UNIVERSITY OF CALIFORNIA

SANTA CRUZ

THE SYLLABLE IN DOMAIN GENERALIZATION: EVIDENCE FROM ARTIFICIAL LANGUAGE LEARNING

A thesis submitted in partial satisfaction
of the requirements for the degree of

MASTER OF ARTS

in

LINGUISTICS

by

Maya C. Wax Cavallaro

September 2023

The Thesis of Maya C. Wax Cavallaro
is approved:

Professor Jaye Padgett, Chair

Professor Rachel Walker

Associate Professor Grant McGuire

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

THE SYLLABLE IN DOMAIN GENERALIZATION: EVIDENCE FROM ARTIFICIAL LANGUAGE LEARNING

by

Maya C. Wax Cavallaro

Domain generalization is an account of certain word-final phonological phenomena, such as devoicing, in which they originate as phrase-final patterns that become phonologized and generalized by learners from the phrase level to the word level. Myers & Padgett (2014) tested this theory empirically through two artificial language learning experiments. They show that participants can learn an utterance-final obstruent devoicing pattern, given relatively short exposure, and generalize it to a word-level final devoicing rule.

Building upon Myers & Padgett (2014), this thesis explores whether the same type of domain generalization can account for syllable-final phenomena, or whether there is something special about the word that makes it the destination of generalization. Three artificial language learning experiments are presented, the third of which shows generalization from the word/utterance level to the syllable. Though syllables may differ from words in important ways, evidence from this study suggests that the syllable is accessible to the grammar for learning, generalization, and derivation of phonological rules.

Acknowledgments

I am immensely grateful to Jaye Padgett for sharing his research and insight and guiding me through this project, Ryan Bennett, Grant McGuire, and Rachel Walker for their ongoing support and mentorship, Amanda Rysling, 2022 LING 290 participants, and all my classmates and colleagues who have heard me discuss this research at many stages and provided fantastic questions and feedback. Special thanks to Mykel Brinkerhoff, Dan Brodkin, Jason Ostrove, and Adrian Tiscareno, for their help with recording stimuli, to Jacob Pollock for insight on statistics and more, and to the MeloMelo Kava Bar, where I wrote the majority of the following text while sipping kombucha or yerba mate. I am also eternally grateful to my Tz'utujil friends and consultants who have shared their language and so much more with me and inspired much of my work. Maltyoox!

I would also like to acknowledge my family, who have always supported me and my academic and professional endeavors. I dedicate this work to the memory of my brilliant grandmother, Lenore Wax.

1. Introduction

Domain generalization is a term coined by Myers & Padgett (2014, from here on M&P) for a fairly widespread theory (e.g. Hyman 1978, Westbury & Keating 1986, Hock 1991, Barnes 2006, Padgett 2015) of the diachronic origins of certain domain-edge phenomena, such as word-final obstruent devoicing. Word-final devoicing is one example of a phonological pattern that makes reference to word edges. It is attested in a variety of languages and language families, including, for example, Catalan, German, Hausa, Russian, Sanskrit, Turkish, and Walloon (Blevins 2006, Myers 2012). While devoicing does not seem to have any phonetic grounding or motivation in word-final position, this phenomenon is often attributed to phonetic tendencies at the ends of longer phrases or utterances.[1]

In phrase-final position, phonetic pressures, such as a decline in the subglottal air pressure necessary for voicing over the course of an utterance or a spreading of the vocal folds in anticipation of non-speech breathing may cause variable final voicelessness (Ohala 1983, Blevins 2006). A domain-generalization-type account posits that this phonetic tendency can become phonologized and generalized from the phrase domain to the word domain. In this case, learners may interpret a tendency toward final voicelessness as a categorical phonological restriction. They then generalize from a ban on final voiced obstruents in utterance-final words to a ban on final voiced obstruents in all words.

[1] In this paper, 'phrase' and 'utterance' will be used interchangeably.

M&P tested this type of domain generalization empirically, through two artificial learning experiments. They show that, even given relatively short exposure, participants can learn an utterance-final restriction and generalize utterance-final devoicing to a word-level rule, applying the rule to utterance-medial words. While M&P and others (e.g. Blevins 2006) claim that phonological rules can be generalized from the phrase/utterance domain to word edges, they tend to leave for future work the question of whether there is something special about words that allows for this generalization.

M&P specifically acknowledge the possibility that utterance-final devoicing might be expected to generalize to syllable-final devoicing, though, "whether this is a good prediction [they] leave as an open question" (427). They also raise the possibility of a general preference for word-based phonology, citing patterns such as stress and vowel harmony that are typically associated with the word domain, rather than larger domains like the phrase.

A number of phonological phenomena, however, seem sensitive to syllable structure as well. In many languages, the consonants which may occur in coda position are restricted to a small subset. In languages such as Mandarin, for example, only nasal consonants may occur in codas (Faytak et al. 2020). Syllable codas are often where contrast neutralizations occur. Different types of syllable-final laryngeal neutralization patterns—both diachronic and synchronic—are found in languages such as German, Korean, Maidu, and Thai (Lombardi 1995). "Lenition" patterns, such as /s/-debuccalization in many varieties of Spanish, shown in (1) may also apply to syllable—rather than just word—edges.

(1) Spanish /s/-debuccalization

	orthography	non-debuccalizing	debuccalizing varieties
a.	español	[esˈpaɲol]	[e**h**ˈpaɲol]
b.	sostener	[sosˈtener]	[**soh**ˈtener]
c.	feliz cumpleaños	[feˈlis kumpleˈaɲos]	[feˈli**h** kumpleˈaɲo**h**]

In these varieties, the onset /s/ in (1b) does not debuccalize, but debuccalization occurs in all codas, whether word-internal or word-final.

This paper explores whether syllable-edge phenomena can, like word-edge phenomena, result from domain generalization. Through a series of artificial language experiments, I show that the syllable can, indeed, be a domain of generalization. In the next section, I discuss potential similarities and differences between the word domain and the syllable domain. I then discuss M&P's work briefly before presenting three experiments. Experiment 1 replicated M&P-style utterance-to-word generalization. Experiment 2 failed to show generalization to the syllable, but it differed from Experiment 1 in many crucial ways. Experiment 3 tested for generalization to the word and generalization to the syllable using identical materials and found both types of generalization. In the final section, I discuss implications of these findings and potential directions for future work.

2. Generalization beyond the word

If generalization by analogy from larger prosodic domains (the phrase) to the word level does, in fact, occur, it seems that generalization beyond the word to smaller domains should also be possible. This is predicted by Blevins (2006) in an evolutionary account of phonological patterns, which states, “For sound changes originating with phrase-final laryngeal gestures and lengthening…the direction of final-devoicing sound patterns is predicted to be utterance > phrase > word > syllable” (140). Blevins gives examples of generalization to the word level, but not to the syllable.

Zec & Zsiga (2022) do, however, provide an example of generalization beyond the word to a smaller domain. They cite domain generalization as an account of patterns of High tone retraction in Štokavian varieties of Serbian. They claim that High tone retraction is motivated at the phrase level, in order to avoid tone crowding due to a phrase-final boundary tone. This phrase-edge process, they explain, has been generalized to word-final position and then, in some varieties, to foot-final position (80).

As discussed briefly in Section 1, many cross-linguistic phonological processes seem to apply to the syllable domain. Though German is often cited as a case of word-final obstruent devoicing, it has been claimed that word-internal, syllable-final obstruent devoicing occurs in both German and Dutch (e.g. Wetzels & Mascaró 2001).

A clear example of syllable-level patterns comes from Tz’utujil, a K’ichean-branch Mayan language spoken in the highlands of Guatemala,[2] which presents a typologically rare case of final devoicing (Wax Cavallaro 2021, 2022). While

[2] This region is also known as *Ixim Ulew*, or Land of the Corn.

Tz’utujil does not utilize contrastive voicing in its phonemic inventory,[3] it does devoice approximants syllable-finally.

(2) Approximant devoicing in Tz’utujil (Dayley 1985)

a.	*way*	[wa**j̊**]	‘tortilla’
b.	*Moysees*	[mɔ**j̊**seːs]	‘Moses’
c.	*kow*	[kɔ**ɸ**]	‘hard'
d.	*tewlaj*	[tɛ**ɸ**lax]	‘very cold'
e.	*uleew*	[uleː**ɸ**]	‘land’
f.	*jul*	[xʊ**l̥**]	‘hole’
g.	*elnaq*	[ɛ**l̥**naqʰ]	‘he has left’
h.	*q’or*	[qˀɔ**r̥**]	‘lazy’
i.	*warnaq*	[wa**r̥**naqʰ]	‘he has gone to sleep’

Though this devoicing occurs word-internally, it is a positional (domain-final) phenomenon, rather than an assimilatory one. In examples such as (2d), (2g), and (2i), sonorant consonants devoice in codas while adjacent on both sides to voiced segments.

It is plausible that the same phonetic tendency towards phrase/utterance-final voicelessness typically cited as the motivation for word-final devoicing could also lead to the phonologization and generalization of final sonorant devoicing. If this is the case, however, Tz’utujil would show generalization beyond the word, to the syllable, as shown in (2b,d,g,i).[4]

---

[3] Tz’utujil contrasts a set of “simple” pulmonic stops and affricates (/p,t,k,q,t͡s,t͡ʃ/) with a glottalized set (/ɓ,ɗ,k’,q’,t͡s’,t͡ʃ’,ʔ/). The consonant inventory also includes voiceless fricatives (/s,ʃ,x/) and sonorants (/m,n,l,r,w,j/).

[4] According to Dayley (1985), nasal consonants devoice only partially and only word-finally in Tz’utujil. Some of my independently-collected data makes me question the presence and nature of nasal devoicing in Tz’utujil, so I will leave this for future work, but I note that this *could* be evidence of generalization first to the word level and then to the syllable level, where nasal devoicing did not generalize beyond the word level.

In order for learners to generalize from the phrase level to the word level and beyond, it seems they must somehow interpret a pattern applying to segments at the phrase edge as a pattern applying to a different, smaller domain (e.g. word or syllable edges). If a pattern like final devoicing originates at the phrase level, learners must hear examples of words with and without final devoicing at the early stages of the change. M&P suggest that learner-directed speech tends to consist largely of shorter, even single-word, utterances. Utterance-final tokens may be, in a sense, overrepresented in the learner's lexicon or memory. If a high enough proportion of words encountered are utterance-final, the restrictions/alternations applying to utterance-final words may be interpreted as patterns applying to all words.

This type of frequency-based calculation fits well within something like Exemplar Theory (Johnson 1996, Pierrehumbert 2000). Exemplar Theory proposes that learners store experienced tokens of language in memory with an amount of phonetic detail that captures a great deal of variability. Decisions in the identification and production of sounds, words, etc. can be based on calculations over these stored forms. An alternative to the stored exemplar approach would be something like the Gradual Learning Algorithm (Boersma 1997), where experienced tokens of language do affect and update a learner’s grammar (by promoting/demoting constraints) without being encoded in longterm memory. For purposes of this paper, I will remain as theory-neutral as possible, while still acknowledging that the grammar needs access to some unit over which to generalize.

If generalization to the syllable is possible, this requires information about syllable structure to be either stored (perhaps in separate exemplar clouds for syllables, or as part of the structure within stored words/utterances) or, minimally, accessible to the grammar for frequency calculations and derivation of phonological rules/patterns. In Section 6 below, I show experimental evidence that this could, indeed, be the case, but there are also differences between the word domain and the syllable domain that merit discussion.

2.1 Factors that may facilitate learning/generalization

In addition to the relative proportion of short/single-word utterances experienced by language learners, there are other reasons why utterance-final words (and, theoretically, syllables) may be disproportionately represented in learners' memory. A number of factors may make certain tokens more likely to be stored (or, depending on one's theory, more heavily-weighted) than others. These may include attention, recency, and even prominence. Pierrehumbert (2000) also suggests learner age, and social considerations, such as context/register and feelings of affiliation, as possible factors affecting the formation of exemplar memories. For the most part, this matter is beyond the scope of this paper, but it is an intended area for future work which affected the design of Experiment 2, so it is worth discussing briefly.

As discussed in M&P, utterance edges have been shown to have a special status for language learners. Sundara et al. (2011), for example, explore observations that young children are more likely to produce inflectional morphemes at utterance edges than utterance-internally. In a study of 22- and 27-month-old English-learning children, they show an effect of utterance position on both perception and production, demonstrating that children show earlier sensitivity to inflections at utterance edges. Johnson et al. (2014) similarly show that infants begin learning to segment words at utterance edges. “Clearly, utterance edges play a crucial role in infants’ early segmentation attempts,” they write, “and any model of early speech perception that does not incorporate this fact will present a distorted view of early word segmentation abilities” (13).

Sundara et al. (2011) point out that utterance-final lengthening allows learners slightly more time to perceive phonological information without subsequent sounds that can potentially mask that information. Utterance-final sounds are also the most recent sounds in an utterance, and therefore may be easier to recall accurately. While utterance-final position benefits from final lengthening and recency, utterance-initial position is associated with articulatory strengthening, which may also aid in perception and encoding.

In considering the typologically-rare case of Tz’utujil sonorant devoicing in (2), a possible factor to consider is that Tz’utujil has fixed final stress. If utterance-final syllables are already especially memorable to learners by nature of being at the utterance edge, perhaps the added prominence of lexical stress[5] gives them the additional salience needed to learn a more marked pattern of sonorant devoicing.

