiment that were not included in the learning stage, but are related to items that were included there. Finley and Badecker , for example, exposed participants to pseudo-words with an alternating final syllable which was *mi* when the preceding vowels were front, and *mu* when the preceding vowels were back (a back harmony pattern). In one condition, called *Mid Hold-Out*, the conditioning vowels in the learning phase included low and high vowels but not mid vowels. In the *Low Hold-Out* condition, on the other hand, the conditioning vowels included mid and high vowels but no low vowels. In the test phase, all three vowel heights were included as conditioning vowels. The researchers found that participants generalized the harmonic pattern to mid vowels in the Mid Hold-Out condition, but not to low vowels in the Low Hold-Out condition. Participants assumed that a generalization that held of low and high vowels would hold of all vowels, while a generalization that held of mid and high vowels could be limited to just the nonlow vowels.

The poverty of stimulus method is an effective way of testing how learners generalize from incomplete data, and under what conditions they are willing to go beyond the data they have. Learners of a language with prepausal phonetic devoicing have evidence about voicing in prepausal position, but the claim is that they generalize beyond that evidence when they extend the voicing restriction to the ends of words. In the experiments to be reported here, participants

were exposed to a familiarization sample in which obstruents contrast in voicing except in utterance-final position, and there are no instances of word-final obstruents except in utterance-final position. They were asked to judge whether novel test items belonged to the same language as they heard in the familiarization phase. The items in the test set included final voiced and voiceless obstruents, in both utterance-final and nonfinal position. The hypothesis is that participants will generalize the pattern that they learned on the basis of utterance-final words to new words that are utterance-medial.

## 2.2 Experiment 1

### 2.2.1 Methods

#### 2.2.1.1 Participants

48 undergraduate students at the University of California, Santa Cruz participated in this experiment. They were all taking an introduction to linguistics course, and they received course credit for participating.

All participants were native speakers of English, with the latest initial exposure to the language being at 7 years. 23 of the participants had also spoken another language besides English as children, but these were all languages without a final voicing neutralization pattern (e.g. Spanish, Hebrew). Data from 3 participants was excluded from the analysis because those 3 were speakers of a language with final voicing neutralization: Russian, Thai, and German.

2 of the participants turned out to have taken a course in phonology which covered phonological patterns such as final devoicing. Their data was also excluded from the analysis.

#### 2.2.1.2 Materials

The stimuli for the study were nonsense sentences consisting of CV(C) syllables, in which C belonged to the set [p, t, s, z, m, n] and V belonged to the set [i, e, a, o, u]. Each sentence had one of the two frames in (3).

(3) a. Utterance-final: ˈsanta ___.
b. Utterance-medial: ˈsanta ___ ˈmizupu.

The blank in (3) was the position of the target word, which was the only item that varied in the sentences. The target words were nonce words of the form ˈCVC(V(C(V))), where codas occurred only at the end of the target word, and were restricted to the set [s, z, m, n]. The target words for the learning stage of the experiment are given in Table 1a, and those for the testing stage are given in Table 1b.

| *s-final* | *z-final* | *nasal-final* | *vowel-final* |
|---|---|---|---|
| pis | puz | pum | pamu |
| tes | tiz | ton | teno |
| sas | sez | zan | zipa |
| zos | zaz | sem | sute |
| mus | moz | mim | mozi |
| nis | nuz | nun | nasu |
| mapes | nasiz | zipon | nipeno |
| notas | mezez | setam | metima |
| puzos | pitaz | mazen | sazupe |
| tisus | tupoz | nosim | zosoti |
| semis | somuz | pumun | pumasu |
| zanes | zaniz | tinom | tinezo |

**Table 1a: Target words for the learning stage**

| *s-final* | *z-final* | *nasal-final* | *vowel-final* |
|---|---|---|---|
| ses | pez | pan | piza |
| zas | taz | tem | tuse |
| pus | zoz | zin | somi |
| tis | suz | sum | zanu |
| mes | miz | mon | nepo |
| nas | nez | nam | mita |
| zopus | pezaz | tepen | mopuse |
| sutis | tisoz | patim | natizi |
| pizes | mupuz | zosun | zesenu |
| tesas | notiz | suzom | sizamo |
| mamos | sanez | niman | punota |
| nonus | zemaz | menam | tomupe |
| pas | paz | pim | pupe |
| tus | tuz | tan | titu |
| sos | soz | som | semo |
| zes | zez | zun | zana |
| mis | mez | men | moze |
| nos | naz | nom | nusu |

**Table 1b: Target words for the testing stage**

Each of the consonants has the same frequency of occurrence in nonfinal position in each of the two target word inventories in Tables 1a and 1b. Each inventory also included equal numbers of items ending in the four different classes of final segment: [s], [z], a nasal, or a vowel.

Each target word in Tables 1a and 1b was recorded in each of the sentence frames in (3) by one of the authors, with a declarative intonation and no pauses between words. The pronunciation

was as in Western varieties of American English, except for the presence of the low central vowel [a]. Stress was located on the first syllable of each nonsense word.

Two learning sets were created from the recordings of the sentences with the items in Table 1a. Both sets included all the nasal-final and vowel-final items in Table 1a in both the final and the nonfinal sentence frames. The **final devoicing** learning set also included all the sentences in which an s-final word occurs in the sentence-final frame in (3a). This learning set therefore had obstruents only in syllable-onset position, except for 12 sentences with [s] in utterance-final position. Thus the final devoicing set included s-final sentences such as *ˈsanta ˈpis*, but no z-final sentences like *ˈsanta ˈpuz* or sentences with medial obstruent codas such as *ˈsanta ˈpis ˈmizupu*. In this set of sentences, both [s] and [z] occurred prevocalically, but [s] was the only obstruent that occurs in utterance-final position.

The **final voicing** learning set was the same, except each s-final sentence in the final devoicing set was replaced with a z-final sentence, consisting of one of the z-final words in Table 1a in the sentence-final frame in (3a). This set thus included z-final sentences such as *ˈsanta ˈpuz***,** but no s-final sentences such as *ˈsanta ˈpis* or sentences with medial obstruent codas. In this set of sentences, both [s] and [z] occurred prevocalically, but [z] was the only obstruent that occurred in utterance-final position.

Each learning set consisted of 60 sentences, with 12 sentences in each of 5 conditions (sonorant-, vowel- and obstruent-final words in the sentence final frame (3a), and sonorant- and vowel-final words in the nonfinal frame in (3b)).

There was one testing set, consisting of 144 sentences: the 72 items in Table 1b in each of the 2 sentence frames in (3).

#### 2.2.1.3 Procedure

The materials were presented to participants through headphones in a sound-attenuated booth in the UC Santa Cruz Phonetics Laboratory, using Superlab experiment presentation software (Cedrus, Version 4.5) with a response pad of which two buttons were used for recorded responses.

In the learning phase of the experiment, participants were told that they would be listening to sentences of a ‘made-up language’. They were instructed to listen carefully to each sentence, repeat the sentence aloud, and then press any button on the response pad to proceed to the next sentence. Sentences were blocked by frame type, with sentences having utterance-final target words (3a) in the first block and those having utterance-medial test words (3b) in the second block. These blocks were repeated in this order two more times, with the sentences presented in a different randomized order within the block in each repetition. Participants were given the opportunity to rest between blocks. The total number of trials in the learning phase was 180 (3 presentations of the 60-sentence learning set).

The participants were divided into two groups. The **final devoicing learning group** were presented with the sentences of the final devoicing learning set during the learning phase, while the **final voicing learning group** heard only the sentences of the final voicing learning set.

After a participant finished the learning phase of the experiment, he or she entered the testing phase. In this phase, participants from both learning groups heard the sentences of the testing set: novel sentences with novel target words (Table 1b), none of which had been encountered in the learning phase by either learning group.

The participants were instructed to listen to each sentence and press a blue button on the right of the button box if they thought the sentence belonged to the language they had learned, and a red button on the left if they thought the sentence did not belong to that language. The presentation of the stimuli was self-paced in both the learning phase and the testing phase. The participant could hear the stimulus only once in a given trial, and the next stimulus was presented only when they pressed a button.

As in the learning phase, the sentences in the testing phase were blocked by sentence frame, with the sentences with sentence-final test words as in (3a) presented in the first block, and those with sentence-medial test words as in (3b) presented in the second block. This was done in order to encourage participants to compare sentences only within a given frame. Sentences of the form (3b) with medial test words were less frequent in each learning set than sentences of the form (3a), since the latter included all 4 classes of test word in Table 1a, while the former only included the nasal- and vowel-final test words. The presentation in blocks was intended to discourage participants from treating sentences of the form (3b) as less well-formed than those of form (3a) purely on the basis of relative frequency.

After the experiment, participants were given a debriefing questionnaire, which asked them what they based their responses on during the experiment. Most had nothing concrete to offer, suggesting that any learning was relatively implicit. None mentioned any kind of restriction on [s] or [z], though three noticed that whether a target word contained [s] or [z] mattered.

#### 2.2.1.4 Hypotheses

The experiment tests the following three hypotheses.

*Hypothesis 1 (Learning).* If the participants are learning the pattern they are exposed to in the learning set, then they should accept more forms consistent with that pattern than forms that are inconsistent with it. Participants in the final devoicing condition (where utterance-final obstruents were always [s]) will accept sentences with utterance-final [s] more often than those with utterance-final [z], and the participants in the final voicing condition should display the opposite tendency. Such a response pattern would be evidence of an effect of the learning set on the participants' responses in the test phase.

*Hypothesis 2 (Domain Generalization).* If learners generalize as expected, participants who learn the utterance-final pattern as shown by their responses in the first test phase will follow the same pattern of responses for utterance-medial words in the second test phase. Thus participants in the

final devoicing condition will accept word-final [s] more often than word-final [z] in utterance-medial as well as utterance-final position. Participants in the final voicing condition will accept word-final [z] in both sentence contexts more often than word-final [s]. Such a pattern of responses would support the claim that learners generalize the domain of such patterns, as presupposed in the analogical generalization account.