[5] Bennett (2016) points out that phonetic correlates of stress in Mayan languages are underdocumented. To my knowledge, there has been no acoustic/instrumental study of stress in Tz’utujil. Baird’s (2014) preliminary analysis of acoustic correlates of stress in closely-related K’iche’ found that F_0 (pitch accent) was the only consistently reliable correlate across different varieties, though vowel duration (quantitative accent) is a correlate of stress in varieties that do not have a phonemic vowel length distinction.

Another possibility is that the prominence of utterance-final syllables makes generalization to the syllable (rather than just to the word) more likely by increasing the likelihood of interpreting an utterance-final pattern as a syllable-level rule.

Pye (1983) shows an effect of stress on child language acquisition in children acquiring K'iche', a closely-related K'ichean language with fixed final stress and similar final devoicing. These children focus on final syllables, passing through a stage of production in which they produce only the final, stressed syllable of multisyllabic words. This means that when CVC verb roots are combined with VC suffixes and resyllabified, children will produce just the final segment of the root with the suffix, choosing perceptual salience over semantic complexity/informativity. These L1 learners pay special attention to stressed, word-final syllables. If child learners are paying special attention to the final, stressed syllables of words, it seems plausible that they might perceive utterance-final patterns as syllable-final patterns, rather than word-final patterns.

Experiment 2, described in Section 5 below, was designed, in part, to explore whether stress might affect learning and/or generalization, but it was not able to do so. I hope to follow this line of investigation in future work.

2.2 The syllable as different from the word

Words exist as both morphosyntactic and phonological/prosodic units, though prosodic word and morphosyntactic word boundaries do not always align perfectly (see e.g. Ito & Mester 2009). The syllable, on the other hand, is a purely phonological domain. This presents a potential point of departure in two different ways.

The first has to do with the structures over which frequency-type calculations can be made. In order to generalize over words or syllables, the grammar must have access to those units. The fact that words are somehow stored in the mind is relatively uncontroversial. Whether one believes in an entirely abstract mental lexicon, a complex map of exemplar clouds, or a combination or variation on these types of representations, words are a unit that is stored in memory and accessible to the grammar.

The status of the syllable, on the other hand, is more contentious. Many phonologists have argued that syllable structure is not part of phonological representation. A great deal of work in Optimality Theory (Prince & Smolensky 1993) claims that syllable structure is determined by phonological constraints, but not contained in underlying representations. Ohala & Kawasaki-Fukumori (1997) and Steriade (1999) propose that apparent syllable-level patterns can be inferred from phonetic factors such as coarticulation and licensing by cue, and that reference to the syllable is not necessary.

Others, however, defend the syllable. Vaux (2003), for example, argues for "a theory of the lexicon in which some, but not all, predictable information is stored." Providing evidence from Armenian, he claims some information about syllable structure must be present underlyingly. To assertions that syllables do not exist in the grammar, Vaux replies, "My response is that we do not yet fully understand the psychological mechanisms involved in accessing information stored in the brain, nor do we understand how these mechanisms interact with the performance abilities required to execute tasks such as marking parts of words or dividing them in parts…" (118).

M&P point out that, given an exemplar-type model where experienced tokens are stored in memory, there are more words than utterances available to the learner because all utterances consist of words. This may lead to a preference for word-level phonology over phrase-level phonology. The same logic could apply to syllables. All words are comprised of syllables (but not vice-versa). This means, theoretically, that the learner has access to more syllables than words, assuming that syllable structure is encoded within representations of words.

Another possibility is that, whether or not the syllable is accessible to the grammar, it is not a potential domain to which generalization by analogy may occur. In work on "the life cycle of phonological processes," Bermudez-Otero & Trousdale (2012) appeal to phonologization and domain narrowing in a way that is both quite similar and significantly different from that discussed in M&P. Work of Bermudez-Otero and colleagues (see also Bermudez-Otero & McMahon 2006) takes a stratal-cyclic approach based in Lexical Phonology to explain phonological changes such as post-nasal /g/-deletion in English.

While this approach does allow for—and even posits—generalization from the phrase level to the word level and beyond to a smaller domain, the domains in question are more morphosyntactic, rather than purely prosodic. Burmudez-Otero and colleagues propose that phonological rules, even those affecting syllable codas, may apply at the phrase stratum, the word stratum, or the morphological stem stratum.

According to this approach, a pattern like phrase-final devoicing would not generalize from phrase-final position to syllable-final position, though it might generalize to the morphological stem level.

Given evidence of generalization from the phrase level to the word level (M&P, Zec & Zsiga 2022, etc.) and the existence of syllable-level phenomena (including patterns like syllable-final devoicing in Dutch, German, and Tz'utujil, that could plausibly result from the generalization of phrase-level phonetic tendencies), it seems likely that learners may generalize beyond the word domain to the syllable level. At the same time, the fact that the syllable is a purely phonological domain means that syllables may be stored differently in memory or may not be accessible to the grammar in the same ways as words. To test whether this is the case, I present three experiments in the following sections.

3. Domain generalization in artificial language learning

Myers & Padgett's 2014 study on domain generalization (M&P) presents two artificial grammar learning (AGL, also known as miniature grammar learning or artificial language learning) experiments. The AGL paradigm allows researchers to control the structure of the language material to which learners are exposed in order to observe the learning that occurs. Participants are exposed to a constructed mini-language during a learning phase. The linguistic knowledge acquired during this phase is measured during a testing phase.

This type of experiment has been employed to observe both syntactic and phonological learning. While the timescale of the learning phase may vary from a few minutes to a few weeks, it is definitely a shorter exposure than in natural language learning. Despite the short duration of this type of experiment, as well as the artificial nature of the "language" and the learning task, AGL has been able to provide convincing insight into how humans acquire linguistic knowledge.

M&P utilized the *poverty of the stimulus* methodology in AGL (Wilson 2003, 2006, Finley & Badecker 2009, White 2013) to test whether learners would generalize from a limited speech sample. In this type of experiment, there are classes of stimuli present in the test phase that were not included during the learning phase, though they are related to classes present in the learning phase. To test for generalization of final obstruent devoicing, for example, M&P included both voiced and voiceless obstruent onsets in their stimuli during both phases of their experiments. Utterance-finally, however, only voiceless obstruents (in addition to vowels and nasals) but not voiced obstruents could occur. This was the pattern participants were meant to learn. Utterance-medially, however, target words ended only in vowels or nasals. This is the poverty of the stimulus: Participants were not given any evidence about obstruents in utterance-medial, word-final position during the learning phase.

During the testing phase, M&P presented participants with new stimuli and asked them to judge whether each item belonged to the language to which they had been exposed. This method proved effective, allowing M&P to show that participants could learn the utterance-final pattern and generalize it to novel, non-final words.

In their first experiment, they trained one group on final devoicing (only voiceless obstruents allowed utterance-finally) and a second group on a less-natural final voicing pattern (only voiced obstruents allowed utterance-finally). The only obstruents in the experiment were sibilant [s] and [z]. Both groups demonstrated evidence of learning and generalization, though the final devoicing group did so more successfully, showing an effect of naturalness.

To capture more aspects of natural language learning, M&P incorporated a larger set of obstruents ([p,b,t,d,k,g]), as well as morphologically conditioned alternations in their second experiment. They again found evidence of both learning and generalization. They also found somewhat weaker evidence that learners have an overall bias against accepting more than one form for a given item, but that such alternating forms may assist in the learning of patterns such as final devoicing.

The experiments presented in the following sections build upon this work by M&P. While there is still work to be done on naturalness and alternations, the focus of this study is on learning and generalization and whether learners can generalize to the syllable domain as they generalized to the word domain in M&P.

4. Experiment 1: Generalization to the word level

In order to build upon M&P, it was first important to replicate their findings with respect to learning and utterance-to-word generalization. After adjusting the experiment design in several ways, discussed below, I had to show that participants would still generalize to the word level before I could test their ability to generalize beyond the word to a smaller domain. Experiment 1 does, in fact, show evidence that participants learned an utterance-final pattern and generalized it to the word level, applying it to novel, utterance-medial words.

4.1 Differences from M&P Experiment 1

Experiment 1 is based very closely on Experiment 1 from M&P (M&P1), however there are some notable differences. First, rather than employing a devoicing-style neutralization, this experiment uses a liquid /r~l/ neutralization. As the goal was eventually, in future experiments, to explore patterns at the syllable level, the concern was that participants' L1 phonology would prevent them from generalizing final devoicing to syllable codas due to voicing assimilation. Participants would likely prefer voiceless codas immediately preceding voiceless consonants and voiced codas before adjacent voiced consonants for reasons unrelated to learning or generalization. Coda liquid neutralization is attested, for example, in varieties of Puerto Rican Spanish, where /l/ and /r/ neutralize to [l] syllable-finally. This attested, but relatively rare, neutralization involves sounds that are perceptible to native English speakers, but English does not have neutralization or assimilation phenomena involving liquid /l/ and /r/.

Because of the choice of neutralization, the consonant inventory for stimuli for this experiment ([t,k,m,n,l,r]) differs slightly from that of M&P1 ([p,t,s,z,m,n]). I also chose to use only "light" (non-velarized) [l] and trilled [r], as well as monophthong vowels and fully-released final stops so that the sounds would all be perceptibly distinguishable and the language would sound very non-English. This strategy was abandoned in later experiments, as discussed below.

Another future goal of this research was to investigate whether/how stress affects learning and/or generalization. While M&P used a combination of monosyllabic, disyllabic, and trisyllabic target words with word-initial stress, I chose to control for stress by using only monosyllabic target words in Experiment 1, where all relevant syllables are stressed (by nature of being the only syllable).

Finally, while M&P1 was run in a sound-attenuated booth in the UC Santa Cruz Phonetics Laboratory, using Superlab experiment presentation software (Cedrus, Version 4.5), the experiment described below was carried out online using PCIbex (Zehr & Schwarz 2018). This was mainly done due to safety concerns around the COVID-19 pandemic, though it also allowed for a more diverse participant pool. Rather than recruiting exclusively from UC Santa Cruz undergraduates, I was able to recruit from all over the United States using Prolific Academic (www.prolific.co).[6]

Of course, running the experiment online had some drawbacks. While participants were instructed to use quality, wired headphones, several reported that they did not. There was far less ability to control for the quality of the audio and participants' focus. There were also frequent technical issues. This has, however, become an accepted method of running certain types of linguistic experiments, especially during the pandemic.

[6] Participants ranged in age from 18-63 years old. About 60% of participants identified as white, and about 60% of participants identified as male. More complete demographic information provided in Appendix A.

4.2 Methods

4.2.1 Participants. As noted above, participants were recruited via Prolific Academic (www.prolific.co) and paid for their participation. Initially, 60 participants were recruited, 30 in each of two groups. After it was determined that data from several of these should be excluded, additional participants were added until both groups had 24 members.

Participants were prescreened by Prolific and identified as native speakers of English residing in the United States with no language-related disorders or hearing difficulties. Despite this, one participant self-reported in a debrief questionnaire that English was not their L1, so their responses were excluded. No participants reported any prior training in linguistics.

4.2.2 *Materials*. The stimuli were nonsense utterances consisting of CV(C) syllables, where C belonged to the set [t,k,m,n,l,r] and V belonged to the set [i,e,a,o,u]. As in M&P1, target words were presented in one of the two frames in (3), where the underline shows the position of the target word.

(3) a. *utterance-final* ˈkanta ___
b. *utterance-medial*[7] ˈkanta ___ miˈtuku

All target words were monosyllables of the shape CVC. All consonants occurred with equal frequency in onset position, while codas were divided into four classes: nasals, stops, [l], and [r], with equal numbers of words of each coda class. The target words

[7] Throughout this paper, *medial* is meant to be synonymous with *non-final*.

for the learning stage are given in (4a), and those for the testing stage are given in (4b).

(4) a. *Target words used during the learning phase of Experiment 1*

nasal-final	stop-final	[l]-final	[r]-final
kin	tit	kal	kar
kum	kut	kil	ker
lam	lek	lel	lar
men	lok	lul	lir
nan	lot	mal	mir
nin	mak	mol	mor
num	mik	nil	nor
ran	mit	nul	nur
rim	nut	rol	rer
rom	ret	rul	rur
tem	tak	tal	tor
ton	kok	tel	tur

b. *Target words used during the testing phase of Experiment 1*

nasal-final	stop-final	[l]-final	[r]-final
kan	kak	kel	kir
kim	kek	kol	kor
kun	kot	kul	kur
lim	let	lal	ler
lom	luk	lil	lor
lon	mat	lol	lur
lum	mut	mel	mar
mam	nat	mil	mer
mim	net	mul	mur
mum	nik	nal	nar
mun	nit	nel	ner
non	nok	nol	nir
ren	rik	ral	rar
ron	rok	rel	rir
run	rot	ril	ror
tan	tet	til	tar
tom	tik	tol	ter
tun	tut	tul	tir

Each target word in (4) was recorded in each of the two utterance frames in (4) by a graduate student with extensive linguistic training who is a native speaker of

American English. Pronunciations were intentionally not-English-like, with non-English vowels [e,o,a], “light” [l] in both onset and coda position, and (dramatically) trilled [r].

Two learning sets were created from these recordings, for two different groups of participants. I call these groups L-to-R and R-to-L because [l] is intended to neutralize to [r] in one group while [r] neutralizes to [l] in the other. Both groups were presented with all of the nasal-final and stop-final target words in (4a) in both utterance frames. This was intended to demonstrate that coda consonants are allowed both utterance medially and utterance finally. The L-to-R group only heard the [r]-final (but not the [l]-final) target words during the learning phase, and they were only presented with those words in the utterance-final frame (3a). This is the poverty of the stimulus design described in Section 3 above. The R-to-L group heard only the [l]-final (but not the [r]-final) target words during the learning phase, also only in utterance-final position.