*Hypothesis 3 (Naturalness)*. Participants in the final devoicing learning group had the task of learning a phonetically natural pattern, which has a robust phonetic precursor and is widely attested, while the participants in the final voicing learning group had to learn an equally simple pattern of final voicing, which has no phonetic basis and is barely attested (Yu 2004). If the latter kind of pattern is harder for people to learn, then we would expect the participants in the final devoicing learning group to have a higher proportion of 'correct' responses (corresponding to the distribution in their learning set) than the participants in the final voicing learning group.

Some studies have found evidence that particular 'natural' patterns were more successfully learned than comparable 'unnatural' patterns (Schane et al. 1975; Wilson 2003; Peperkamp et al. 2006; Wilson 2006; Moreton 2009; Carpenter 2010). However, other studies have found no difference in learnability between the two kinds of patterns (Pycha et al. 2003; Pycha et al. 2006). Moreton and Pater (2012a, b) distinguish between cases in which the two patterns differ in their formal complexity and those in which they differ in phonetic grounding. In their review of the literature, they find ample evidence of greater learning success for formally simpler patterns (analytic bias), but no convincing evidence of greater learning success for patterns with a stronger phonetic motivation (channel bias). The two patterns to be learned in this experiment do not differ in formal complexity, since they differ just in which of two classes of obstruents is allowed in utterance-final position, so if there is a difference in learning performance between the two learning groups that might provide evidence for a role of phonetic grounding in learnability.

### 2.2.2 Results

Each participant responded to 144 test sentences in the testing phase, of which 72 had obstruent-final target words. Only the sentences with obstruent-final target words were included in the analysis, since the hypotheses are all about the responses to voicing in word-final obstruents.

One participant in the final voicing learning set accepted every stimulus, and a participant in the final devoicing learning set accepted all but 2. These participants were evidently for some reason not responding to the stimuli, so their responses were excluded from the analysis. That left 41 participants, 19 belonging to the final devoicing group and 22 to the final voicing group.

The proportion of 'accept' responses for sentences with obstruent-final target words is presented in Figure 1 by final segment class, learning group, and the position of the target word in the sentence. Figure 1a summarizes responses to stimuli in which the target word was in sentence-final position. It can be seen that participants in the final devoicing group accepted sentences with utterance-final [s] more often than they accepted sentences with utterance-final [z], while participants in the final voicing group had the opposite pattern.

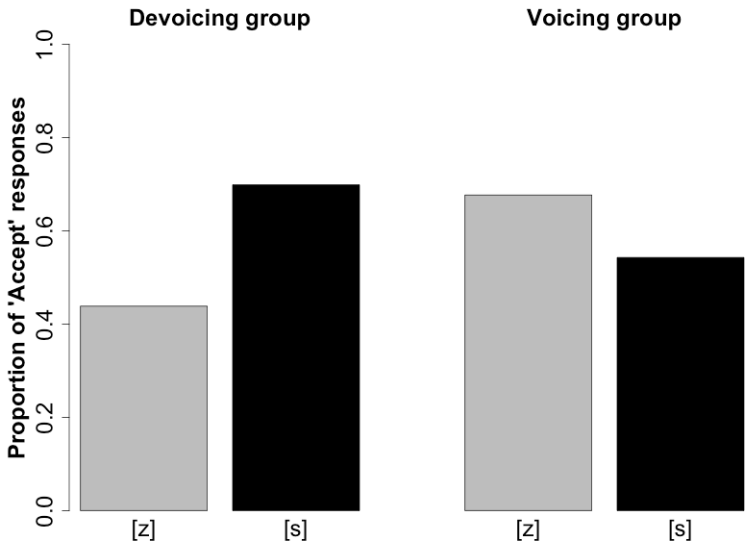


**Fig. 1a: Proportion of 'accept' responses by word-final voicing category and learning group (Sentence-final position)**

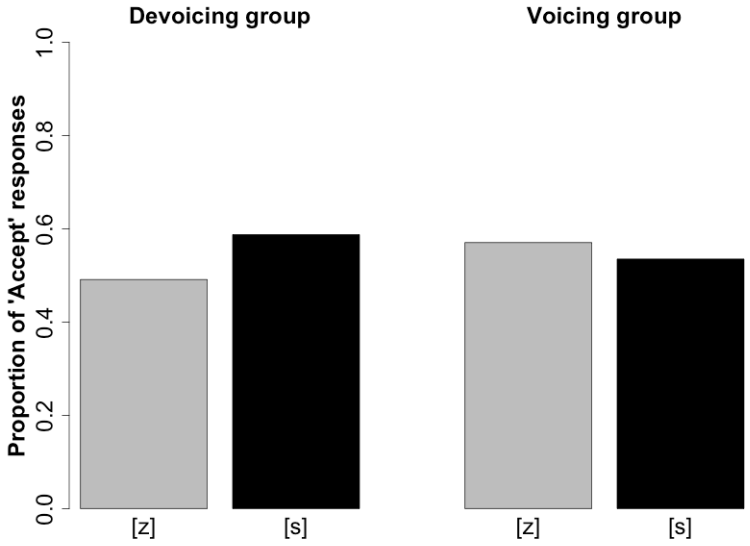


**Fig. 1b: Proportion of 'accept' responses by word-final voicing category and learning group (Sentence-medial position)**

Figure 1b presents the responses to stimuli in which the target word was nonfinal in the sentence. There were no obstruent-final words in this position in either learning set, but it still can be seen that the participants in the final devoicing group accepted more sentences with [s] at the end of the test word than sentences with [z] in that position, while participants in the final voicing group had the opposite pattern. The difference in response between [s] and [z] is however smaller in the nonfinal cases in Figure 1b than it is in the final cases in Figure 1a.

The data were modeled by means of a mixed model logistic regression analysis, employing the *lme4* package in R (Baayen et al. 2008; Bates et al. 2012; Barr et al. 2013). The dependent

variable was the response – 'no' or 'yes' – to the question of whether the stimulus belonged to the language the participant had heard in the learning set. 'Yes' was treated as the marked value of the response variable (coded as 1), and 'no' as the default (coded as 0). The random effects were the participants and the items (the sentences). The fixed effects were the word-final segment category (voiced or *voiceless*), the sentence position of the test word (final or *nonfinal*), and the learning group (final devoicing or *final voicing*). In each of these factors, the first level was treated as the default, and the second as the marked value. Intercepts were included in the model for both random effects, as well as slopes for each participant for segment class and sentence position. Alternative models differing in the random effects structure (e.g. those lacking either intercept or either slope) were found to be less well-fitting in ANOVA comparisons based on log likelihood (Baayen et al. 2008). The results for the fixed effects are given in Table 2.

| **Factor** | **b** | **z** | **p** |
|---|---|---|---|
| Intercept | -0.27 | -1.2 | .22 |
| **Voicing** (voiceless) | **1.31** | **5.3** | **<.0001*** |
| Sentence position (nonfinal) | 0.25 | 0.9 | .36 |
| **Learning group** (final voicing) | **1.08** | **3.9** | **<.0001*** |
| **Voicing * Sentence position** | **-0.80** | **-2.7** | **<.007*** |
| **Voicing * Learning group** | **-1.93** | **-6.6** | **<.0001*** |
| **Sentence position * Learning group** | **-0.68** | **-2.0** | **.04*** |
| **Voicing * Sentence position * Learning group** | **1.24** | **3.8** | **<.0002*** |

**Table 2**: Fixed effects (significant effects ($p < .05$) marked by boldface)

There were significant main effects of voicing and learning group, reflecting an overall greater proportion of 'yes' responses for voiceless-final test words and for the final voicing learning group. All of the interactions were also significant.

To explore these interactions, the dataset was broken down into subsets according to sentence position and learning condition. Table 3 gives the results for 4 tests, with the same random effects structure as in the previous analysis, and the single fixed effect of voicing.

| **Sentence position** | **Learning class** | **b** | **z** | **p** |
|---|---|---|---|---|
| **Final** | **Devoicing** | **0.82** | **3.2** | **.001*** |
| **Final** | **Voicing** | **-0.62** | **-2.4** | **.02*** |
| **Nonfinal** | **Devoicing** | **0.51** | **2.4** | **.02*** |
| Nonfinal | Voicing | -0.14 | -0.6 | .54 |

**Table 3**: Effects of voicing in data subsets defined by learning group and sentence position

Looking at the sentence-final cases in Table 3, it can be seen that voicing had a significant effect on responses for both learning groups. In the devoicing learning group, a sentence with a final voiceless obstruent was significantly more likely to be accepted than a sentence with a final voiced obstruent, while the opposite was the case in the voicing learning group. This pattern of

responses reflects the distribution of obstruents in the respective learning sets, and indicates that some learning of the distribution pattern did occur.

Turning to the nonfinal cases in Table 3, voicing had a significant effect on responses for the devoicing learning group but not for the voicing learning group. The significant effect in the former case supports the hypothesis that the learners applied the final devoicing pattern in these cases, for which there was no model in their learning set. The lack of a significant effect in the voicing learning group means that the null hypothesis cannot be rejected in this case, the null hypothesis being that the voicing pattern of the learning set is not applied in this case.

A reviewer notes that the effect of learning group could reflect a simple preference for [s] or [z], due to the fact that there are 24 more instances of [s] than [z] in the devoicing learning set, and the same difference the other way for the voicing learning set. To test whether this difference in segment frequency led to a preference for a given voicing independent of word position, we tested the effect on responses of nonfinal [s] and [z] in test words, as an extra factor added to the models in Table 3. For the devoicing learning group, in which final [z] was excluded, a factor was included encoding whether the word included a nonfinal [z], and for the voicing learning group a factor was included encoding whether the word included a nonfinal [s]. If subjects were responding to the frequency of these two sounds, regardless of that sound's position in the word, then it would be expected that these nonfinal instances would affect the responses. But this factor had no significant effect in any of the four data subsets treated in Table 3. We can conclude that the sensitivity to voicing between the learning groups was specifically sensitivity to the voicing of the word-final obstruent.

To test whether one of the two learning set patterns was more successfully learned than the other, the responses were recoded in terms of correctness, where a response was treated as 'correct' if it corresponded to the learning set pattern, and incorrect otherwise. For the devoicing learning group, it was correct to reject a sentence with a z-final test word, or to accept a sentence with any other sort of test word. For the voicing learning group, it was correct to reject a sentence with an s-final test word, or to accept a sentence with any other sort of test word. Other responses were incorrect.

| | **Sentence-final** | **Nonfinal** |
|---|---|---|
| *Devoicing learning group* | 63% | 55% |
| *Voicing learning group* | 57% | 52% |

**Table 4**: Percent of correct responses by learning group and sentence position.

As can be seen in Table 4, there was a higher percent of correct responses for the sentence-final cases than the nonfinal cases, and a higher percentage of correct responses for the devoicing learning group than for the voicing group. 74% of the devoicing group gave correct responses to over half the stimuli in both final and nonfinal positions, while 26% met that criterion in the final condition but not the nonfinal. In the voicing group, 41% gave correct responses to over half the stimuli in both conditions, 23% achieved that only for the final condition, and 32% did not reach that for either condition. One subject in the voicing condition had 36% correct in the final condition, and 53% correct in the nonfinal condition.

The effect of both learning group and position were found to be significant in a mixed model regression analysis with the correctness of the response as the dependent variable. Random effects intercepts were included in the model for both participant and sentence, with a slope factor for voicing within participant. The fixed effects were Learning group and Sentence position. Both had significant main effects, and their interaction was not significant, as can be seen in the summary in Table 5.

| **Factor** | **b** | **z** | **p** |
|---|---|---|---|
| **Intercept** | **-0.62** | **-5.9** | **<.0001*** |
| **Sentence position** (nonfinal) | **0.40** | **3.3** | **.001*** |
| **Learning group** (final voicing) | **0.30** | **2.1** | **.03*** |
| Sentence position * Learning group | -0.18 | -1.1 | .27 |

**Table 5**: Mixed model logistic regression analysis results (Experiment 1)

There were significantly more incorrect responses in nonfinal compared to final position, indicating that participants learned the distribution pattern more successfully in the position where they had direct evidence of the pattern in their learning sets. The fact that there were significantly more incorrect responses for the final voicing group than for the final devoicing group supports the hypothesis that the final devoicing pattern is more easily learned than the final voicing pattern.

A reviewer suggests that the difference between the final and nonfinal cases could be due to the order of presentation, since the block of sentences with final test words were always presented before the nonfinal block. Thus the difference in correctness between final and nonfinal could be due to fatigue increasing as the experiment progressed, to forgetting the training generalizations, or to the effect of exposure to pattern-incongruent sentences during the testing phase. The only direct response to this would be to run the experiment with the blocks in the opposite order, which we were unable to do. However, we did examine the order of presentation of the stimuli within the blocks (which varied randomly across participants). The position (1 – 72) of each test stimulus in the within-block order of presentation was included as an additional factor in a model otherwise identical to that in Table 5. It did not prove to be a significant effect ($b = 0.002$, $z = 0.9$, $p = .39$). The test thus provided no evidence for the interpretation of the difference between the final and nonfinal cases as a gradual order of presentation effect.

### 2.2.3 Discussion of Experiment 1

The results of the first experiment show that the participants learned the patterns of distribution in their learning sets on relatively brief exposure (60 sentences repeated three times), as evidenced by their responses to novel items. None of the participants completely mastered the patterns of distribution in the learning sets (the highest proportion of correct responses for any condition being just 78%), but they showed a significant tendency to accept pattern-congruent items more often than pattern-violating items.

Given only evidence of a restriction in sentence-final position in the learning phase, their responses in the test phase clearly showed that they had extended the pattern to sentence-medial word-final position, at least in the case of the final devoicing pattern. The experiment thus demonstrates that domain generalization is a significant tendency among language learners. The lack of evidence of such generalization in the case of the voicing group could just reflect the weaker performance in both the voicing group (compared to the devoicing group) and the nonfinal sentences (as compared to the final ones) combining to make the effect too small to be significant.

The participants' responses reflected the distributional pattern in the learning set both for sentence-final targets and sentence-medial ones, but the strength of this effect was significantly weaker in the sentence-medial case. There are several possible explanations for this fact. Participants had direct evidence in their learning sets about the sentence-final case, but they had no direct evidence about the distribution of voicing in sentence-medial cases. The sentence-medial cases required a step of extrapolation from their learning experience which was not required in the sentence-final cases. Alternatively, trials with utterance-medial target words may have been harder because they were longer or because participants had to locate the target word between non-target words.

It was also found in this experiment that participants learned final devoicing more successfully than final voicing. This is then a case in which a widely attested, phonetically natural pattern proved more learnable than an equally simple pattern that is barely attested and has no phonetic basis. There are several possible interpretations of this finding. First, the greater learnability of the final devoicing pattern could be due to channel bias (Moreton 2008; Moreton & Pater 2012a, b) in the experiment. The normal partial devoicing of utterance-final voiced obstruents could have affected our stimuli, leading to a tendency for experiment participants to misidentify final voiced obstruents as voiceless (Myers 2012). Such a tendency would tend to undermine the final voicing pattern for those participants, since some of the final voiced obstruents would be misidentified as voiceless. There is no corresponding tendency for utterance-final voiceless obstruents to be voiced or misidentified as voiced, so the final devoicing pattern would not be undermined in the same way. If this was the case, the less natural pattern was not harder to learn as a pattern (reflecting a substantive learning bias, see below), but was just subject to more phonetic ambiguity in category identification than the more natural pattern.

To test this interpretation, we submitted all the obstruent-final stimuli to 17 native speakers of American English, who were asked to identify the final consonant of the varying test word as either *s* or *z*. Just as in Experiment 1, the materials were presented to participants through headphones in a sound-attenuated booth in the UC Santa Cruz Phonetics Laboratory, using Superlab experiment presentation software (Cedrus, Version 4.5) with a response pad of which two buttons were used for recorded responses. The stimuli were blocked by the position of the varying test item (final/nonfinal), and half the participants were given the utterance-final block first and the other half had the blocks in the opposite order. Data were excluded from two bilingual speakers who had less than 75% correct overall.

The proportion of correct identification responses for each final consonant and each sentence position are given in Table 6.

| | Final sentence position | Nonfinal sentence position |
|---|---|---|
| Word-final [s] | 90 | 85 |
| Word-final [z] | 86 | 78 |

**Table 6**: Percentage of correct responses by final consonant and sentence position (stimulus identification test)

It can be seen that the word-final voiceless obstruents were correctly identified more often than the voiced ones, in both final and nonfinal sentence position. Unexpectedly, the obstruents in final position were identified correctly more often than in nonfinal position. The data were submitted to a mixed model logistic regression analysis, with random effect intercepts for participant and item, and slopes within participant for both voicing and position. The dependent variable was the accuracy of the response (correct/incorrect), with *correct* treated as the default.

The results are given in Table 7:

| **Factor** | **b** | **z** | **p** |
|---|---|---|---|
| **Intercept** | **-2.84** | **-8.3** | **<.0001*** |
| Voicing (voiced) | 0.46 | 1.1 | .25 |
| Sentence position (nonfinal) | 0.67 | 1.6 | .10 |
| Voicing * Sentence position | 0.07 | 0.1 | .88 |

**Table 7**: Mixed model logistic regression analysis results (stimulus identification test)

There was no effect either of the voicing of the test consonant or the position of the test word in the sentence. The test therefore provided no evidence for the channel bias account of the learning group effect.

Alternatively, the more successful learning of final devoicing compared to final voicing could reflect a substantive learning bias (Wilson 2006). This need not be understood as an innate or universal bias in favor of a phonetically natural pattern, however. Second language patterns (such as the ones in our experiment) might be more learnable to the extent they coincide with the phonetic, phonological, or lexical tendencies in one's first language. For example, English allows both voiced and voiceless obstruents in word- and utterance-final positions (e.g. *maze/mace*), but there is a strong tendency toward partial devoicing of utterance-final obstruents (as noted in Section 1.1). It could be that the participants in our study, who all were speakers of English, had a preference for a categorical pattern corresponding to that phonetic tendency, over a categorical pattern that runs opposite to the phonetic tendency in their first language. Alternatively, a bias due to lexical frequency could exist due to a greater number of English words ending in [s]

compared to [z]. Yet a count of all [s]- and [z]-final wordforms in the WebCelex lexical database (http://celex.mpi.nl/) shows that the opposite is true: there are 12,770 ending in [s] and 22,023 ending in [z], for an [s]-to-[z] ratio of 0.58. If we restrict the search to words ending in [s] or [z] preceded by a vowel, since these words better resemble our experimental stimuli, the counts are 4,385 and 12,587 respectively, for a ratio of 0.35. Final [z] outnumbers final [s] because of the great number of plural noun and third person singular verb forms ending in [z].[4]

We refer to the target items in the experiment as words, but one might wonder whether participants in our experiments really understood experimental stimuli such as *ˈsanta pis ˈmizupu* (where *pis* is the target word) as three words contained within an utterance, as opposed to something else, e.g., morphemes contained within a word, words within words (compounds), or various combinations of these things. What we have called target words are the varying sequences set in the invariant frames: *ˈsanta* ___ and *ˈsanta* ___ *ˈmizupu*. Participants were thus encouraged to treat *ˈsanta, ˈmizupu* and the target words as units smaller than the utterance. The context units *ˈsanta* and *ˈmizupu* occurred in all of the sentences, while the target words belonged to a broader inventory of 36 sequences, each of which occurred 4-8 times in the learning set. Within the target words, there weren't any recurring sequences that could be interpreted as smaller units. So the target words of the experiment were sequences smaller than the utterance which could not be broken down into smaller recurring units. This is what makes them the terminal units of this very simple grammar, and so analogous to words in natural languages. Whether the participants thought of them as words or not, they did treat them as a unit smaller than the utterance that was the domain of a generalization about the distribution of voicing.[5]

### 2.3 Experiment 2

The patterns that participants learned in Experiment 1 were reflected in distributional generalizations, but not in any phonological alternations among differing versions of the same item. Such alternation patterns are clearly an important aspect of phonological knowledge, so Experiment 2 is designed to test whether domain generalization also occurs when the distributional pattern is reflected in alternations. Experiment 2 also differs from Experiment 1 in testing learning and domain generalization across the consonant stops [p,t,k] vs. [b,d,g] rather than over [s] and [z] only. A finding of domain generalization in this second experiment would heighten our confidence that the result is robust.