4.2.3 Procedure. Materials including text instructions were presented to participants online on their own personal computers using PCIbex (Zehr & Schwarz 2018). Participants were instructed to use quality, wired headphones and to locate themselves in a setting in which they could hear clearly and focus for the duration of the activity (about 30-45 minutes).

During the learning phase, participants were told they would be listening to sentences in a made-up language. They were instructed to repeat each sentence aloud to get a feel for how the language sounds. This phase was self-paced—participants clicked a button on their screen to hear the next sentence, though they could only hear each utterance once per trial. Stimuli were presented in blocks by sentence position/frame, with target words presented in the utterance-final (3a) frame first and utterance-medial (3b) frame second. Each block was repeated three times, with the sentences presented in a different randomized order each time. The learning phase comprised 180 total trials (three presentations of the 60-utterance learning set). Participants were given the opportunity to take a short break between blocks and between the learning phase and the testing phase.

Participants in the L-to-R group were trained on the L-to-R learning set, meaning they heard utterance-final [r] but not utterance-final [l]. Participants in the R-to-L group were trained on the R-to-L learning set, meaning they heard utterance-final [l] but not utterance-final [r]. The other items in the two learning sets were identical.

During the testing phase, both groups were presented with the same set of stimuli: All of the target words in (4b) presented in each of the two frames in (3). Rather than repeat utterances aloud, as in the learning phase, participants were instructed to listen to each sentence and answer the question presented on the screen: *Does this sentence sound like it belongs to the language you have been learning?*

To answer 'yes' (i.e. to *accept* the item), participants could click a button on their screen or press the 'd' key on their keyboard. To answer 'no' (i.e. to *reject* the item), participants could click a button or press the 'k' key on their keyboard. The 'yes' button was always presented on the left, and the 'no' button on the right. As in the learning phase, this phase was self-paced. Participants could hear each utterance only once, and the next utterance was presented after they selected ‘yes’ or ‘no’.

Again, utterances were presented in blocks by frame. Following M&P, this was done so that participants would compare items within each frame, to prevent them from making their acceptability judgments based on the position of the target word in the sentence. Utterances in the medial frame (3b) were less frequent in the learning set (because only nasal-final and stop-final—but not liquid-final—target words were presented in this frame), so there was concern that participants would perceive the medial frame as less well-formed than the final frame on the basis of relative frequency.

Each block was presented one time during this phase, with the utterance-final frame presented first. The order of the utterances within the block was randomized. Participants were given the opportunity to take a short break between the two blocks.

After the testing phase, participants were given a debriefing questionnaire, which asked what they believed the experiment was investigating and what strategy they used to decide which sentences to accept/reject. The majority of participants did not mention anything like a restriction on final liquids, suggesting that learning was relatively implicit. Four participants did consciously and explicitly figure out the pattern. Their data was excluded from the analysis, as described in the results section below.

4.2.4 Hypotheses. *Like M&P1, this experiment tested three hypotheses.*

Hypothesis 1: Learning. If the participants in this experiment learn the pattern they are exposed to during the learning phase, their 'accept' responses should reflect the pattern in the testing phase. This means that participants in the L-to-R group (who heard [r] but not [l] utterance-finally during the learning phase) will accept test sentences with utterance-final [r] more frequently than sentences with utterance-final [l]. The R-to-L group (who heard [l] but not [r] utterance-finally) will do the opposite, accepting utterance-final [l] more often than utterance-final [r]. This type of pattern would be evidence of an effect of the learning set on participants' responses during the testing phase.

Hypothesis 2: Generalization. The learning phase does not provide participants with evidence about which liquids are allowed in utterance-medial, word-final position. If participants generalize the pattern they learned from utterance-final position to word-final position, then the pattern should hold for test sentences with the utterance-medial frame. This means the L-to-R group will accept word-final, utterance-medial [r] more frequently than they accept word-final, utterance-medial [l], and the R-to-L group will accept word-final, utterance-medial [l] more frequently than they accept word-final, utterance-medial [r]. This would show generalization of the learned pattern to the smaller, word-level domain.

Hypothesis 3: 'Naturalness'. If either [r] → [l] neutralization or [l] → [r] neutralization is more phonetically or phonologically 'natural' in some way, then this could affect learning. If there is such an effect, it is expected that one group will have more ‘correct’ answers than the other, meaning that one group will come closer to accepting all of the attested (during the learning phase) liquid and none of the unattested liquid in utterance-final position.

M&P1 found that their phonetically-natural final devoicing pattern was learned more successfully than the phonetically-unnatural (but equal in formal complexity) final voicing pattern, consistent with claims that natural patterns are more successfully learned than unnatural ones (e.g. Wilson 2003, 2006). They do, however, discuss other possible explanations for this outcome. Despite the apparent effect of naturalness, both groups did also demonstrate learning and generalization.

It is less clear in the case of liquid neutralization whether there is a more natural direction, but it is important to make sure that participants can, in fact, learn both patterns. A difference in performance between the L-to-R group and the R-to-L group might suggest that either [l]-to-[r] or [r]-to-[l] neutralization is more natural, if the more natural pattern is learned more successfully, as in M&P1. If both groups show a preference for the same liquid ([r] or [l]) in final position, that might suggest that learning is not be occurring but that a preference for [r] or [l] is coming from elsewhere (e.g. prior language background, universal markedness, ease of articulation, etc.).

4.3 Results

While each participant responded to a total of 144 sentences during the testing phase, only the 72 sentences containing target words with coda liquids are included in the analysis, as the hypotheses all refer to word-final liquids.

It was determined that the quickest one could possibly read all instructions and complete the experiment was approximately 16 minutes, so one participant was excluded for finishing in under 16 minutes. Seven participants accepted every item. Because these participants were not responding to the stimuli, their responses were excluded.[8] Additionally, four participants figured out the pattern consciously, and explained it in their debrief. While these participants did behave as hypothesized in their acceptability judgments, it is likely that their behavior does not reflect the same type of (implicit) *learning* this experiment is trying to measure, so their results were excluded. It is notable, however, that these participants all characterized the pattern as a restriction on *word*-final sounds, rather than *utterance*-final sounds, showing evidence of generalization.

As noted above, additional participants were added until there was data from 24 participants in each group. For each participant, there were 18 observations per condition (18 utterance-final, [r]-final target words, 18 utterance-final [l]-final target words, 18 utterance-medial, [r]-final target words, and 18 utterance-medial, [l]-final target words). The proportion of 'accept' responses for utterances with liquid-final target words is presented in Figure 1, by word-final segment, learning group, and position of the target word in the utterance.

[8] The instructions were adjusted in the following experiments to prevent this.

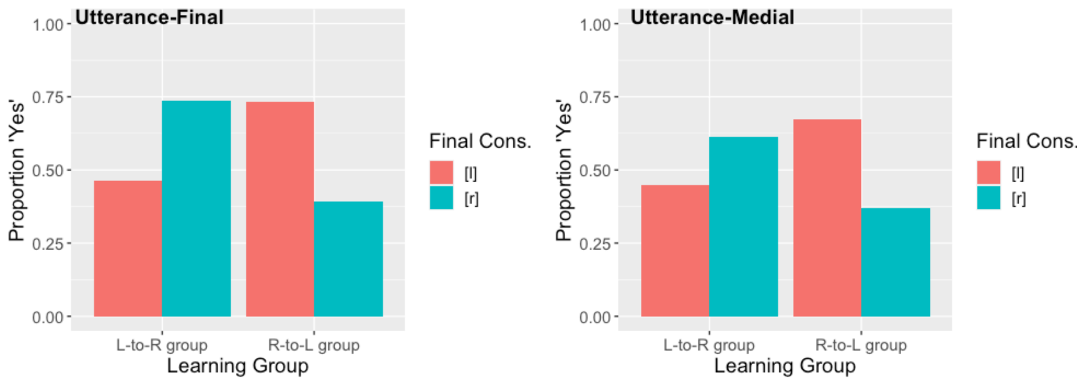

Figure 1: Proportion of 'yes'/'accept,' responses by liquid consonant and learning group: (left) utterance-final position; (right) utterance-medial position. (Experiment 1)

The chart on the left summarizes responses to utterances with the target word in final position. Participants in the L-to-R group accepted (responded 'yes' to) utterance-final [r] more frequently than utterance-final [l], while participants in the R-to-L group did the opposite, accepting utterance-final [l] more frequently than utterance-final [r].

The chart on the right summarizes responses to utterances with the target word in utterance-medial position. Though participants had not heard any word-final liquids in this position during the learning phase, participants in the L-to-R group still accepted utterances with word-final (utterance-medial) [r] more often than utterances with word-final [l], and participants in the R-to-L group did the opposite. This suggests that participants generalized the pattern learned during the learning phase to a different domain, though the difference in proportion of 'accept' responses for final [l] and final [r] was smaller in the utterance-medial condition.

The data were modeled by means of a mixed effects logistic regression analysis, using the *lme4* package in R (Barr et al. 2013, Bates et al. 2015). The dependent variable was the response ('no' or 'yes') to the question, *Does this sentence sound like it belongs to the language you have been learning?* 'Yes' was treated as the marked value of the response variable (coded as 1), and 'no' as the default (coded as 0). The fixed effects were the word-final liquid consonant (L or *R*), utterance position of the target word (final or *medial*), and learning group (L-to-R or *R-to-L*). In each of these factors, the first level was treated as the default value and the second (in italics) as the marked value. Random effects included individual participants and items (the utterances). Intercepts were included in the model for both of these random effects, as well as slopes for Final consonant and Utterance position for each participant. The results[9] for the fixed effects are given in Table 1.

	b	*z*	*p*
Intercept	-0.10632	-0.511	0.609
Final consonant ([r])	1.19583	4.520	6.18e-06 *
Utterance position (medial)	-0.19305	-0.963	0.335
Learning group (R-to-L)	1.13639	3.970	7.17e-05 *
Final consonant × Utterance position	-0.20344	-1.301	0.193
Final consonant × Learning group	-2.66862	-7.619	2.56e-14 *
Utterance position × Learning group	0.04202	0.160	0.873

*Table 1: Fixed effects in a logistic regression model of responses (Experiment 1). Significant values (p < 0.05) are indicated by *.*

[9] The model failed to converge when including a three-way interaction between Final consonant × Utterance position × Group (as in M&P1), so I only look at two-way interactions here.

There were significant main effects of Final consonant and Learning group. This does not necessarily reveal anything particularly relevant to the hypotheses for this experiment, especially given the complexity of the model. Across both utterance positions and both groups, final [r] was accepted slightly more often than final [l] (913 vs. 884). The L-to-R group also responded 'yes' 40 more times than the R-to-L group did (977 vs. 933) over all conditions. It is possible that [r] is preferable to [l] word-finally based on some sort of naturalness criterion, or that the trilled [r] was somehow more salient than [l], leading to slightly different learning patterns. L1 phonology may also have played a role, as final /l/ is generally velarized in American English, while trill /r/ is not an American English phoneme and, therefore, has no expected alternations.

More relevant to the hypotheses is the interaction between Final consonant and Learning group. It was predicted that the L-to-R group would accept more [r] codas than [l] codas, while the R-to-L group would do the opposite. As Figure 1 shows, this was indeed the case, suggesting that the patterns were learned in both groups. No significant effect of Utterance position is also potentially indicative of generalization.

To explore this further, the data were broken into subsets according to Utterance position and Learning group. Table 2 gives the results for four tests, with the same random effects structure as in the previous analysis, and the single fixed effect of Final consonant.

Utterance position	Learning group	*b*	*z*	*p*
final	L-to-R	1.4975	3.958	7.55e-05 *
final	R-to-L	-1.6453	-5.587	2.31e-08 *
medial[10]	L-to-R	0.8216	3.313	0.000925 *
medial	R-to-L	-1.4338	-5.205	1.94e-07 *

*Table 2: Effects of Final consonant in data subsets defined by Utterance position and Learning group (Experiment 1). Significant values (p < 0.05) are indicated by *.*

As can be seen in Table 2, final liquid consonant ([l] vs. [r]) had a significant effect on responses for both learning groups in the final condition. This shows that participants were able to learn the liquid neutralization pattern in utterance-final position, based on their exposure to the pattern during the learning phase.

In the utterance-medial condition, Final consonant also had a significant effect on responses for both learning groups, suggesting that the learned pattern was generalized from utterance-final words to a word-level rule, though this effect was somewhat weaker for the L-to-R group. No significant effect of Utterance position or interaction between Utterance position and Final consonant was found for the R-to-L group (Table 3), but a significant interaction ($p < 0.05$) between Utterance position and Final consonant was found for the L-to-R group (Table 4).

[10] For this subset, the model would not fit unless it was simplified by removing the random effect for item.

R-to-L group	*b*	*z*	*p*
Intercept	1.0858	6.544	5.97e-11*
Final consonant ([r])	-1.6086	-6.104	1.03e-09*
Utterance position (medial)	-0.3254	-1.615	0.106
Final consonant × Utterance position	0.1375	0.622	0.534

*Table 3: Fixed effects within the R-to-L group (Experiment 1). Significant values (p < 0.05) are indicated by *.*

L-to-R group	*b*	*z*	*p*
Intercept	-0.19668	-0.739	0.4601
Final consonant ([r])	1.39926	4.571	4.85e-06*
Utterance position (medial)	-0.02236	-0.095	0.9242
Final consonant × Utterance position	-0.57688	-2.554	0.0106*

*Table 4: Fixed effects within the L-to-R group (Experiment 1). Significant values (p < 0.05) are indicated by *.*

Participants in the L-to-R group accepted fewer items with word-final [r] in the utterance-medial condition, suggesting some uncertainty in the condition for which they had no evidence of permissible final liquids during the learning phase.