[4] If we limit the searches to words having a spoken frequency greater than or equal to 25 occurrences per million words, [z]-final words remain more preponderant. The ratios corresponding to those above are 0.76 (all words) and 0.55 (preceding vowel). These searches assume that type frequency is the relevant measure. We might also consider the average token frequency of the forms. Excluding wordforms with a listed frequency of 0 occurrences per million words, the average COBUILD spoken frequency of [z]-final words is 20.0, versus 22.2 for [s]-final words, not a large difference. These numbers are 26.5 versus 32.5 respectively when there is a preceding vowel.

[5] It is also relevant that all of our target words bore initial stress. English speakers have been shown to perceive word boundaries before stress (Cutler & Norris 1988).

### 2.3.1 Methods
#### 2.3.1.1 Participants

There were 25 participants in Experiment 2, who were all undergraduate students at the University of California, Santa Cruz. 17 of them were taking an introduction to linguistics course, and they received course credit for participating. The rest had taken no linguistics courses and were volunteers.

All participants were English speakers who had acquired English as children. Many had experience with a language other than English, but none with a language with a final laryngeal neutralization pattern.

Unlike in Experiment 1, all participants in Experiment 2 were presented with the same learning set. There were thus no participant groups and all comparisons were within-subject.

#### 2.3.1.2 Materials

The stimuli in both the learning and testing phase were sentences belonging to one of the two frames in (4), where the blank is the position of the test word, the only varying item in the sentence. The sentences consisted of CV(C) syllables, where C belonged to the set [m, n, p, t, k, b, d, g], and V belonged to the set [i, e, a, o, u].

(4) a. Utterance-final position: biˈtomi ___.
b. Utterance-medial position: biˈtomi ___ ˈnama.

The varying target words for this experiment consisted of pairs of items in which one had the form ˈCVC (the *singular* form), and the other ˈCVCi (the *plural* form). In *nonalternating* pairs, such as *pet – peti***,** the initial CVC (the *stem*) was identical in the singular and plural. In alternating pairs, such as *git – gidi***,** the initial CVC differed just in voicing of the final consonant.

The target word pairs in the learning phase were of three types: (1) nonalternating pairs in which the stem-final C was nasal (*nasal nonalternating, e*.g. *min* - *mini*), (2) nonalternating pairs in which the stem-final C was a voiceless stop [p, t, k] (*voiceless nonalternating*, e.g. *kip – kipi*), and (3) alternating pairs in which the singular ended in a voiceless stop and the plural had a corresponding voiced stop (*voiceless-voiced alternating*, e.g. *git – gidi*). The full set of target words used in the learning phase is given in Table 8. There were 10 word pairs in each of the three types, for a total of 30 word pairs in the learning set.

| *Nasal non-alternating* | | *Voiceless non-alternating* | | *Voiceless-voiced alternating* | |
|---|---|---|---|---|---|
| *sg* | *pl* | *sg* | *pl* | *sg* | *pl* |
| min | mini | kip | kipi | git | gidi |
| pim | pimi | bik | biki | mip | mibi |
| dem | demi | gek | geki | tik | tigi |
| ken | keni | pet | peti | met | medi |
| gan | gani | pap | papi | kap | kabi |
| tam | tami | mat | mati | bak | bagi |
| dom | domi | dok | doki | got | godi |
| pon | poni | not | noti | dop | dobi |
| bun | buni | guk | guki | puk | pugi |
| num | numi | tup | tupi | tut | tudi |

**Table 8: Target words used in learning phase of Experiment 2**

The sentences of the learning set included all of the nasal-final items in Table 8 in both contexts (4a) and (4b), but the obstruent-final items (voiceless nonalternating and alternating) only in the utterance-final context (4a). Thus, as in Experiment 1, there are no coda obstruents in the learning set except in sentence-final position. There were 40 singular-plural sentence pairs in the learning set: 30 in the sentence-final frame (4a), and 10 (all with nasal-final target-word stems) in the sentence-medial frame (4b).

The sentences of the testing stage consisted of the same two sentence frames in (4), with the target words listed in Table 9. The sentences included voiceless nonalternating and voiceless-voiced alternating pairs of novel test words, as in the learning set, but also *voiced nonalternating* pairs such as *teb – tebi* and *voiced-voiceless alternating* pairs such as *teb – tepi*.

| *Voiceless non-alternating* | | *Voiced-voiceless alternating* | | *Voiceless-voiced alternating* | | *Voiced non-alternating* | |
|---|---|---|---|---|---|---|---|
| *sg* | *pl* | *sg* | *pl* | *sg* | *pl* | *sg* | *pl* |
| tep | tepi | teb | tepi | tep | tebi | teb | tebi |
| gep | gepi | geb | gepi | gep | gebi | geb | gebi |
| nok | noki | nog | noki | nok | nogi | nog | nogi |
| buk | buki | bug | buki | buk | bugi | bug | bugi |
| nit | niti | nid | niti | nit | nidi | nid | nidi |
| bit | biti | bid | biti | bit | bidi | bid | bidi |
| ket | keti | ked | keti | ket | kedi | ked | kedi |
| bet | beti | bed | beti | bet | bedi | bed | bedi |
| tat | tati | tad | tati | tat | tadi | tad | tadi |
| pat | pati | pad | pati | pat | padi | pad | padi |
| dot | doti | dod | doti | dot | dodi | dod | dodi |
| mot | moti | mod | moti | mot | modi | mod | modi |
| kut | kuti | kud | kuti | kut | kudi | kud | kudi |
| gut | guti | gud | guti | gut | gudi | gud | gudi |

**Table 9: Target words used in the testing phase of Experiment 2**

There were 14 word-pairs in each of the 4 classes in Table 9, for a total of 56 word pairs in the test set. Each of the 56 word pairs in Table 9 occurred in each of the 2 contexts in (4), for a total of 112 stimulus sentence pairs.

The sentences were produced by a phonetically-trained speaker with a Canadian English pronunciation of consonants (e.g. aspiration of syllable-initial voiceless stops), a declarative intonation and no internal pauses.

### 2.3.1.3 Procedure

Experiment 2 was carried out using the same equipment as in Experiment 1.

In both the learning phase and the testing phase, sentences were presented in pairs in which the two sentences were identical except that one had the singular member of a test-word pair, and the other the corresponding plural member, e.g. *bitomi min – bitomi mini*. The sentence-pairs were accompanied by schematic drawings as in Figure 2. The sentences with singular forms (CVC) were accompanied by a simple drawing of a stick figure pointing at a single triangle (as in Fig. 2a), while during the sentences with plural forms (CVCi), the same figure pointed at three triangles (as in Fig. 2b). The triangles in the diagrams were black in the sentence-final condition and yellow in the sentence-medial condition. The pictures were meant to suggest schematic meanings for the *–i* 'suffix' (plural) and for *nama* (yellow).

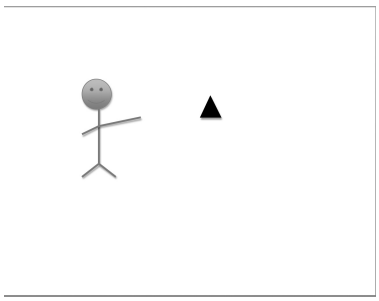

**Fig. 2a: Image accompanying sentences with singular forms in final position.**

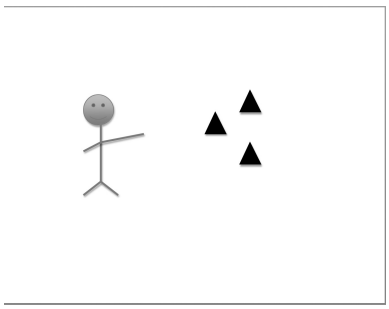

**Fig. 2b: Image accompanying sentences with plural forms in final position.**

At the beginning of the learning stage, participants were told that they would be hearing pairs of sentences of a 'made-up language'– the first sentence with the singular form of a word, and the second sentence with the plural form of that word. In the first run-through of the learning stage, the participants were instructed to repeat aloud each sentence that they heard. The sentence-pairs were blocked by sentence frame: the subjects heard all the sentence-pairs in the learning set that had the sentence-final frame (4a), followed by all the sentence-pairs in the learning set that had the sentence-medial frame (4b). The sentence-pairs within a block were presented in randomized order, different for each participant.

Having heard and repeated all the sentence-pairs in the learning set, participants then entered a feedback phase of the learning stage. In this phase, they first heard the sentence with the plural form. They were instructed to listen to this sentence, and then speak aloud the corresponding second sentence in the pair – the one with the corresponding singular form. After saying this sentence aloud, they were instructed to press any response button, at which point they heard the sentence with the correct singular form. As in the first phase, all the sentences with the final frame (4a) were presented, followed by all the sentences with the nonfinal frame (4b). The feedback phase was repeated three times. Thus a participant encountered each of the 40 singular-plural sentence pairs in the learning set 4 times: once in the initial listening-and-repeating phase, and 3 times in the feedback phase.