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. This means that for the L-to-R group, it was correct to accept coda [r] and to reject coda [l]. For the R-to-L group, it was correct to accept coda [l] and to reject coda [r].

	utterance-final	utterance-medial
L-to-R	63.7%	58.2%
R-to-L	67.0%	65.0%

Table 5: Percentage of correct responses by Learning group and Utterance position (Experiment 1).

As expected based on the results above, both groups had a higher percentage of correct responses in the final condition than in the medial condition. There was also a higher percentage of correct responses in the R-to-L group than in the L-to-R group, though this difference was not statistically significant, as shown in Table 6 below.

The recoded results were modeled using a mixed model regression analysis with Correctness of the response as the dependent variable and Learning group and Utterance position as the fixed effects. Random effects intercepts were included in the model for both participant and item, with a slope factor for Final consonant within participant. Only Utterance position had a significant main effect. Neither Learning group, nor the interaction between Learning group and Utterance position was significant.

	b	*z*	*p*
Intercept	0.6389	4.612	3.99e-06*
Utterance position (medial)	-0.2734	-2.539	0.0111*
Learning group (R-to-L)	0.1895	0.970	0.3322
Utterance position × Learning group	0.1750	1.152	0.2494

Table 6: Fixed effects in a logistic regression model of response correctness (Experiment 1).

The main effect of Utterance position is consistent with imperfect generalization to the utterance-medial position, for which participants had not received direct evidence of liquid goodness during the learning phase.

While the R-to-L group, qualitatively, appears to have learned and generalized slightly more successfully than the L-to-R group, this effect did not reach significance. Both groups demonstrated evidence of both learning and generalization.

4.4 Discussion

As in M&P1, these results show that participants were able to learn a phonological distribution pattern in their learning sets on relatively brief exposure (60 utterances repeated three times), as evidenced by their responses to novel items. None of the participants demonstrated complete "mastery" of the pattern from the learning sets, but they did show a significant tendency to accept pattern-congruent items more often than pattern-violating ones.

Participants were given evidence for the restriction on final [l] or [r] only in utterance-final position, but their preference was carried over to novel, utterance-medial items, demonstrating generalization of the pattern to the word level. This generalization was slightly weaker–though still significant–in the L-to-R group, potentially suggesting a naturalness effect (Hypothesis 3). However, another possible explanation for this asymmetry was revealed in participants' comments in the exit survey. Participants were asked about their strategies for choosing good/bad utterances. As discussed above, data from the four participants who consciously figured out the pattern were excluded from analysis, but a number of participants did make mention of [r] in their comments. Several of these expressed frustration with, or dispreference for, trilled [r]. This rhotic had been selected because it is articulatorily and perceptually distinct from [l], as well as non-English-sounding. It is a phoneme that does not exist in the inventory of American English, yet L1 English speakers are easily able to perceive it. This may, however, have made it overly salient, especially because the recorded pronunciations were somewhat dramatically trilled. Because participants had to repeat the items during the learning phase, many were frustrated by their own inability to pronounce that sound. For these reasons, trill [r] was abandoned in favor of an English approximant [ɹ] in subsequent experiments.

Overall, Experiment 1 was able to replicate the results of M&P1, showing that participants can learn a pattern and generalize it from the utterance domain to the word domain given relatively short exposure in an AGL experiment. This experiment was very similar in design to M&P1, but with monosyllabic target words and a different domain-final neutralization pattern: final liquid ([r]~[l]) neutralization rather than final devoicing. Given these results, the next step was to investigate whether learners can generalize beyond the word level to the syllable level, resulting in syllable-level phonological phenomena, such as Tz'utujil devoicing.

5. Experiment 2: No generalization to the syllable

While the main goal of Experiment 2 was to test whether learners can generalize phonological patterns to the syllable level, I also wanted to explore whether final stress (like that in Tz'utujil) plays a role in facilitating learning and/or generalization of domain-final phenomena, as discussed above. The stimuli described below were, therefore, recorded with both penultimate stress and final stress so that two participant groups could be created: one that heard only stimuli with final stress and one that heard only stimuli with penultimate stress. A significant difference in learning and/or generalization between these two groups might suggest that stress can affect these processes.

The hypothesis was that final stress aids in learning and/or generalization, so it was expected that the final-stress group would show more successful learning/generalization than the penultimate-stress group. Because it was uncertain whether either group would generalize to the syllable, and in order to avoid spending unnecessary time and research funds, the final-stress group was run first. The plan was to recruit more participants and run the penultimate-stress group only if the final stress group showed evidence of generalization to the syllable, so that both learning and generalization could be compared between the two groups. If the final stress group did not show evidence of learning or generalization, it is highly unlikely that the penultimate stress group would do so. As discussed in the results section below, the first group showed learning but not generalization, so a second group was never recruited.

Though this result might suggest that generalization to the syllable is not possible (at least in this type of experiment), crucial differences between Experiments 1 and 2 posed enough questions and potential confounds that a subsequent third experiment was devised. I will briefly describe the design and results of Experiment 2 below, before discussing Experiment 3 in more detail.

5.1 Methods

5.1.1 Participants. As in Experiment 1, participants were recruited from across the United States via Prolific Academic and paid for their time. There were 25 participants,[11] ranging in age from 22-58. All were native speakers of English residing in the United States with no language-related disorders or hearing difficulties.

5.1.2 Materials. In this experiment, stimuli consisted of disyllabic nonce words, of the two shapes in (5), where C belonged to the set [ʒ,m,l,ɹ] and V belonged to the set [i,e,o,u]. Target words were not presented in carrier sentences, but they were grouped into two conditions, based on location of the coda consonant within the word, as shown in (5). Target coda consonants are bolded and underlined for clarity.

(5)	a.	*word-final*	.CV.CV<u>**C**</u>.
	b.	*word-medial*	.CV<u>**C**</u>.CV.

In an attempt to simplify the stimuli, the consonant inventory included only one nasal ([m]), one obstruent ([ʒ]), and the two liquids ([l] and [ɹ]).[12] Each of these consonants occurred with equal frequency in onset position, but [l] never occurred in in coda position, and [ɹ] occurred only in word-final codas. Participants, therefore, never heard liquid codas word-internally during the learning phase.

---

[11] Participants in Experiment 2 identified predominantly (72%) as white/of European descent and 64% male. I do not currently believe that ethnic identity or gender, alone, should affect performance on this task, but demographic data is reported in case it is relevant to future work. For more demographic data, see Appendix A.
[12] It was later determined that this strange and unnatural inventory presented a potential confound, as discussed below, and it was abandoned in Experiment 3.

The voiced post-alveolar fricative was chosen as the one obstruent because of concerns about resyllabification. Due to their native phonotactics, L1 English speakers will often perceive obstruent-sonorant (or sibilant-consonant) clusters as complex onsets, rather than coda-onset sequences, leading them to perceive CVCCV words as .CV.CCV., rather than .CVC.CV. The voiced post-alveolar fricative, however, never occurs as the C1 in English onset clusters, so it would better demonstrate that word-internal codas are permitted in this made-up language.

The stimuli for the learning stage are given in (6a), and those for the testing stage are given in (6b-c). Each target word in (6) was recorded with stress on two different syllables—penultimate and final—by a PhD linguist who is a native speaker of American English. Unlike the stimuli for Experient 1, these stimuli were intentionally pronounced with English-like vowels and liquids. For example, vowels written as <e> or <o> in (6) were actually diphthongized to sound more like American English vowels. Stress, too, was realized as in the speaker's L1 American English, but without vowel reduction. The strongest correlates to stress were pitch accent and vowel duration.

(6) a. *Stimuli used during the learning phase of Experiment 2*

Block 1 .CV.CVC.			Block 2 .CVC.CV.	
nasal-coda	obs-coda	[ɹ]-coda	nasal-coda	obs-coda
.lu.lem.	.mi.meʒ.	.mu.ɹeɹ.	.ʒem.ɹi.	.ɹeʒ.mu.
.le.ʒem.	.lo.meʒ.	.ʒi.ʒeɹ.	.mem.ʒi.	.meʒ.li.
.ʒu.ɹem.	.ɹo.ɹeʒ.	.ɹi.leɹ.	.lem.ɹu.	.leʒ.mo.
.mo.mim.	.ʒo.miʒ.	.ɹe.ʒiɹ.	.lim.le.	.ɹiʒ.ɹo.
.mu.ɹim.	.mo.ʒiʒ.	.ʒu.miɹ.	.ʒim.ʒo.	.liʒ.mi.
.lo.ʒim.	.mi.liʒ.	.me.ɹiɹ.	.ɹim.ɹu.	.ʒiʒ.me.
.ʒi.ʒom.	.ʒe.moʒ.	.ʒu.loɹ.	.lom.ʒu.	.moʒ.mi.
.lu.ɹom.	.lo.ɹoʒ.	.ɹo.moɹ.	.ɹom.ʒe.	.ʒoʒ.ɹu.
.ʒo.lom.	.me.ʒoʒ.	.li.loɹ.	.mom.le.	.ɹoʒ.me.
.ɹe.mum.	.ɹi.luʒ.	.ɹi.luɹ.	.lum.lo.	.muʒ.ɹe.
.mu.lum.	.ɹu.muʒ.	.li.ʒuɹ.	.ʒum.ʒi.	.ɹuʒ.lo.
.ʒo.ʒum.	.ɹo.ɹuʒ.	.lo.ɹuɹ.	.mum.ʒo.	.ʒuʒ.lu.

b. *Stimuli used during the testing phase of Experiment 2 - Block 1*

Block 1 .CV.CVC.			
[m]-coda	[ʒ]-coda	[ɹ]-coda	[l]-coda
.lo.ɹum.	.le.miʒ.	.ʒi.ʒiɹ.	.mi.ʒel.
.ʒo.mim.	.li.leʒ.	.li.ʒoɹ.	.ʒe.mil.
.li.ʒom.	.mo.liʒ.	.mo.muɹ.	.ɹi.mil.
.ʒi.ʒem.	.ɹu.ɹuʒ.	.mi.ʒuɹ.	.me.mul.
.lu.ʒum.	.lu.liʒ.	.ʒi.ɹeɹ.	.le.ʒel.
.ɹo.ʒim.	.lo.leʒ.	.me.moɹ.	.mu.ʒol.
.mu.ɹum.	.li.ɹiʒ.	.ʒo.ɹoɹ.	.li.ɹil.
.mo.ʒim.	.me.moʒ.	.me.ʒiɹ.	.ɹi.lol.
.ɹu.ɹom.	.ʒu.loʒ.	.ɹi.meɹ.	.mo.lol.
.le.ʒom.	.le.ɹoʒ.	.le.luɹ.	.ʒu.ʒil.
.mu.mim.	.ɹi.ɹoʒ.	.ʒu.meɹ.	.mu.lol.
.ɹe.ɹim.	.mu.ɹeʒ.	.mi.muɹ.	.lu.ɹul.
.ɹe.lum.	.ʒi.meʒ.	.lo.luɹ.	.lo.ʒul.
.ʒu.lom.	.ɹu.ʒeʒ.	.ʒe.ʒuɹ.	.ɹo.ɹel.
.le.lem.	.ɹe.miʒ.	.ʒe.ɹoɹ.	.ʒu.mol.
.me.lim.	.mi.ɹoʒ.	.ʒu.moɹ.	.lo.mel.
.ʒe.ɹum.	.ʒi.ɹeʒ.	.ɹe.muɹ.	.ɹe.lil.
.ɹe.mem.	.ɹo.luʒ.	.mu.leɹ.	.ʒi.ɹel.

c. *Stimuli used during the testing phase of Experiment 2 - Block 2*

Block 2 .CVC.CV.			
[m]-coda	[ʒ]-coda	[ɹ]-coda	[l]-coda
.mum.ɹi.	.luʒ.lu.	.loɹ.li.	.ɹol.ɹi.
.ʒim.lo.	.muʒ.mu.	.ɹiɹ.me.	.mul.ʒe.
.ɹim.ɹe.	.moʒ.mo.	.ɹuɹ.lo.	.lel.ʒu.
.ɹom.ɹi.	.loʒ.li.	.loɹ.lu.	.lel.ɹu.
.ʒem.ʒi.	.ʒeʒ.mu.	.miɹ.mo.	.mel.ɹu.
.ʒim.ʒu.	.ʒiʒ.ɹu.	.ɹiɹ.lu.	.lul.me.
.lum.ɹu.	.ɹiʒ.mi.	.ɹoɹ.me.	.lil.ɹi.
.ʒum.le.	.luʒ.le.	.ʒeɹ.lo.	.ʒil.ɹe.
.mum.lo.	.miʒ.ɹu.	.loɹ.ʒi.	.ʒul.mu.
.ʒum.ɹi.	.ɹiʒ.mu.	.ʒuɹ.ʒi.	.mul.ɹi.
.mem.ʒe.	.leʒ.ɹo.	.ʒuɹ.lo.	.ɹel.ʒe.
.lim.ɹu.	.ɹeʒ.ɹi.	.ɹoɹ.ʒo.	.ɹil.ɹo.
.ʒim.ʒe.	.loʒ.mi.	.meɹ.lu.	.mul.ɹo.
.lum.ʒi.	.ɹoʒ.li.	.loɹ.mo.	.lil.ʒu.
.ɹom.lo.	.ɹeʒ.lu.	.ʒuɹ.me.	.mil.ɹi.
.ɹem.ɹu.	.luʒ.mo.	.ʒeɹ.me.	.ɹol.mi.
.mim.ʒo.	.loʒ.mo.	.ʒiɹ.li.	.ʒel.me.
.ɹem.ʒi.	.ɹeʒ.ɹe.	.meɹ.ʒo.	.mol.ʒu.

It was intended that one group would be presented only with final-stress stimuli and the other group would be presented only with penultimate-stress stimuli. Both groups would otherwise hear the same words and perform identical tasks. After the final-stress group failed to show generalization to the syllable, however, the penultimate-stress stimuli were never used.

5.1.3 Procedure. The procedure was the same as in Experiment 1, though the instructions were slightly modified for clarity. Participants were told they were listening to *words* in a made up language, rather than *sentences*.

5.1.4 Hypotheses. In this experiment, all participants were exposed to the same pattern of liquid neutralization where [ɹ] but not [l] occurred in word-final position (and neither liquid occurred in word-internal codas) during the learning phase.

Hypothesis 1: Learning. If the participants in this experiment learn the pattern they are exposed to during the learning phase, their 'accept' responses should reflect the pattern in the testing phase. This means that participants will accept test words with word-final [ɹ] more frequently than those with word-final [l], showing an effect of the learning set on participants' responses during the testing phase.

Hypothesis 2: Generalization. The learning phase does not provide participants with evidence about which liquids are allowed in word-medial, syllable-final position. If participants generalize the pattern they learned from word-final position to syllable-final position, then the pattern should hold for test items with word-medial coda consonants. This means participants will accept syllable-final, word-medial [ɹ] more frequently than they accept syllable-final, word-medial [l]. This would show generalization of the learned pattern to the smaller, syllable-level domain.

Hypothesis 3: Stress. If final stress facilitates learning and/or generalization by making word-final syllables more salient, then the final-stress group should show more "correct" answers than the penultimate-stress group, in the word-final condition (learning), the word-medial condition (generalization), or both (learning & generalization). This means participants in the final-stress group will show a larger effect of coda consonant ([ɹ] vs. [l]), accepting more [ɹ] codas and/or rejecting more [l] codas than the penultimate-stress group.

5.2 Results

As in Experiment 1, the data were modeled by means of a mixed model logistic regression analysis, with the same dependent variable (response 'yes'/accept or 'no'/reject). The fixed effects were the Coda liquid consonant (L or *R*) and Coda position (word-final or word-*medial*). The Learning group condition was not included in the analysis, as only one group (final stress) was run. Random effects included individual participants and items (the utterances). Intercepts were included in the model for both of these random effects, as well as slopes for Coda consonant and Coda position for each participant. The results for the fixed effects are given in Table 7.

	b	*z*	*p*
Intercept	-0.7595	-2.609	0.00909 *
Coda consonant ([ɹ])	1.8276	7.298	2.92e-13 *
Coda position (word-medial)	0.6082	2.315	0.02060 *
Coda consonant × Coda position	-1.8957	-5.961	2.51e-09 *

*Table 7: Fixed effects in a logistic regression model of responses (Experiment 2). Significant values ($p < 0.05$) are indicated by *.*

There were significant main effects of Coda consonant and Coda position, as well as a significant interaction between Coda consonant and Coda position. To explore these main effects further, the data were broken into subsets by Coda position. Results are shown in Table 8.

Coda position	*b*	*z*	*p*
final	1.9524	5.570	2.55e-08 *
medial	-0.0599	-0.246	0.806

Table 8: Effects of Coda consonant in data subsets defined by Coda position (Experiment 2).

A significant effect of Coda consonant was found only in word-final position, but not word-medially, suggesting learning but not generalization occurred. Figure 2 shows participant-by-participant 'yes' (accept) responses by Coda position. While nearly all participants showed some preference for coda [ɹ] over coda [l] word-finally, the pattern did not hold for the word-medial items.

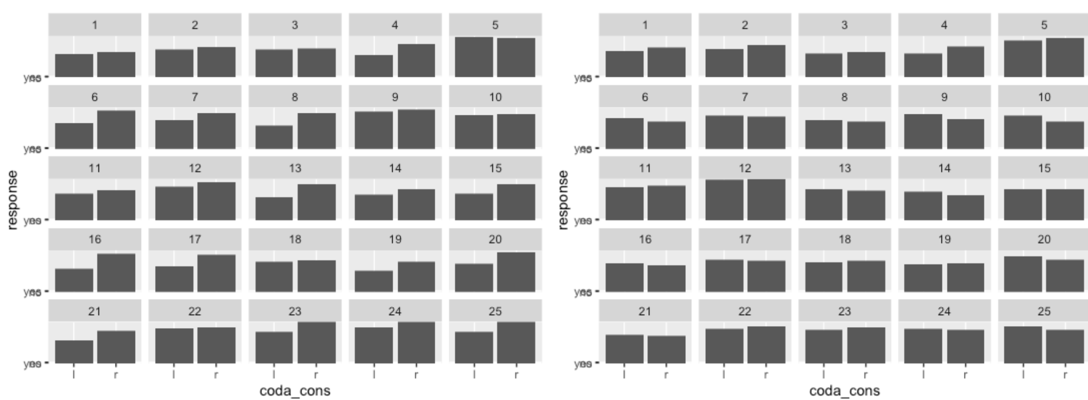

Figure 2: By-participant responses for coda [l] and coda [ɹ] in word-final position (left) and word-medial position (right). 'Yes' or 'accept' response is 1, and 'no'/'reject' is 0.

5.3 Discussion

As in Experiment 1 (and M&P), participants were able to learn the pattern from the learning set from relatively brief exposure during the learning phase of the experiment. Though none of them performed "perfectly" (accepting all [ɹ]-final words and rejecting all [l]-final words), participants did show a preference for word-final liquids consistent with the pattern they were exposed to.

They did not, however, generalize this pattern to word-medial codas. This could be evidence of a preference for word-level phonology over syllable-level phonology. It may suggest that syllables are not available in the same way as words for frequency calculations or derivation of phonological rules/constraints/distributions.

Another possibility is that generalization to the syllable is somehow a more complex (or just slower) process than generalization to the word, and it simply cannot occur in this type of AGL experiment.

There were, however, many differences between Experiment 1 and Experiment 2, any one or more of which may pose a potential confound when comparing the utterance-to-word generalization of Experiment 1 to the lack of word-to-syllable generalization found in Experiment 2. Perhaps the most glaring difference is in the size and makeup of the segment inventories for the stimuli in the two experiments. Experiment 1 employed a small, but plausible, consonant inventory with two stops, two nasals, and two liquids ([t,k,m,n,l,r]), and a widely-attested five-vowel ([i,e,a,o,u]) inventory. Experiment 2, on the other hand, used only one obstruent, one nasal, and the two liquids ([ʒ,m,l,ɹ]), and only four vowels ([i,e,o,u]). Such vowel inventories are certainly attested, but the slightly smaller vowel inventory, combined with the smaller consonant inventory, makes for an overall smaller inventory.

Relatedly, the ratio of liquids to other consonants changed. In Experiment 1, liquids comprised one third of the consonant inventory, but in Experiment 2, they comprised one half of the consonants.

Even the consonants and vowels that occurred in the stimuli in both experiments were pronounced somewhat differently, as Experiment 1 had trill [r], only 'light' [l], and non-English monophthongs [e,o], while Experiment 2 featured more English-like pronunciations.

While post-alveolar /ʒ/ is a phoneme of English, and thus both pronounceable and perceptible by L1 English speakers, it is both typologically (cross-linguistically) rare and, also, the least-frequently-occurring phoneme in the English lexicon (Hayden 1950). There are no attested languages whose obstruent inventory contains only /ʒ/. This makes the inventory in Experiment 2 rather marked/unusual, and, based on exit survey comments, the [ʒ] stood out to many participants. This may have somehow distracted from the liquids, which were the focus of the experiment.

Apart from sounds, there was another potentially consequential difference between the two studies. Experiment 1 presented target words in carrier sentences. Within each block, the carrier sentence remained the same while only the target words changed. The same target words were repeated in the utterance-medial and utterance-final frames. Experiment 2, on the other hand, did not have carrier sentences, as each item/utterance was a single, disyllabic word. Stimuli for the word-medial and word-final conditions consisted of entirely different target words. Participants in Experiment 1 heard the same 24 non-liquid-final target words repeated three times each in the final frame and three times each in the medial frame (a total of six repetitions per word), as well as 12 liquid-final target words repeated three times in the final frame. Participants in Experiment 2 heard 60 different target words during the learning phase, but they did not hear any one target word more than three times.

Experiment 2 appears to present a null result with regards to the generalization hypothesis (Hypothesis 2 above). Thus far, it appears that, in an AGL experiment, generalization from the utterance level to the word level is possible, but generalization to the syllable is not. Before concluding this, however, it seems important to investigate whether these differences of experiment design can account for the differing results.

6. Experiment 3: Generalization to the word and the syllable

Experiment 3 was designed to address the problem of the many differences between Experiments 1 and 2 by re-testing for both generalization to the word and generalization to the syllable using identical stimuli. Participants were broken into two groups who heard identical stimuli and performed the same learning and testing tasks. The only difference between the groups was in the instructions they were given and, thus, what they believed they were hearing and learning.

One group was told repeatedly that they were listening to *three-word sentences* in a made up language. This group is referred to from here on as the utterance-to-word (UtW) group because it was hypothesized that they would generalize from utterance-final position to word-final position.

The other group was told repeatedly that they were listening to *three-syllable words* in a made up language. This group will be called the word-to-syllable (WtS) group because it was hypothesized that they would generalize from the word level to the syllable level.

As often as possible, both verbal and written instructions referred to either '(three-word) sentences' or '(three-syllable) words,' depending on the group. The buttons participants clicked to hear the next item during the learning phase said either, 'Next Sentence,' for the UtW group, or, 'Next Word,' for the WtS group. The question on screen after each item in the testing phase was either 'Does this sentence sound like it belongs to the language you have been learning?' or 'Does this word sound like it belongs to the language you have been learning?' Judging by comments in the exit survey, the UtW group believed they were listening to three-word sentences, and the WtS group believed they were listening to three-syllable words.

To improve data quality and encourage participants to thoroughly read and follow instructions, three checks were added. After the first block of the learning phase, participants were asked whether they had been repeating the items (words or sentences) aloud and reminded that doing so is important. At two points during the testing phase, participants were presented with a simple yes-no question[13] as an attention check. It was decided ahead of time that only failure of *both* attention checks would result in a participant's responses being excluded from analysis. Seven participants failed a single attention check, but nobody failed both.

[13] The attention check questions were, '*Are you a giant, talking spider? Please answer truthfully. This is an attention check.*' and '*Are you a human? Please answer truthfully. This is an attention check.*' As in the rest of the test phase, participants could click 'yes' or press the 'd' key or click 'no' or press the 'k' key.

Another difference between Experiment 3 and the prior experiments was the participant pool. While the experiment was still carried out online using PCIbex, participants were all undergraduate students at the University of California, Santa Cruz, enrolled at the time of their participation in linguistics courses. They received course credit, rather than monetary compensation. This was—in some ways—a less diverse group of participants, and there is a concern that they had some prior training in linguistics and phonology, meaning they may be more consciously aware of phonological patterns than participants recruited externally. The majority of these participants, however, had taken only introductory-level linguistics courses,[14] and all were instructed not to attempt any analysis. Comments on their exit surveys suggest that learning was still relatively implicit, though there was more awareness of "R" and the right edge than in the exit surveys from Experiments 1 and 2. This is discussed further in the discussion section below.

An advantage of using the undergraduate participant pool was the opportunity to meet via video conferencing with participants. Participants were briefed in groups of 1-4 on Zoom before they were given the link to the experiment. This allowed an opportunity to provide both print and verbal instructions, to check that participants had an appropriate setup in terms of equipment (computer, headphones) and also a quiet space where they could focus, and to answer any questions participants might

[14] Many participants were recruited from Ling 50 Introduction to Linguistics or Ling 80k Language and Culture, introductory courses that only briefly introduced phonetics and phonology without much analysis of phonological alternations or patterns. There were, however, a number of students who had taken Phonetics 1 and Phonology 1 and other linguistics courses. Only one student had taken Phonology 2.

have before beginning. I also remained available in the Zoom "room" in case participants had questions or technical difficulties during the experiment or wanted to discuss it in more detail afterwards. Because the participant pool was comprised of students, some information about the experiment and what it was intended to observe was provided for educational purposes upon completion of the testing phase and the exit survey. Students also had the opportunity to ask questions via Zoom or email if they wanted to.

The main reason for switching to the student participant pool was cost, though recruiting from undergraduate students was the standard before the COVID-19 pandemic. It is also how M&P recruited participants for their experiments.

6.1 Methods

6.1.1 Participants. Initially, 88 participants were recruited, though results were not obtained from twelve of these due to technical problems. All students were granted course credit in exchange for their participation, whether or not they were able to complete the entire experiment. This left 76 participants, divided evenly between two groups. Due to the nature of the subject pool, participants could not be prescreened or disqualified as easily as in the prior experiments. More than a quarter of the participants were native speakers of languages besides English. Rather than exclude all of their data, a decision was made to exclude data only from participants who reported that English was not their *dominant* language. This left 71 participants: 36 in the word-to-syllable group and 35 in the utterance-to-word group.