When they completed the learning phase, participants entered the test phase. During this phase, participants heard a sentence containing a plural form from Table 9 followed by a sentence with the corresponding singular from Table 9. The target items were novel in that none of them had occurred in the learning stage. When they heard the second sentence, participants were instructed to press the blue, right-hand button on the response pad if they judged that the pair of sentences belonged to the language they had heard during the learning stage, and otherwise to press the red, left-hand button. As in the learning stage, each sentence was accompanied by the corresponding

schematic drawing. The drawing for the singular form was labeled with a prompt question 'Belongs to the language?' together with the two possible answers: 'No' (in red at the left edge of the window) and 'Yes' (in blue at the right edge of the window).

There were two blocks in the testing phase: all the sentences with the sentence-final frame (4a), followed by all the sentences with the utterance-medial frame (4b). Participants heard all pairs of forms given in Table 9, in both sentence positions in (4). Therefore participants encountered 112 trials in total during the test phase: the 56 target word-pairs in Table 9 in the 2 sentence frames in (4).

After the test phase was completed, participants completed a debriefing questionnaire which asked them how they had distinguished sentences that belonged to the language from those that did not. Many noticed that the basis of the difference lay in the way the singular differed from the plural, and some noticed that the crucial pattern had to do with pairs of consonants such as *t* and *d*. None succeeded in verbalizing the actual pattern of distribution of voiced and voiceless obstruents in the learning set.

#### 2.3.1.4 Hypotheses

Experiment 2 tests the following three hypotheses:

*Hypothesis 1 (Learning)*. It was expected that the participants' responses would be affected by the pattern of distribution in the learning set. Thus participants would accept sentence pairs with utterance-final voiceless obstruents more often than those with utterance-final voiced obstruents.

*Hypothesis 2 (Domain Generalization)*. It was expected that participants would generalize the pattern of distribution from utterance-final to word-final position. This would be reflected in the participants accepting sentence pairs with word-final voiceless obstruents more often than those with word-final voiced ones in utterance-medial position as well as utterance-final position. Utterance-final position was the only position in which they had experienced coda obstruents during the learning set, so such a pattern of response would show an extension of the pattern beyond the context for which they had direct evidence.

*Hypothesis 3 (Alternation Resistance)*. It was expected that participants would accept more nonalternating pairs than alternating pairs, regardless of whether they conformed to the pattern of sound distribution in the learning set. Tessier (2006) found avoidance of alternation in an artificial language learning experiment testing the learning of a voicing pattern with 4-year-old English learners. Kerkhoff (2004) found that Dutch-speaking children as old as 7;8 were still making voicing errors in a task of producing the plural corresponding to a given singular form, and that these errors were generally in the direction of preserving the same voicing in the plural as in the singular (e.g. giving incorrect nonalternating **bɛtən* as the plural of *bɛt* 'bed', instead of the correct alternating form *bɛdən*).

### 2.3.2 Results

Of the 25 participants, one responded 'yes' to every single stimulus, and another responded 'yes' to all but two. These participants did not seem to be responding to the stimuli, so their responses were excluded from the analysis. That left 23 participants, each responding to 112 sentence-pairs in the testing stage. One participant twice pressed a key besides the two labeled ones, so those two responses were excluded as uninterpretible, leaving a total of 2574 responses in the analysis.

The acceptance rates for stimuli by condition are presented in Figure 3. It can be seen that sentence pairs with voiceless-final stems were accepted more often than those with voiced-final stems. Within the same position class (final/nonfinal), non-alternating pairs were accepted more often than alternating ones.

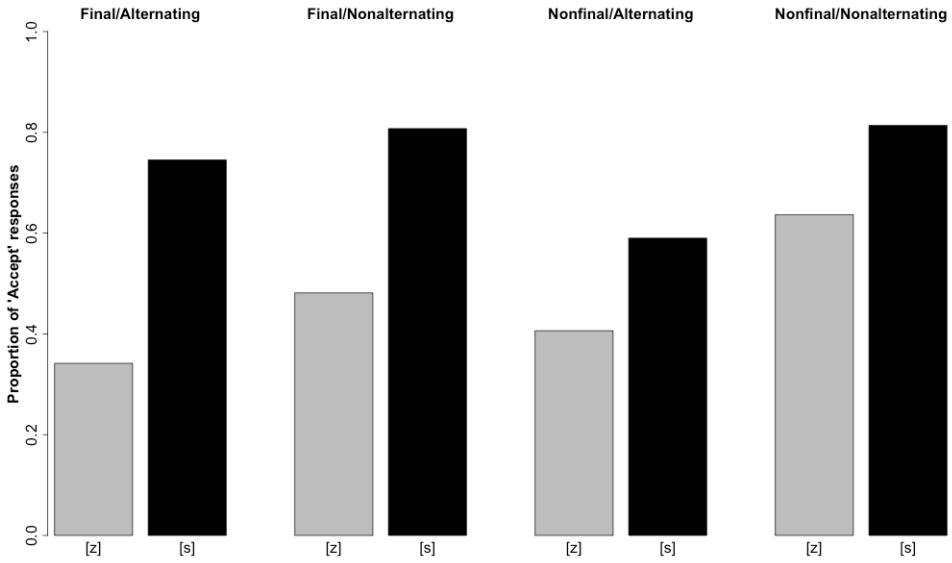


**Fig. 3: Proportion of 'accept' responses by stem-final voicing category, utterance position and alternation status**

The hypotheses were tested in a mixed-model logistic regression analysis, as in Experiment 1. The dependent variable was the response, with 'yes' coded as the marked value. The random effects were participant and item (sentence pair). The fixed factors were voicing (voiced/*voiceless*), utterance position (final/*medial*) and alternation status (alternating/*non-alternating*). In each pair of levels within a factor, the first was default (coded 0) and the second was marked (coded 1). The model included intercepts for both random effects (participant and item), as well as slopes within participant for all three fixed effects. It was found through ANOVA comparison with respect to log likelihood that any model with any simpler random effects structure was significantly less well-fitting than this maximal model. The results of the statistical test are given in Table 10.

| Factor | b | z | p |
|---|---|---|---|
| **Intercept** | **-0.83** | **-2.7** | **.006*** |
| **Voicing** (voiceless) | **2.54** | **5.1** | **<.0001*** |
| Alternation (nonalternating) | 0.67 | 1.6 | .11 |
| Sentence position (nonfinal) | 0.37 | 1.5 | .15 |
| Voicing * Alternation | -0.37 | -1.1 | .29 |
| **Voicing * Sentence position** | **-1.41** | **-4.3** | **<.0001*** |
| **Alternation * Sentence position** | **0.69** | **2.1** | **.03*** |
| Voicing * Alternation * Sentence position | 0.43 | 0.89 | 0.37 |

**Table 10**. Fixed effects for Experiment 2 (significant effects highlighted in boldface)

There was a significant main effect of voicing, with the likelihood of a 'yes' response significantly greater with a voiceless-final test words than with test words ending in a voiced obstruent. The main effects of alternation and position were not significant, but there were significant interactions of voicing and alternation with sentence position.

The results provide only qualified support for the hypothesis of a bias against alternating sentence pairs. There was not a significant tendency to accept nonalternating over alternating sentence pairs in general, though the significant interaction between alternation and sentence position reflects that there was such a tendency for sentence pairs with test words in the nonfinal sentence position.

To explore the interactions, the data were divided into subsets according to alternation and sentence position. Mixed-model logistic regression analyses were run to test whether voicing had a significant effect on responses in each of these subsets. Response (no/*yes*) was the dependent variable, with *yes* as the marked value. Random effects were subject and sentence pair. An intercept was included in the model for each random effect, as well as a slope variable within subject for voicing. The only fixed effect was voicing (voiced/*voiceless*). The results are given in Table 11.

| Sentence position | Alternation class | b | z | p |
|---|---|---|---|---|
| **Final** | **Alternating** | **2.57** | **4.3** | **< .0001*** |
| **Final** | **Nonalternating** | **2.75** | **4.6** | **< .0001*** |
| **Nonfinal** | **Alternating** | **1.25** | **2.2** | **.03*** |
| **Nonfinal** | **Nonalternating** | **1.44** | **3.0** | **.002*** |

**Table 11**. Effect of voicing in position and alternation subsets.

Sentence pairs with word-final voiceless obstruents were significantly more likely to get accepted than pairs including test words with final voiced obstruents, whether the pair was alternating or not, and whether the test word was utterance-final or not. This supports the hypothesis that the pattern of final devoicing was generalized beyond utterance-final position to include utterance-medial word-final position.

As in the case of Experiment 1, it would be possible that the participants had a preference for voiceless obstruents over voiced obstruents, regardless of position, due to the greater frequency of occurrence of voiceless obstruents in the learning set. To test for this possibility, the model presented in Table 10 was augmented by two additional fixed effects: the presence of a nonfinal [p/t/k] in the test word (present/absent), and the presence of a nonfinal [b/d/g] in the test word (present/absent). In each case, absence was the default value and presence the marked value. Neither factor had a significant effect: nonfinal [p/t/k] ($b = 0.04$, $z = 0.2$, $p = .82$), nonfinal [b/d/g] ($b = 0.29$, $z = 1.9$, $p = .06$). Note that the effect of a nonfinal [b/d/g] just misses significance ($\alpha = .05$), but the effect is in the wrong direction for the hypothesis, since the presence of a nonfinal [b/d/g] weakly favored acceptance of the sentence, rather than rejection. These results thus fail to provide evidence for a general preference for [p/t/k] over [b/d/g], regardless of the consonant's position in the word.

To test the hypotheses about the effects of position and alternation on the correctness (pattern-congruence) of the responses, responses were classified as 'correct' if they corresponded to the pattern in the learning set, and 'incorrect' otherwise. Thus a 'yes' response was correct if the sentence pair belonged to the voiceless set, and a 'no' response was correct otherwise. The proportion of correct responses for each alternation and position subset is given in Table 12.

| | **Sentence-final** | **Nonfinal** |
|---|---|---|
| *Alternating* | 70% | 59% |
| *Nonalternating* | 66% | 59% |

**Table 12**: Percent of correct responses by alternation group and sentence position.

For final alternating pairs, 80% of the participants gave correct responses in over half the cases, compared to 76% of the participants for final nonalternating pairs. For nonfinal alternating pairs, on the other hand, only 48% of the participants had correct responses more than half the time, and 52% achieved that level for nonalternating pairs.

A mixed-model logistic regression analysis was performed, with correctness of each response as the dependent variable and *incorrect* the marked value of that variable. The random effects were participant and item (sentence pair), with intercepts for both of those factors and slopes for voicing, alternation group and sentence position within participant. The fixed effects were sentence position and alternation status. The results of the test for the fixed effects are given in Table 13.

| Factor | b | z | p |
|---|---|---|---|
| **Intercept** | **-1.4** | **-4.4** | **<.0001*** |
| **Alternation** (nonalternation) | **0.54** | **2.2** | **.03*** |
| **Sentence position** (nonfinal) | **0.86** | **3.3** | **.0009*** |
| Alternation * Sentence position | -0.37 | -1.2 | .23 |

**Table 13**: Fixed effects for a mixed-model logistic regression analysis of correctness of responses (significant effects ($p < .05$) marked by boldface).

There were main effects for alternation and sentence position, and no significant interaction between the two. There were thus significantly more correct responses in alternating and final sentence pairs.

Since the final sentence pairs were presented in a block before the nonfinal ones, the greater number of errors in nonfinal position could be due to fatigue over the course of the testing stage, or exposure to pattern-incongruent items in the test phase. To test for this possibility, a model was constructed that was identical to that presented in Table 13, except with the addition of one factor: the within-block position of each stimulus item in the presentation for that subject. This factor had no significant effect ($b = -0.002$, $z = -0.7$, $p = .48$). The test thus provides no support for the interpretation of the difference between the final and nonfinal blocks as an effect of the time elapsed from the presentation of the learning set.

### 2.3.3 Discussion of Experiment 2

The results of Experiment 2 support the first two hypotheses. Participants learned the utterance-final pattern, distinguishing voiced-final from voiceless-final items (Learning), and they generalized this learning to word-final, utterance-medial position (Domain Generalization). These results show that domain generalization occurs whether the generalized pattern is reflected in alternations or not, and they replicate the results of Experiment 1 for the more representative set of consonants [p,t,k,b,d,g].

It is possible that the participants, in preferring words ending in voiceless obstruents over those ending in voiced ones, were influenced by their previous experience with English as well as by their experience with the learning set. For example, this could be due to differences in the type frequency of final [p,t,k] versus [b,d,g] in the English lexicon. There are 31,721 wordforms in the WebCelex lexical database (http://celex.mpi.nl/) ending in any of [p,t,k], compared to 30,046 ending in [b,d,g], for a voiceless-to-voiced ratio of 1.06. If we restrict the searches to wordforms in which the final consonant is preceded by a vowel (so that they are more similar to the experimental stimuli), the counts are 19,510 versus 17,803 respectively, for a ratio of 1.10. These are not large differences. However, the differences increase if we restrict the search to wordforms having a COBUILD spoken frequency greater than or equal to 25 occurrences per million words: 559 versus 286 (all words, ratio = 1.95), and 312 versus 179 (preceding vowel, ratio = 1.74). In addition, words ending in [p,t,k] have an average spoken frequency of 28.9, compared to 15.1 for those ending in [b,d,g] (ignoring words with a listed frequency of 0). For words in which the final consonant is preceded by a vowel those frequencies are 39.4 and 9.7 respectively. It is therefore conceivable that our experimental participants were influenced by

English lexical statistics. Against this interpretation, though, we note that the occurrence of [p,t,k] vs. [b,d,g] elsewhere in a stimulus word had no effect on participant judgements, as discussed above, and that the English-speaking participants in Experiment 1 in the final voicing condition preferred final voiced obstruents over final voiceless ones. Subjects seem to be more affected by asymmetries in the experimental learning set than by native lexical statistics.

The results provide weak evidence for the hypothesis that learners have a bias against accepting more than one form for a given item, as argued by Kerkhoff (2004) and Tessier (2006). Such a bias was found in nonfinal sentence position, but not in sentence-final position. Resistance to alternations is an important aspect of phonology. It is reflected in the Faithfulness family of constraints in Optimality Theory (Prince & Smolensky 1993 [2004]; McCarthy & Prince 1995). In rule-based models, it is reflected in the fact that phonological rules only apply when their structural description is satisfied, so that the default is no change.

Finally, it is interesting that there was a small but significant increase in accuracy for the alternating condition, suggesting that the presence of alternating forms may assist in learning a pattern like final devoicing. Learners may resist accepting more than one form for a given item, but once they have done so, the alternation is informative to them, as it is to phonologists.

**3. General discussion and conclusion**

Historical accounts that relate word-final phonological restrictions to the phonetic properties of utterance-final position crucially posit generalization from utterance-final to word-final position. The two experiments in this paper provide the first direct empirical evidence that learners do tend to generalize in this way, and thus provide support for the traditional view that domain generalization is a natural diachronic process.

Why do language learners generalize in this way from utterance-final position? Why didn't the participants in the experiments make the conservative assumption that the restriction on voicing only applied to utterance-final position, since that was the only position for which they had evidence of a restriction?

We might posit that domain generalization is a kind of 'cyclic' or 'paradigm uniformity' effect, reflecting a preference that the utterance-medial variant of a word have the same form as the devoiced utterance-final variant of the same word (Steriade 1997). But in the experiments described above, participants were responding to novel forms, so there was no established utterance-final form of a particular lexical item that could be extended to an utterance-medial form of the same item. This is analogous to the way in which a Russian speaker applies final devoicing to previously unfamilar words despite the lack of any utterance final model, as discussed in Section 1. The effect of the learning set on the test responses must be interpreted as a generalization about the distribution of sounds, not about the form of individual lexical items.

For a learner who encounters utterance-final words with final obstruent devoicing, there is inherent ambiguity about the domain of devoicing – such data are consistent with utterance-final devoicing or with word-final devoicing. Suppose that learners have both interpretations equally available to them. For example, in the context of Optimality Theory Flack (2009) argues that the

existence of a constraint such as *D]$_{Wd}$ (prohibiting voiced obstruents word-finally) necessarily implies the existence of *D]$_{Utt}$ as well. A reviewer points out that, given a learning model such as the Gradual Learning Algorithm in Stochastic Optimlity Theory (Boersma 1997; Boersma & Hayes 2001) or other algorithms that similarly gradually 'promote' constraints based on relevant input encountered, any evidence that leads to the promotion of *D]$_{Utt}$ will necessarily lead to the promotion of *D]$_{Wd}$ as well, while the reverse is not true. This state of affairs will cause *D]$_{Wd}$ to be promoted higher than *D]$_{Utt}$ assuming the presence of any input, in non-utterance-final position, that causes *D]$_{Wd}$ to be promoted. This might happen if a language happens to have more forms with word-final voiceless obstruents than forms with voiced ones. Alternatively, 'paradigm uniformity' based on utterance-final forms (see above) could act as a bridge.[6] In this way utterance-final devoicing might be generalized to all words. However, in the context of our experiments it is not obvious what input would give the advantage to *D]$_{Wd}$, since participants only ever encountered word-final obstruents in utterance-final position. Note also that this approach predicts domain generalization to every other phonological domain, since constraints such as, e.g., *D]$_{Phrase}$ and *D]$_{Syllable}$ also exist (Flack 2009). Given the implicational logic entertained above, utterance-final devoicing might be expected to generalize to *syllable*-final devoicing. Whether this is a good prediction we leave as an open question.

We suspect that the tendency toward domain generalization is a manifestation of a more general bias toward *word*-based phonology. There are certainly cases of phrase-level phonology, but the environments of most of the restrictions on the distribution of sounds are limited to within words. The word is, for example, the most common domain for patterns of stress (Hayes 1995), and vowel harmony almost never extends beyond a single word (a striking rare exception being Somali (Andrzejewski 1955)). Alternations are typically reflected within words in word paradigms, which is why phonologists usually begin a phonological inquiry with such paradigms (Kenstowicz & Kisseberth 1979). Again, however, if there is a general bias toward word-based phonology, where does it come from?

Assuming a model of acquisition in which phonotactic generalizations are built from a store of lexical representations (Pierrehumbert 2003; Edwards et al. 2004), learners are more likely to induce generalizations over words than over larger units like phrases or utterances, because they have more stored words than stored phrases over which to induce generalizations. This is because we encounter many more words than utterances (since words make up utterances). Words may also be more successfully stored, because they tend to be shorter than utterances, and because a given word is reinforced in memory more often by repeated exposure than a given utterance. In this view, domain generalization is simply the consequence of inferring the generalizations that are most robustly supported.

[6] As the same reviewer points out, if cast within Harmonic Grammar (Legendre et al. 1990a, b), in which the harmony of a candidate form can depend on the additive effect of *both* *D]$_{Wd}$ and *D]$_{Utt}$, this account also provides a possible explanation for the fact that our participants did not learn final devoicing as well in utterance-medial position as in utterance-final position – devoiced variants are supported by *both* constraints in utterance-final position. However, in the discussion of Experiment 1 we noted other possible explanations for this difference.