Participants ranged in age from 18-29 years old (mostly 18-22), and were overwhelmingly California natives, with some exceptions. Unlike the participants recruited via Prolific in Experiments 1 and 2, who skewed male, UCSC student participants were majority (about 70%) female-identifying. One participant in the UtW group reported a slight bilateral hearing loss. Though this is typically grounds for exclusion from this type of experiment,[15] the participant did not express concern about their ability to complete the activity. Because this experiment is carried out on participants' personal computers with headphones, it is possible for participants with some degree of hearing loss to adjust their sound settings to essentially correct hearing.

6.1.2 Materials. Stimuli consisted of tri-syllabic nonce utterances (to be interpreted as either words or sentences, depending on the group), of one of the two shapes in (7), where C belonged to the set [t,k,m,n,l,ɹ] and V belonged to the set [i,a,u]. As in Experiment 2, there were no carrier sentences, but stimuli were grouped into two conditions based on the location of the coda consonant within the item.

(7)	a.	*final*	.CV.CV.CV**<u>C</u>**.
	b.	*medial*	.CV.CV**<u>C</u>**.CV.

Target coda consonants are bolded and underlined for clarity.

After Experiment 2, a more plausible (less-marked) consonant inventory was chosen for these stimuli. Mainly, the voiced post-alveolar fricative ([ʒ]) was replaced with two stops ([t,k]). This did raise the concern that some obstruent codas in the

[15] The author notes that she has a mild-to-moderate bilateral hearing loss, as well as an attention disorder, and hopes to challenge the ways we choose participants to include/exclude in experimental work in the future.

medial frame (7b) would be perceived as C1 in an onset cluster, rather than codas, due to participants' English phonotactics. It was decided that this affected a small enough percentage of the stimuli[16] that participants should still be able to learn that codas are allowed in word/utterance-medial position. Obstruent codas are, otherwise, not part of the analysis, so this issue should not affect results.

While Experiment 2 featured an /ɹ/ → [l] neutralization, where [l] never occurred in coda position during the learning phase, Experiment 3 switched the direction of the neutralization. Each consonant occurred with equal frequency in onset position, but [ɹ] never occurred in in coda position, and [l] occurred only in final codas. Participants, therefore, never heard liquid codas word/utterance-internally during the learning phase. The stimuli for the learning stage are given in (8a), and those for the testing stage are given in (8b-c).

Each item in (8) was recorded by the author, a linguistics PhD student and native speaker of American English.[17] As in Experiment 2, these stimuli were intentionally pronounced with English-like vowels and liquids and English-like stress. All utterances were produced with declarative intonation and final stress that could be interpreted either as word-final stress (for the WtS group) or sentence stress (for the UtW group).

---

[16] Only two of the twenty-four medial-coda (.CV.CVC.CV.) target words in the learning phase contained obstruent-[ɹ] sequences that could, potentially, be perceived as tautosyllabic onset clusters. An effort was made to pronounce these in a way that did not sound like an English onset cluster (obstruents were released, and [ɹ] was fully voiced).

[17] For Experiments 1 and 2, colleagues were asked to record stimuli to avoid any potential bias in the author's production of the utterances. This was a challenging and time-consuming ask, however, and it was not clear how the author might possibly produce biased stimuli. Therefore, the author recorded these utterances.

(8) a. *Stimuli used during the learning phase of Experiment 3*

Block 1 .CV.CV.CVC.			Block 2 .CV.CVC.CV.	
nasal-coda	obs-coda	[l]-coda	nasal-coda	obs-coda
.ni.ni.man	.ti.nu.lut.	.ka.na.nil.	.ti.mam.ki.	.ra.muk.na.
.ki.la.kum.	.nu.ti.rak.	.ta.mu.mul.	.na.kin.ka.	.ru.kik.mu.
.lu.ki.kan.	.ra.ru.lak.	.lu.ta.nal.	.la.lan.ra.	.li.nat.li.
.ta.ki.nim.	.ki.ma.rut.	.ra.lu.mil.	.ku.nim.li.	.mu.tat.nu.
.nu.nu.tin.	.mu.ku.lik.	.na.ti.nil.	.ti.kum.ka.	.li.nut.ku.
.tu.la.lan.	.na.la.tuk.	.ki.ri.rul.	.ka.run.ma.	.na.nak.ta.
.ri.ka.kim.	.mu.ti.tat.	.ti.mi.kil.	.ti.mun.li.	.na.rik.ta.
.ra.mi.kin.	.ra.lu.tak.	.ka.ma.tal.	.ma.lum.tu.	.mi.tat.ni.
.ri.li.kam.	.ti.ku.mit.	.mu.ra.lal.	.la.tun.mu.	.ki.lik.mu.
.nu.ku.rin.	.mi.tu.ruk.	.li.ni.nul.	.mu.run.ti.	.ru.mit.li.
.lu.na.lum.	.ka.ru.tut.	.mu.ta.mil.	.ki.ram.nu.	.tu.lit.ri.
.la.ra.num.	.mi.ri.mat.	.lu.mu.rul.	.nu.kam.ra.	.ru.tik.ru.

b. *Stimuli used during the testing phase of Experiment 3 - Block 1*

Block 1 .CV.CV.CVC.			
nasal-coda	obst-coda	[l]-coda	[ɹ]-coda
.mu.lu.tim.	.ta.ti.kat.	.ma.nu.mul.	.ni.ki.mir.
.la.nu.run.	.tu.ri.lak.	.ku.la.nul.	.ki.li.tar.
.nu.ta.lan.	.ti.nu.tit.	.lu.mi.ril.	.li.ra.kar.
.ri.ki.tun.	.lu.ka.muk.	.ka.la.kul.	.ri.ma.tir.
.ti.ti.tam.	.mi.ni.nit.	.ta.ka.lil.	.lu.la.kir.
.li.ki.nam.	.ri.ta.mit.	.mu.mu.kul.	.na.na.tur.
.ma.mi.mun.	.ma.ta.kak.	.ra.tu.ril.	.ku.ra.lar.
.li.li.kum.	.na.ru.nak.	.ka.lu.kal.	.lu.ta.mur.
.ma.ra.rim.	.ra.mu.lut.	.ti.nu.kal.	.tu.mi.rir.
.li.ki.rin.	.ki.na.mik.	.ki.mu.kal.	.ru.ta.tar.
.na.ra.nin.	.nu.ri.tut.	.ti.ru.ral.	.ki.tu.mur.
.mu.nu.lam.	.tu.mi.mat.	.mi.mi.lil.	.ri.mi.lar.
.mi.la.nin.	.la.mu.mit.	.tu.ka.rul.	.nu.ka.tir.
.tu.nu.tum.	.ra.ri.kik.	.ma.mi.lul.	.ra.ru.lir.
.na.ru.nun.	.ru.la.nak.	.ku.li.nul.	.mi.ku.nar.
.ku.ni.ram.	.ra.na.lut.	.ti.lu.lil.	.lu.ki.tar.
.la.li.nam.	.na.ru.nuk.	.ki.ku.mil.	.na.tu.mur.
.ri.tu.run.	.na.ka.rak.	.nu.na.rul.	.ku.ti.kir.

c. *Stimuli used during the testing phase of Experiment 3 - Block 2*

Block 2 .CV.CVC.CV.			
nasal-coda	obs-coda	[l]-coda	[ɹ]-coda
.ka.tum.na.	.ti.ruk.tu.	.ri.tal.ti.	.tu.tar.ni.
.ni.tan.ru.	.ni.lak.mu.	.li.tul.mi.	.mi.mar.na.
.ra.kim.ku.	.ta.tik.ri.	.ki.tal.ni.	.mi.rur.ku.
.li.rin.ri.	.li.lut.ra.	.mu.ral.ma.	.ka.nar.nu.
.ru.mim.ta.	.ri.mik.ri.	.lu.tul.ti.	.nu.rir.ki.
.mu.lim.tu.	.ma.nat.mu.	.ru.nil.ma.	.ru.kir.tu.
.ni.mun.tu.	.na.mit.ki.	.na.tal.ni.	.mu.nur.ki.
.ru.rim.lu.	.ra.luk.ru.	.ta.lil.ta.	.mi.lur.nu.
.mu.man.ti.	.ti.nut.ku.	.ma.nil.ku.	.na.nur.ku.
.la.mun.ka.	.lu.kit.ku.	.li.lul.nu.	.ki.nar.la.
.ra.ram.lu.	.li.nak.li.	.nu.lal.ti.	.tu.mar.la.
.la.lin.ma.	.lu.rat.ma.	.li.mul.ku.	.na.kar.ni.
.ta.kim.la.	.nu.rik.li.	.ni.nal.na.	.ka.tur.la.
.ru.rin.ri.	.ru.rit.ki.	.ti.ril.ma.	.ta.nar.ma.
.ru.mam.ru.	.ka.muk.ta.	.mu.lul.ri.	.ka.kir.ti.
.na.man.la.	.ki.kut.li.	.ku.nul.mu.	.ki.kur.ma.
.ti.kum.lu.	.ta.tit.nu.	.ku.lal.ri.	.tu.kir.na.
.la.kun.la.	.ma.kak.ra.	.ku.tul.ri.	.mi.lir.ma.

6.1.3 Procedure. The procedure was essentially the same as in Experiments 1 and 2, except for the changes described in Sections 4 and 5 above.

6.1.4 Hypotheses. All participants were exposed to the same pattern of liquid neutralization where [l] but not [ɹ] occurred in final position (and neither liquid occurred in word/sentence-medial codas) during the learning phase.

Hypothesis 1: Learning. If the participants in this experiment learn the pattern they are exposed to during the learning phase, their ‘accept’ responses should reflect the pattern in the final condition (7a) in the testing phase. This means that participants in the UtW group will accept test items with sentence-final [l] more frequently than those with sentence-final [ɹ], and participants in the WtS group will accept test items with word-final [l] more frequently than those with word-final [ɹ], showing an effect of the learning set on participants' responses during the testing phase.

Hypothesis 2: Generalization. The learning phase does not provide participants with evidence about which liquids are allowed in word/sentence-medial codas. If participants generalize the pattern they learned from the right edge of the larger domain (word or sentence) to the smaller domain (syllable or word), then the pattern should hold for test items in the medial condition (7b). This means participants in the UtW group will accept utterance-medial, word-final [l] more frequently than they accept utterance-medial, word-final [ɹ], and participants in the WtS group will accept word-medial coda [l] more frequently than they accept word-medial coda [ɹ]. This would show generalization of the learned pattern to the smaller domain.

Hypothesis 3: Syllable as domain. If the syllable and the word are both domains to which phonological patterns can be generalized, then both groups should show evidence of learning and generalization. A difference in performance between the two groups could suggest differences in how utterances, words, and syllables are treated in the grammar. Specifically, if both groups show evidence of learning but only the UtW group shows generalization to medial position, that would suggest that generalization to the syllable is not possible (at least in this AGL experiment) or involves a different process than generalization to the word.

6.2 Results

It was somewhat more difficult than in the preceding experiments to decide whether results from any participants should be excluded, due to the different participant pool and restrictions on prescreening. As discussed above, this experiment was not limited to *native* English speaker participants, but instead to English-dominant participants. Some participants had some undergraduate-level training in phonology, but this did not seem to affect performance.[18]

After excluding only non-English-dominant participants, there were a total of 71 participants (36 WtS and 35 UtW). As in the previous experiments, only the testing phase utterances with coda liquids were included in the analysis. For each participant, there were 18 observations per condition (18 final [ɹ] codas, 18 final [l] codas, 18 medial [ɹ] codas, and 18 medial [l] codas). The proportion of 'accept' responses is presented in Figure 3, by liquid coda consonant, learning group, and position of the target coda in the utterance/word.

[18] Ten participants had taken a number of linguistics classes, including at least one course in phonology. When compared with the other participants, however, there was not a significant difference in their proportion of 'correct' responses. Analysis can be found in Appendix B.

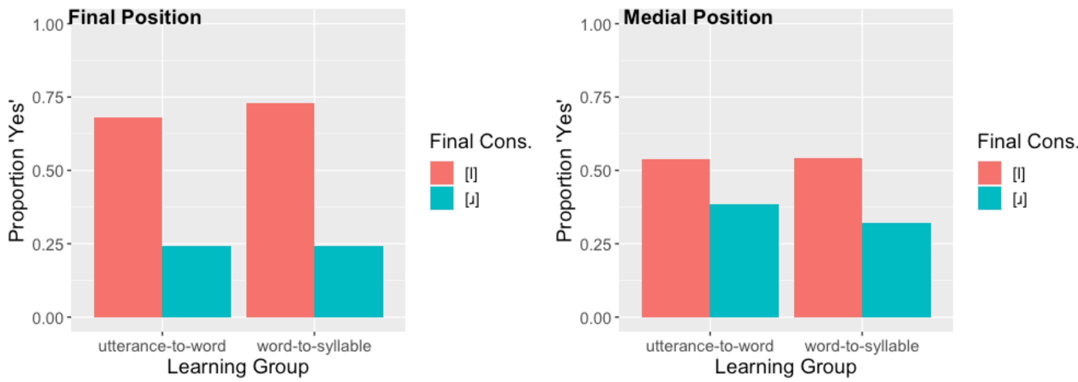

Figure 3: Proportion of 'yes,' or 'accept,' responses by liquid consonant and learning group: (left) final position; (right) medial position (Experiment 3 - All participants).

Qualitatively, there appears to be clear evidence of learning in both groups, with participants accepting substantially more [l] codas than [ɹ] codas in final position. In medial position, there is still a preference for [l] over [ɹ] codas in both groups, though this preference is weaker than the final condition, and also weaker than the preference seen in Experiment 1.