If it is correct that domain generalization happens because learners make generalizations over their mental lexicon, then we would not expect generalization from utterance-final position to domain-final position in domains that are larger than the word, such as phonological phrases. Since phrases consist of words as utterances do, the logic described above will still favor generalization to word-final position. As far as we know, there are no established cases of such generalization, but this should be investigated further. This account also does not imply that generalization should lead to *syllable*-final devoicing, since stored words with final devoicing do not by themselves provide any evidence about syllables. (We do not assume, as in account above, a theory that forces learners to entertain a constraint like *D]$_{Syll}$ simply because they entertain *D]$_{Wd}$.) These are predictions that can be tested in further experiments.

A question any account of domain generalization must address concerns the existence of direct counterevidence to domain generalization. In our experiments, learners had no evidence about word-final obstruent voicing in utterance-medial position. In real life, if a language has utterance-final devoicing, then the learner's input contains word-final voiced obstruents in non-utterance-final position. Why does domain generalization occur despite such counterevidence?[7]

There is reason to believe that utterance-final words, including words produced in isolation, have a special status for language learners. Infants learn to segment words from phrases at utterance edges before they can do so utterance-medially (Seidl & Johnson 2006; Johnson et al. 2014). Johnson et al. refer to utterance edges as 'anchors of reliability' for infants trying to segment speech. Brent and Siskind (2001) show (for English speakers around 9 months of age) that infants' likelihood of producing specific words depends on the frequency with which their caregivers uttered the same words earlier – but only when they uttered them in isolation. Sundara et al. (2011) demonstrate that infants (at around two years of age) both perceive and produce the English third person singular suffix *–s* better in utterance-final than in utterance-medial position, and they further argue that the production difference must be at least partly explained by the perceptual difference. Sundara et al. suggest three reasons why utterance-final forms may be better perceived. First, the added length of sounds in final position due to final lengthening may render these sounds more perceptible. Second, sounds followed immediately by more sounds can be masked. Third, sounds in utterance-final position are recalled more accurately and often. (Sundara et al. note that there is also much evidence in the literature for the salience of utterance-*initial* words.) The relevance of these findings here is that they imply that utterance-final word variants could figure more prominently in early lexicons because they are more accurately perceived and more frequently experienced (because they can be segmented from utterances).

Finally, we mention a few other implications of the domain generalization idea. If domain generalization is driven by a general bias for word-based generalizations over sound categories, then we do not expect it to be limited to word-final devoicing. We predict that other cases of domain generalization will work the same way, so that, for example, if learners were to be exposed to an utterance-final restriction on tone or vowel length, these too would be generalized

[7] We might also ask whether domain generalization ever goes in the opposite direction – from utterance-medial to utterance-final position. If this happened in a language with utterance-final devoicing, it would simply wipe out the devoicing pattern, leaving no trace. But the following discussion gives reason to doubt such 'outward' domain generalization.

to word-final position (see discussion in the Introduction). Moreover, if the basis of domain generalization is a general preference for word domains, then we expect it to apply if the restriction to be learned is on sounds at the *onset* of the utterance, rather than the offset. That is, domain generalization should be equally relevant to word-*initial* generalizations, as in fact argued for by Hualde (2013). More generally, a restriction on the distribution of sounds within words (e.g. one conditioned by affixes) should be more easily learned, all else being equal, than an equivalent restriction on the distribution of sounds within phrases (e.g. one conditioned by a neighboring word). It would require further experiments to determine whether these expectations are supported.

**References**


Andrzejewski, B. (1955). The problem of vowel representation in the Isaaq dialect of Somali. Bulletin of the School of Oriental and African Studies 17, 567-580.

Baayen, H. R., D. J. Davison & D. M. Bates (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of memory and language 59, 390-412.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tilly (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. Journal of memory and language 68, 255-278.

Bates, Douglas, Martin Maechler & Ben Bolker (2012). lme4: linear mixed-effects models using S4 classes. (R package version 0.999999-0.). http://CRAN.R-project.org/package=lme4 (accessed 9 July 2103).

Becker, L. (1979). A contribution to an explanation of the Neo-Stokavian accent retraction. Zbornik za filologiju i linguistiku 22, 87-94.

Becker, Lee (1977). Perceptually motivated phonetic change. In Beach, W., S. Fox & S. Philosoph (eds.) Papers from the Thirteenth Regional Meeting of the Chicago Linguistic Society. Chicago: Chicago Linguistic Society, 45-57.

Benua, Laura (1995). Identity effects in morphological truncation. In Beckman, J., S. Urbanczyk & L. Walsh (eds.) Papers in Optimality Theory. Amherst, MA: GLSA, 77-136.

Blevins, Juliette (2006). A theoretical synopsis of Evolutionary Phonology. Theoretical linguistics 32.2, 117-166.

Bloomfield, Leonard (1933). Language. Chicago: University of Chicago Press.

Boersma, Paul (1997). How we learn variation, optionality, and probability. Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam 21, 43-58.

Boersma, Paul & Bruce Hayes (2001). Empirical tests of the gradual learning algorithm. Linguistic inquiry 32, 45-86.

Braine, M. (1963). On learning the grammatical order of words. Psychological review 70, 323-348.

Brent, Michael R. & Jeffrey Mark Siskind (2001). The role of exposure to isolated words in early vocabulary development. Cognition 81, B33-B34.

Carpenter, Angela C. (2010). A naturalness bias in learning stress. Phonology 27, 345-392.

Chafe, Wallace (1959). Internal reconstruction in Seneca. Language 35, 477-495.

Coetzee, Andries (2009). Learning lexical indexation. Phonology 26, 109-145.

Cutler, A. & D. Norris (1988). The role of strong syllables in segmentation for lexical access. Journal of experimental psychology: human perception and performance 14, 113-121.

Docherty, G. (1992). The timing of voicing in British English obstruents. Berlin: Foris.

Edwards, Jan, Mary Beckman & Benjamin Munson (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. Journal of speech, language, and hearing research 47, 421-436.
Ernestus, Mirjam & Harald Baayen (2007). Intraparadigmatic effects on the perception of voice. In van der Weijer, J. & E. van der Torre (eds.) Voicing in Dutch. Amsterdam: John Benjamins, 153-174.
Ewert, Alfred (1943). The French language. London: Faber.
Finley, Sara & William Badecker (2009). Artificial language learning and feature-based generalization. Journal of memory and language 61, 423-467.
Flack, Katherine (2009). Constraints on onsets and codas of words and phrases. Phonology 26, 269-302.
Gafos, Adamantios (2006). Dynamics in grammar: comment on Ladd and Ernestus and Baayen. In Goldstein, L., D. H. Whalen & C. T. Best (eds.) Laboratory phonology 8. Berlin: Mouton de Gruyter, 51-79.
Gómez, R. (2002). Variability and detection of invariant structure. Psychol Sci 13, 431-436.
Haggard, M. (1978). The devoicing of voiced fricatives. Journal of phonetics 6, 95-102.
Hansen, K. C. & L. E. Hansen (1969). Pintupi phonology. Oceanic Linguistics 8, 153-170.
Hayes, Bruce (1995). Metrical Stress Theory: Principles and Case Studies. Chicago: The University of Chicago Press.
Herman, Rebecca (1996). Final lowering in Kipare. Phonology 13, 171-196.
Hock, Hans Heinrich (1991). Principles of historical linguistics. New York: Mouton de Gruyter.
Hualde, José Ignacio (2013). Intervocalic lenition and word-boundary effects: evidence from Judeo-Spanish. Diachronica 30.2, 223-266.
Hyman, Larry (1978a). Historical tonology. In Fromkin, V. (ed.) Tone: A Linguistic Survey. New York: Academic Press, 257-269.
Hyman, Larry (1978b). Word demarcation. In Greenberg, J. (ed.) Universals of human language, volume 2: phonology. Stanford: Stanford University Press, 443-470.
Hyman, Larry (2013). Enlarging the scope of phonologization. In Yu, A. (ed.) Origins of sound change: approaches to phonologization. Oxford: Oxford University Press, 3-28.
Hyman, Larry M. (1976). Phonologization. In Juilland, A. (ed.) Linguistic Studies Offered to Joseph Greenberg on the Occasion of his Sixtieth Birthday. Saratoga, CA: Anma Libri, 407-419.
Ingram, David (1989). Phonological disability in children. London: Whurr.
Iverson, Gregory & J. Salmons (2007). Domains and directionality in the evolution of German final fortition. Phonology 24, 121-145.
Jassem, Wiktor & Lutoslawa Richter (1989). Neutralization of voicing in Polish obstruents. Journal of phonetics 17, 317-325.
Jespersen, Otto (1926). Lehrbuch der Phonetik. Leipzig: B. G. Teubner.
Jessen, Michael (1998). Phonetics and phonology of tense and lax obstruents in German. Amsterdam: John Benjamins.
Jessen, Michael & Catherine Ringen (2002). Laryngeal features in German. Phonology 19.2, 189-218.
Johnson, Elizabeth K., Amanda Seidl & Michael D. Tyler (2014). The edge factor in early word segmentation: utterance-level prosody enables word form extraction by 6-month-olds. PLoS One 9.1: e83546.