Not all participants, however, showed evidence of having learned the desired pattern. There are a number of reasons this may have occurred (as it did in the prior experiments as well). It was determined that participants who did not have a difference of at least four (out of a possible 18) between the number of [l] codas they accepted and the number of [ɹ] codas they accepted in the final condition did not show a preference for [l] over [ɹ] and, therefore, did not learn the intended pattern from the learning phase. It would, therefore, not be worthwhile to include these participants in any analysis of generalization, as they cannot generalize a pattern they did not learn. After excluding these participants,[19] 30 participants remained in the UtW group and 31 remained in the WtS group. The proportion of 'accept' responses for this smaller set of participants is presented in Figure 4, by liquid coda consonant, learning group, and position of the target coda in the utterance/word.

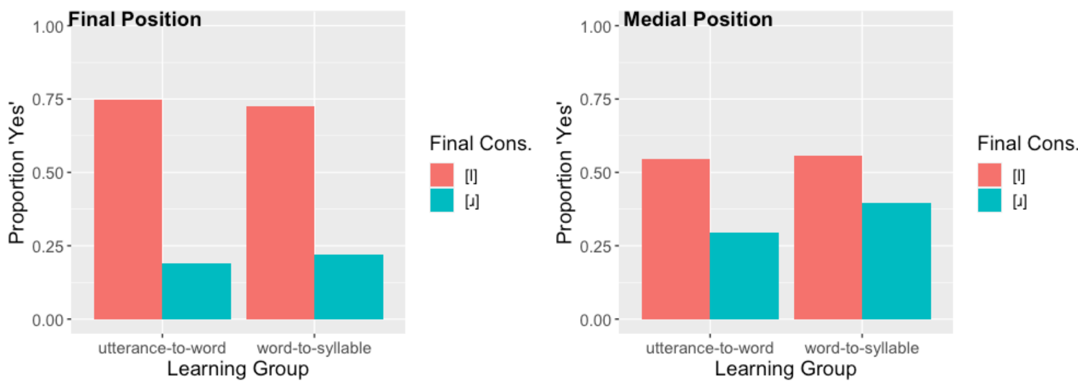

Figure 4: Proportion of 'yes,' or 'accept,' responses by liquid consonant and learning group: (left) final position; (right) medial position (Experiment 3 - Learners).

As in Experiment 1, the data were modeled by means of a mixed effects logistic regression analysis. The dependent variable was the response ('no' or 'yes') to the question, *Does this sentence/word sound like it belongs to the language you have been learning?* 'Yes' was treated as the marked value of the response variable (coded as 1), and 'no' as the default (coded as 0). The fixed effects were the liquid coda consonant (L or *R*), utterance/word position of the target coda (final or *medial*), and learning group (UtW or *WtS*). In each of these factors, the first level was treated as the default value and the second (in italics) as the marked value. Random effects included individual participants and items. Intercepts were included in the model for

[19] Excluding these participants did not change the significant main effects of the regression analysis summarized in Table 9 below.

both of these random effects, as well as a slope for Coda consonant[20] for each participant. The results for the fixed effects are given in Table 9.

	b	*z*	*p*
Intercept	1.2457	6.307	2.84e-10*
Coda consonant ([ɹ])	-3.0106	-10.518	< 2e-16*
Coda position (medial)	-1.0295	-5.222	1.77e-07*
Learning group (WtS)	-0.1021	-0.427	0.669
Coda consonant × Coda position	1.7204	5.980	2.23e-09*
Coda consonant × Learning group	0.1912	0.550	0.582
Coda position × Learning group	0.1454	0.743	0.458
Coda consonant × Coda position × Learning group	0.2198	0.754	0.451

*Table 9: Fixed effects in a logistic regression model of responses (Experiment 3). Significant values ($p < 0.05$) are indicated by *.*

There were significant main effects of Coda consonant and Coda position but not Learning group. There was also significant interaction between Coda consonant and Coda position but there was no significant interaction involving Learning group, suggesting that the UtW group and the WtS group behaved similarly. To further explore these main effects, the data were broken into subsets according to Coda position and Learning group. Table 10 gives the results for four tests, with the same random effects structure as in the previous analysis, and the single fixed effect of Coda consonant.

[20] The model was slightly simplified so that it would converge. A slope for Coda position for each participant was removed.

Coda position	Learning group	*b*	*z*	*p*
final	UtW	-2.8886	-10.387	< 2e-16*
final	WtS	-2.9561	-7.583	3.37e-14*
medial	UtW	-1.3102	-4.856	1.2e-06*
medial	WtS	-0.9029	-4.076	4.57e-05*

*Table 10: Effects of Coda consonant in data subsets defined by Coda position and Learning group (Experiment 3). Significant values (p < 0.05) are indicated by *.*

There was a significant effect of liquid coda consonant ([l] vs. [ɹ]) on responses for both learning groups in the final condition, showing that participants learned the pattern presented in learning phase (that [l] but not [ɹ] may occur in codas). The significant effect of Coda consonant in the medial condition in both groups shows that this pattern was generalized to the smaller domain (the word domain in the UtW group and the syllable domain in the WtS group).

To test whether one of the groups was more successful at learning or generalizing the liquid neutralization pattern, the responses were again recoded in terms of correctness, where a response was treated as 'correct' if it corresponded to the learning set pattern, and incorrect otherwise. This time, for both groups, it was 'correct' to accept coda [l] and to reject coda [ɹ].

	final	medial
UtW	78%	63%
WtS	75%	58%

Table 11: Percentage of correct responses by Learning group and Coda position (Experiment 3).

In both groups, there was a higher percentage of correct responses in the final condition. There was also a higher percentage of correct responses in the UtW group than in the WtS group.

The effect of Coda position was 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 item, with a slope factor for Coda consonant within participant. The fixed effects were Learning group and Coda position. Only Coda position had a significant main effect. Neither Learning group, nor the interaction between Learning group and Coda position was significant, as can be seen in the summary in Table 12.

	b	*z*	*p*
Intercept	1.43847	10.275	< 2e-16*
Coda position (medial)	-0.87895	-6.072	1.27e-09*
Learning group (WtS)	-0.07397	-0.453	0.651
Coda position × Learning group	-0.08530	-0.587	0.557

Table 12: Fixed effects in a logistic regression model of response correctness (Experiment 3).

The significant effect of Coda position suggests imperfect generalization to the utterance/word-medial position. Participants in both groups learned the distribution pattern more successfully in the position for which they had direct evidence of the pattern in their learning sets. While the higher percentage of correct responses in the UtW group than in the WtS group may suggest that generalization to the syllable is more difficult, less natural, or a slower process than generalization to the word level, this effect did not reach significance.

6.2 Discussion

This experiment shows that generalization to the syllable can occur in the same type of experiment as generalization to the word domain.

Results from the UtW group confirm that the learning of an utterance-final distribution pattern and generalization to utterance-medial, word-final position found in Experiment 1 can be replicated with slightly different experiment design. This experiment had a different rhotic ([ɹ], instead of [r]), a slightly different segment inventory overall, a different participant pool, and novel sentences for each item, rather than target words in repeated carrier sentences. Despite these differences, the results still demonstrated both learning and generalization.

Results from the WtS group showed that the syllable is also a domain to which such distributional patterns may be generalized. Because this group was presented with single-word utterances, the pattern to which they were exposed during the learning phase was both word-final and utterance-final. Teasing these apart may be a fruitful area for future research. Importantly, however, the domain to which the restriction on final liquids was generalized was the syllable. Participants believed they were listening to single-word utterances, yet their learned preference for coda [l] over coda [ɹ] carried over to word-internal codas in the testing phase.

The lack of a main effect of Learning group, or any interaction involving Learning group in either responses (Table 9) or correctness (Table 12) shows that the two groups behaved quite similarly in terms of both learning and generalization. It would not be surprising if generalization to the word and generalization to the syllable were slightly different processes. The syllable is, after all, a smaller domain than the word (contained within the word). The results of this experiment, however, do not show a significant difference between the two groups.

While this experiment convincingly presents evidence of generalization to the syllable, it raises a new question: If generalization to the syllable was possible in this experiment, why was it not found in Experiment 2?

Both Experiment 2 and Experiment 3 tested for generalization of a learned phonological pattern from word(/utterance)-final position to syllable-final position. While Experiment 2 did not find such generalization, Experiment 3 did.

It is likely that the strange sound inventory interfered with participant behavior in Experiment 2. Like Experiment 2, Experiment 3 employed English-like [ɹ], but it returned to the more natural type of consonant inventory from Experiment 1 with two stops, two nasals, and two liquids. There was no trill [r] (as in Experiment 1), nor was there a marked fricative [ʒ] (as in Experiment 2).

Like Experiment 2, this experiment did not present stimuli in carrier sentences. Each target item (word/sentence) consisted of a unique combination of three syllables. Utterances in the medial and final conditions were not the same. This means that, compared to Experiment 1, participants heard a greater number of tokens during the learning phase, though they heard some of the target items fewer times. The fact that participants in the WtS group were able to learn the pattern at the word(/utterance) level and generalize it to the syllable level suggests that this aspect of the experiment design is not what caused the null result in Experiment 2. The fact that participants in the UtW group were able to learn the pattern at the utterance level and generalize it to the word level shows that the carrier sentences design was not (at least solely) responsible for the generalization behavior found in Experiment 1.

Accepting that the null result of Experiment 2 was caused by problematic stimuli, one might still question the mechanisms at play in Experiment 3. Adults cannot, and do not, explicitly inform child learners that they will be hearing a series of individual trisyllabic words or three-word sentences. This element of the AGL experiment is quite artificial, and may raise concern about whether the two groups—which performed so similarly—really believed they were doing different tasks. It may be worth digging further into this question, but responses to the exit survey suggest that participants were, indeed, thinking of items as multi-word sentences in the UtW group and individual words in the WtS group.[21] This suggests that generalization to the syllable (as well as generalization to the word in the UtW group) did, in fact, occur.

[21] Participants in the WtS group referred to the items as 'words' in their exit surveys. They did not make any mention of sentences/phrases, word-order, etc. Participants in the UtW group referred to 'sentences' in their responses, mentioning 'words' as subparts of the multi-word items (e.g. "the last word").

7. Discussion and future directions

7.1 Summary

Experiment 1 replicated M&P's findings that participants in an artificial learning experiment can learn a phonological pattern at the right edge of utterances and generalize it to the word level. This may be evidence that learners generalize from larger phonological domains to smaller ones, including prosodic domains such as the foot and the syllable. It is also possible, however, that this type of phonological learning and generalization is affected by the phonology's interactions with morphosyntax and/or semantics. Words may be special targets for learning and generalization because they are semantically meaningful, or because they comprise morphosyntactic units as well as phonological ones. Syllables, on the other hand, are purely phonological, and their status within the phonology is even contentious (see Section 2.2).

The results of Experiment 3, however, show that generalization to the syllable is possible. Participants were presented with a word-final restriction on liquid consonants ([l] but not [ɹ] may occur word-finally), which they generalized to novel word-internal syllables as a restriction on codas ([l] but not [ɹ] may occur in syllable codas). Participants in the word-to-syllable group accepted word-internal coda [l] more frequently than word-internal coda [ɹ], despite a lack of direct evidence about liquid coda goodness during the learning phase of the experiment (due to the poverty of the stimulus design).

The ability to generalize to syllable codas demonstrates that this type of prosodic domain can, in fact, be the target of generalization by analogy (cf. Bermudez-Otero & McMahon 2006, Bermudez-Otero & Trousdale 2012). It also shows that syllables and syllable structure must be available to the grammar on some level for the derivation of such rules. In an exemplar-type model, this could mean that syllables get their own exemplar clouds, or simply that information about syllable structure is contained within other stored representations, such as words.

The null result in Experiment 2 may be worth exploring further in future work. Though it is likely that the strange consonant inventory caused participants to focus on the wrong aspects of the artificial language, this did not prevent them from learning the pattern. They simply failed to generalize to word-internal codas. In addition to the consonant inventories, another difference between Experiment 2 and Experiment 3 is that Experiment 2 presented a pattern of L-to-R neutralization, while Experiment 3 presented a pattern of R-to-L neutralization. Given the findings of Experiment 1, participants should be able to learn and generalize both of these patterns, but there may be differences between the English approximant [ɹ] used in Experiments 2 and 3 and the trill [r] used in Experiment 1. Figuring out why learners might generalize to the syllable in some cases but not others may shed light on potential differences between generalization to the word and generalization to the syllable.

7.2 Future directions

In Experiment 2 and the WtS condition of Experiment 3, stimuli were single-word utterances. This means that the restriction presented in the learning phase was both word-final and utterance-final. The generalization found in Experiment 3 could be described as either word-to-syllable or utterance-to-syllable. This raises the question of whether both of these types of generalization are possible. In other words, must domain-edge phenomena begin at the phrase/utterance level, and must generalization be stepwise, from a larger domain to the next-largest domain, etc.?

It should be possible to test purely word-to-syllable generalization by presenting stimuli in carrier sentences, where only utterance-internal words contain the pattern to be learned/generalized. In this way, word-to-syllable generalization, if it were found, would not also be utterance-to-syllable generalization. Whether these generalizations need to originate at the phrase level, however, is somewhat more complicated to answer.