Kapatsinski, Vsevolod (2009). Testing theories of linguistic constituency with configural learning: the case of the English syllable. Language 85, 248-277.
Kenstowicz, Michael & Charles Kisseberth (1979). Generative Phonology: Description and Theory. New York: Academic Press.
Kerkhoff, A. (2004). Acquisition of voicing alternations. In Baauw, S. & J. van Kampen (eds.) Proceedings of GALA 2003, Vol. 2. Utrecht: LOT, 269-280.
Klatt, D. H. & L. Klatt (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. Journal of the Acoustical Society of America 87, 820-857.
Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky (1990a). Harmonic grammar - a formal multi-level connectionist theory of linguistic wellformedness: an application. Proceedings of the Twelfth Annual Conference of the Cognitive Science Society. Cambridge, MA: Lawrence Erlbaum, 884-891.
Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky (1990b). Harmonic grammar - a formal multi-level connectionist theory of linguistic wellformedness: theoretical foundations. Proceedings of the Twelfth Annual Conference of the Cognitive Science Society. Cambridge, MA: Lawrence Erlbaum, 388-395.
Lehtonen, J. (1970). Aspects of quantity in Standard Finnish. Jyväskylä University Press.
Liberman, Mark & Janet Pierrehumbert (1984). Intonational invariance under changes in pitch range and length. In Aronoff, M. & R. T. Oehrle (eds.) Language sound structure. Cambridge, MA: MIT Press, 157-233.
Lightner, Theodore M. (1972). Problems in the Theory of Phonology, Vol 1: Russian Phonology and Turkish Phonology. Edmonton: Linguistic Research, Inc.
Lindblom, B. (1983). Economy of speech gestures. In MacNeilage, P. (ed.) The Production of Speech. New York: Springer-Verlag, 217-245.
Lisker, Leigh, Arthur Abramson, F. Cooper & M. Schvey (1969). Transillumination of the larynx in running speech. Journal of the Acoustical Society of America 45, 1544-1546.
Lombardi, Linda (1995). Laryngeal neutralization and syllable wellformedness. Natural language and linguistic theory 13.1, 39-74.
McCarthy, John J. (2005). Optimal Paradigms. In Downing, L. J., T. A. Hall & R. Raffelsiesen (eds.) Paradigms in phonological theory. Oxford: Oxford University Press.
McCarthy, John J. & Alan Prince (1995). Faithfulness and Reduplicative Identity. In Beckman, J., L. Walsh Dickey & S. Urbanczyk (eds.) University of Massachusetts Occasional Papers in Linguistics 18. Amherst, MA: GLSA Publications, 249-384.
Miller, George (1958). Free recall of redundant strings of letters. Psychological review 56, 485-491.
Moreton, Elliott (2008). Analytic bias and phonological typology. Phonology 25, 83-127.
Moreton, Elliott (2009). Underphonologization and modularity bias. In Parker, S. (ed.) Phonological argumentation: essays on evidence and motivation. London: Equinox, 79-101.
Moreton, Elliott & Joe Pater (2012a). Structure and substance in artificial-phonology learning, part 1: structure. Language and linguistics compass 6.11, 686-701.
Moreton, Elliott & Joe Pater (2012b). Structure and substance in artificial-phonology learning, part 2: substance. Language and linguistics compass 702-718, 686-701.

Morgan, J., R. Meier & Elissa Newport (1987). Structural packaging in the input to language learning: contributions of prosodic and morphological marking of phrases to the acquisition of language. Cognitive psychology 19, 498-550.
Myers, Scott (1999). Tone association and $f_0$ timing in Chichewa. Studies in African linguistics 28, 215-239.
Myers, Scott (2005). Vowel duration and neutralization of vowel length contrasts in Kinyarwanda. Journal of phonetics 33, 427-446.
Myers, Scott (2012). Final devoicing: production and perception studies. In Borowsky, T., S. Kawahara, T. Shinya & M. Sugahara (eds.) Prosody matters: essays in honor of Elisabeth Selkirk. London: Equinox Publishing.
Myers, Scott & B. Hansen (2007). The origin of vowel length neutralization in final position: evidence from Finnish speakers. Natural language and linguistic theory 25, 157-193.
Ohala, John J. (1993). The phonetics of sound change. In Jones, C. (ed.) Historical Linguistics: Problems and Perspectives. London: Longman Academic, 237-278.
Onishi, K., K. Chambers & C. Fisher (2002). Learning phonotactic constraints from brief auditory experience. Cognition 83, B13-B23.
Padgett, Jaye (2012). The role of prosody in Russian voicing. In Borowsky, T., S. Kawahara, T. Shinya & M. Sugahara (eds.) Prosody matters: essays in honor of Lisa Selkirk. London: Equinox, 181-207.
Peperkamp, Sharon, K. Skoruppa & Emmanuel Dupoux (2006). The role of phonetic naturalness in phonological rule acquisition. In Bamman, D., T. Magnitskaia & C. Zaller (eds.) Proceedings of the 30th Annual Boston University Conference on Language Development. Somerville, MA: Cascadilla Press, 464-475.
Pierrehumbert, Janet (2003). Probabilistic phonology: discrimination and robustness. In Bod, R., J. Hay & S. Jannedy (eds.) Probabilistic linguistics. Cambridge, MA: MIT Press, 177-228.
Pierrehumbert, Janet & Mary Beckman (1988). Japanese Tone Structure. Cambridge, MA: MIT Press.
Port, Robert F. & Michael L. O'Dell (1985). Neutralization of syllable-final voicing in German. Journal of phonetics 13, 455-471.
Pothos, E. & T. Bailey (2000). The role of similarity in artificial grammar learning. Journal of experimental psychology: learning, memory, and cognition 26, 847-862.
Prieto, Pilar, J. van Santen & Julia Hirschberg (1995). Tonal alignment patterns in Spanish. Journal of phonetics 23, 429-451.
Prince, Alan & Paul Smolensky (1993 [2004]). Optimality theory: constraint interaction in generative grammar. Ms., Rutgers University, New Brunswick, NJ, and University of Colorado, Boulder [Published by Blackwell, 2004].
Pycha, Anne, Pawel Nowak, Eurie Shin & Ryan Shosted (2003). Phonological rule-learning and its implications for a theory of vowel harmony. In Garding, G. & M. Tsujimura (eds.) Proceedings of the West Coast Conference on Formal Linguistics 22. Somerville, MA: Cascadilla Press, 423-435.
Pycha, Anne, Eurie Shin & Ryan Shosted (2006). Directionality of assimilation in consonant clusters: an experimental approach. UC Berkeley Phonology Lab annual report, 152-166.
Reber, Arthur S. (1967). Implicit learning of artificial grammars. Journal of verbal learning and verbal behavior 6, 855-863.

Schane, Sanford, Bernard Tranel & H. Lane (1975). On the psychological reality of a natural rule of syllable structure. Cognition 3, 351-358.
Seidl, Amanda & Eugene Buckley (2005). On the learning of arbitrary phonological rules. Language learning and development 1, 289-316.
Seidl, Amanda & Elizabeth K. Johnson (2006). Infant word segmental revisited: edge alignment facilitates target extraction. Developmental science 9.6, 565-573.
Shadle, C. (1997). The aerodynamics of speech. In Hardcastle, W. & J. Laver (eds.) The handbook of phonetic sciences. Oxford: Blackwell, 33-64.
Sievers, E. (1901). Grundzüge der Phonetik, 5th edition. Leipzig: Breitkopf and Hartel.
Silverman, K. & Janet Pierrehumbert (1990). The timing of prenuclear high accents in English. In Kingston, J. & M. Beckman (eds.) Papers in laboratory phonology I: between the grammar and physics of speech. Cambridge: Cambridge University Press, 72-106.
Slifka, J. (2006). Some physiological correlates to regular and irregular phonation at the end of an utterance. Journal of voice 20, 171-186.
Smith, C. (1997). The devoicing of /z/ in American English: effects of local and prosodic context. Journal of phonetics 25, 471-500.
Smith, C. (1999). Marking the boundary: utterance-final prosody in French questions and statements. In Ohala, J., Y. Hasegawa, M. Ohala, D. Granville & A. Bailey (eds.) Proceedings of the XIVth International Congress of Phonetic Sciences. Berkeley: Linguistics Department, University of California, Berkeley, 1181-1184.
Smith, C. (2003). Vowel devoicing in contemporary French. French language studies 13, 177-194.
Steriade, Donca (1997). Phonetics in phonology: the case of laryngeal neutralization. Ms., UCLA.
Steriade, Donca (2000). Paradigm uniformity and the phonetics-phonology boundary. In Broe, M. B. & J. B. Pierrehumbert (eds.) Papers in laboratory phonology V: acquisition and the lexicon. Cambridge: Cambridge University Press, 313-334.
Sundara, Megha, Katherine Demuth & Patricia K. Kuhl (2011). Sentence-position effects on children's perception and production of English third person singular -s. Journal of speech, language, and hearing research 54, 55-71.
Sweet, Henry (1877). Handbook of phonetics. Oxford: Oxford University Press.
Tessier, Anne-Michelle (2006). Testing for O-O faithfulness in artificial phonological acquisition. In Bamman, D., T. Magnitskaia & C. Zaller (eds.) Proceedings of the 30th Annual Boston University Conference on Language Development. Somerville, MA: Cascadilla Press, 619-630.
Vennemann, Theo (1974). Words and syllables in Natural Generative Phonology. In Bruck, A., R. Fox & M. La Galy (eds.) Papers from the Parasession on Natural Phonology. Chicago: Chicago Linguistic Society, 346-374.
Wackernagel, J. (1957). Altindische grammatik. Band I: lautlehre (reprinted from the 1896 edition). Göttingen: Vandenhoeck and Ruprecht.
Warner, Natasha, Allard Jongman, Joan Sereno & Rachèl Kemps (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. Journal of phonetics 32, 251-276.
Westbury, John R. & Patricia A. Keating (1986). On the naturalness of stop consonant voicing. Journal of linguistics 22, 145-166.

White, James (2013). Evidence for a learning bias against saltatory phonological alternations. Cognition 130, 96-115.
Wilson, Colin (2003). Experimental investigation of phonological naturalness. In Garding, G. & M. Tsujimura (eds.) Proceedings of WCCFL 22. Somerville, MA: Cascadilla Press, 101-114.
Wilson, Colin (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. Cognitive science 30.5, 945-982.
Yu, Alan C. L. (2004). Explaining final obstruent voicing in Lezgian: phonetics and history. Language 80.1, 72-97.