One type of argument for domain generalization is the presence of a phenomenon at different domain levels in different, closely-related language varieties, suggesting different stages of a multi-step generalization process. In the Mayan language family, for example, both the final sonorant devoicing described in (2) above, as well as a more widespread pattern of final obstruent[22] aspiration can be found in various languages, though the domain in which these occur varies. While

---

[22] These languages tend to contrast a set of glottalized (ejective and implosive) stops and affricates with a set of simple (pulmonic) stops and affricates. It is the simple stops, and sometimes the simple affricates, which undergo final aspiration. See Bennett (2016) for overview.

word-final aspiration is the most common obstruent aspiration pattern in the Mayan family (e.g. Popti', Yucatec Maya), languages such as Tz'utujil and Poqomchi' consistently aspirate word-internal, preconsonantal[23] obstruents (Bennett 2016).

AnderBois (2011) connects word-final devoicing in Yucatec Maya to utterance-final [h]-epenthesis, proposing that both are motivated by the same "final laryngeal strengthening." Warkentin & Brend (1974) describe word-final aspiration in Ch'ol as stronger or "more heavily aspirated" utterance-finally. These utterance-level effects might suggest that both final aspiration and final devoicing originated at the utterance level and generalized to the word level. Then in some varieties, the pattern generalized further, to the syllable level.

It is unclear, however, to what extent this process is stepwise. Do all domain-edge rules begin at the utterance level and generalize to successively smaller domains? Can learners generalize directly from the utterance level to the syllable level without an intermediate word-final stage? Zec & Zsiga (2022) show generalization to the foot level, claiming that the prosodic foot is also a domain of generalization, yet descriptions of final devoicing and aspiration rarely, if ever, make reference to foot structure. This suggests that generalization can skip from the word level to the syllable level, without an intermediate foot stage. It may also be possible, then, to skip directly from an utterance-final phenomenon to a syllable-level generalization.

[23] There is some debate as to whether these are best described in terms of syllable structure (syllable-final) or linear context (pre-consonantal). Bennett (2016) suggests this may vary from language to language.

Testing whether generalization directly from the utterance level to the syllable is possible in an AGL experiment would require multi-word utterances, as well as multisyllabic words. To ensure that the learned pattern is utterance-final but not also word-final, it would need to occur in the last word of a multi-word utterance. At least some words would need to be more than one syllable long with word-internal codas, in order to test for generalization to the syllable, rather than the word level. As stimuli become longer and/or more complex, there may be challenges in terms of the duration of the experiment and the limits of memory. A factor that may contribute to this type of generalization, however, is lexical stress.

7.2.1 Stress Experiment 2 initially intended to test whether the position of lexical stress could help facilitate learning and/or generalization to the syllable. As discussed in Section 2.1, utterance-final position has been shown to be important to language learners. Utterance-final syllables are associated with final lengthening, giving speakers more time to pronounce and sustain sounds and learners more time to perceive these realizations. The lack of immediately following sounds prevents coarticulation and masking, and the recency of utterance-final position makes utterance-final position easier to remember and access.

Final stress might make utterance-final syllables even longer, louder, and/or more memorable, helping learners to perceive and learn utterance-final patterns. On the other hand, stress is associated with strengthening and preservation of contrasts, potentially motivating stressed syllables to resist variation in phonetic realization. If, however, a change does begin to occur in utterance-final position, stress may help motivate generalization specifically to the syllable, rather than the word, since stress is realized on syllables.

To test whether final stress can facilitate learning and/or generalization, one could use the same (or similar) stimuli from Experiment 3, but controlling for stress. All participants would be told they were listening to trisyllabic words, but one group would hear the items with final stress, while the other heard items with penultimate stress. If final stress aids in learning or generalization, the final stress group should be more successful at learning, generalization, or both.

7.3 Conclusion

This study presents empirical evidence of phonological generalization to the syllable domain. This suggests that the syllable is a relevant structure at some levels of memory and grammar, contra to claims that the syllable is not a linguistic unit. I have outlined some areas for future research that may provide further insight into language acquisition, diachronic change, stress, and the syllable.

Appendix A

Participant demographics

Numbers in parentheses show the number of participants reporting a given response.

Experiment 1

Gender	Male (29) Female (16) Agender (1) Nonbinary (1) Decline to state (1)
Ethnicity	White/European (31) (East) Asian (4) Black/African American (4) Hispanic/Latine (3) Mixed/other/decline to state (6)
Native language	English (48)
Dominant language	English (48)
Other languages	None (36) Spanish - beginner (3), intermediate/heritage (1) French - beginner (5) Spanish, German, French - unspecified (1) Korean - beginner (1) Japanese - beginner (1)
Parents' language	English (21) English/Spanish (3) English/Tagalog (2) English/Polish (1) English/Vietnamese (1) English/Punjabi (1)
State of origin	AZ(1), CA(6), FL(3), IL(5), KY(1), LA(1), MA(1), MD(1), MI(2), MO(2), NC(2), NH(2), NY(2), OH(4), PA(1), TX(2), VA(4), WA(1), WI(1), WV(1), IL/WI/FL(1), Unspecified USA(4)

Table 13: Experiment 1 participant demographics.

Experiment 2

Gender	Male (16) Female (8) Decline to state (1)
Ethnicity	White/European (18) Asian (3) Hispanic/Latine (2) Decline to state (2)
Native language	English (24) English & Sindhi (1)
Dominant language	English (25)
Other languages	None (20) Spanish - native (1), intermediate (1), basic (1) Mandarin - near-native, 10y/o (1) Italian - intermediate (1)
Parents' language	English (21) English & Sindhi (1) English & Spanish (1) English & Fujianese (1) Italian & Turkish (1)
State of origin	AL(1), AZ(1), CA(4), GA(1), IA(1), IL(2), IN(1), KY(2), LA(1), MD(2), NJ(1), NY(1), OH(1), PA(1), RI(1), TX(2), VA(1), Unspecified USA (1)

Table 14: Experiment 2 participant demographics.

Experiment 3

Gender	Female (49) Male (15) Nonbinary/other (3) Decline to state (4)
Ethnicity	White (24) Asian (19) Hispanic/Latine (11) Black/African American (2) Mixed/other (9) Decline to state (6)
Native language	English (55) Spanish (9) Vietnamese (2) Armenian(1) Chinese (1) Japanese (1) Urdu (1) English/Hebrew (1)
Dominant language	English (69) English/Spanish (2)
Other (non-English) languages	**None** (17) **French** - intermediate (1) **Hebrew** - basic (1), native (1) **Hindi** - native/advanced (1) **Japanese** - basic (1), intermediate (1), intermediate/advanced (1) **Mandarin** - basic (1), intermediate (2), native but English-dominant (2), early immersion but basic (1) **Russian** - native (1) **Spanish** - basic (3), intermediate (5), advanced (4), native (3), native but English-dominant (1) heritage/advanced (2), heritage/receptive (1), dual immersion school/intermediate (1) **Vietnamese** - native/intermediate (1), first grade fluency (1) More than one other language: **Tagalog** (basic) and **Japanese** (intermediate) (1)

	Spanish (advanced/fluent), **Japanese** (basic), **French** (basic) (1) **Spanish** (intermediate) and **French** (intermediate) (1) **Spanish** (intermediate) and **Tamil** (receptive) (1) **Spanish** (intermediate/advanced) and **Italian** (basic) (1) **Spanish** (intermediate/advanced) and **Vietnamese** (heritage) (1) **Spanish** (native/fluent) and **ASL** (intermediate) (1) **Spanish** (intermediate) and **Cantonese** (heritage/intermediate) (1)
Parents' language	English (29) Cantonese and Mandarin (1) Cantonese and Spanish (1) Cantonese/Taishanese and English (1) Hebrew (1) Hindi (1) Hindi, Bengali, English (1) Hindko, Pashto, Punjabi, Urdu (1) Japanese and English (1) Mandarin and English (2) Russian (1) Spanish (6) Spanish and English (6) Tagalog and English (4) Tamil (1) Vietnamese (2) Vietnamese and English (2)
State of origin	AZ(1), CA(65), CO(2), NY(1) China(1), Japan (1)

Table 15: Experiment 3 participant demographics.

Appendix B

Comparison between participants with/without phonology background (Experiment 3)

Participants who had taken Phonology 1 and other linguistics courses were coded as 'ling major', while those who had taken only introductory-level linguistics classes and no phonology were coded as 'intro'. Percent of 'correct' responses—responses consistent with the pattern from the learning phase—were calculated for each coda position for the two background/experience groups.

	final	medial
ling major	77.8%	62.5%
intro	76.4%	59.9%

Table 16: Percentage of correct responses by Coda position and Phonology experience.

The data were then modeled using a mixed model regression analysis with correctness as the dependent variable and fixed effects of Coda position and Phonology experience. Random effects intercepts were included for participant and item, with a slope factor for Coda position within participant. While there was a main effect of Coda position (as reported above), neither the main effect of Phonology experience nor the interaction of Phonology experience and Coda position reached significance. Neither experience group was more successful at learning or generalization.

	b	*z*	*p*
Intercept	1.37586	11.688	< 2e-16 *
Coda position (medial)	-0.92995	-7.258	3.92e-13 *
Phonology experience (Ling major)	0.13609	0.612	0.541
Coda position × Phonology experience	0.04398	0.223	0.823

Table 17: Fixed effects in a logistic regression model of response correctness (Experiment 3).

These calculations were made after excluding participants who did not show evidence of learning, as described in Section 6.2 above. Two of the excluded participants had taken Phonology 1 and would be classified as 'ling major'. With all 71 English-dominant participants included, the results are essentially the same.

	final	medial
ling major	73.4%	61.1%
intro	73.0%	58.9%

Table 18: Percentage of correct responses by Coda position and Phonology experience (Experiment 3, all English-dominant participants included)

Again, no statistically significant effect of Phonology experience was found on the percent of correct responses.

References

AnderBois, Scott. 2011. Strong positions and laryngeal features in Yukatek Maya. In Suzi Lima, Kevin Mullin & Brian Smith (eds.), *North East Linguistic Society (NELS)* 39.41–54. Amherst, MA: GLSA.

Baird, Brandon. 2014. Acoustic correlates of stress in K'ichee': A preliminary analysis. *MIT Working Papers in Linguistics* 74.21-34.

Barnes, Jonathan. 2006. *Strength and weakness at the interface: Positional neutralization in phonetics and phonology*. Berlin, New York: Mouton de Gruyter.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68.255–278.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67.1-48.

Bennett, Ryan. 2016. Mayan phonology. *Language and Linguistics Compass* 10.469-514.

Bermudez-Otero, Ricardo, and April McMahon. 2006. English phonology and morphology. In Aarts, B. & A. McMahon (eds.), *The handbook of English linguistics* (pp. 382-410). Oxford: Blackwell.

Bermúdez-Otero, Ricardo, and Graeme Trousdale. 2012. Cycles and continua: On unidirectionality and gradualness in language change. In T. Nevalainen, & E. Closs Traugott (Eds.), *The Oxford handbook on the history of English* (pp. 691-720). [55] (Oxford Handbooks Online). Oxford University Press.

Blevins, Juliette. 2006. A Theoretical Synopsis of Evolutionary Phonology, *Theoretical Linguistics* 32:117-65.

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.

Dayley, Jon. 1985. *Tzutujil grammar*, vol. 107 University of California Publications in Linguistics. Berkeley, CA: University of California Press.

Faytak, Matthew, Suyuan Liu, and Megha Sundara. 2020. Nasal coda neutralization in Shanghai Mandarin: Articulatory and perceptual evidence. *Laboratory Phonology* 11.1-29.

Finley, Sara, and William Badecker. 2009. Artificial language learning and feature-based generalization. *Journal of Memory and Language* 61.423–437.

Hayden, Rebecca. 1950. The relative frequency of phonemes in General-American English, *Word* 6.217-223.

Hock, Hans Henrich. 1991. *Principles of historical linguistics*. 2nd edn. Berlin & New York: Mouton de Gruyter.

Hyman, Larry. 1978. Word demarcation. In J. H. Greenberg (ed.) *Universals of Human Language*. Vol. 2. Phonology. Stanford University Press, pp. 447-470.

Ito, Junko, and Armin Mester. 2009. The extended prosodic word. In Kabak, Baris, and Jaent Grijzenhout, eds. *Phonological Domains: Universals and Derivations.* The Hague, The Netherlands: Mouton de Gruyter. 135-194.

Johnson, Elizabeth K., Amanda Seidl, and 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.

Johnson, Keith. 1996. Speech perception without speaker normalization. in Johnson, K. and Mullennix (eds.) *Talker Variability in Speech Processing*. San Diego. Academic Press.

Lombardi, Linda. 1995. Laryngeal Neutralization and Syllable Wellformedness. Natural Language & Linguistic Theory, 13(1), 39–74.

Myers, Scott. 2012. Final devoicing: Production and perception studies. In T. Borowsky, S. Kawahara, T. Shinya & M. Sugahara (eds.) *Prosody matters: Essays in honor of Elisabeth Selkirk* (pp. 148-180). London: Equinox.

Myers, Scott, and Jaye Padgett. 2014. Domain generalisation in artificial language learning. *Phonology* 31.399-433.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (ed.), *The Production of Speech*. New York: Springer. 189-216.

Ohala, John, and Haruko Kawasaki-Fukumori. 1997. Alternatives to the sonority hierarchy for explaining the shape of morphemes. In *Language and its ecology: Essays in memory of Einar Haugen*, ed. by Stig Eliasson and Ernst H. Jahr, 343-365. Berlin: Mouton de Gruyter.

Padgett, Jaye. 2015. Word-edge effects as overphonologization of phrase-edge effects. In U. Steindl, T. Borer, H. Fang, A. García Pardo, P. Guekguezian, B. Hsu, C. O'Hara, & I. Chuoying Ouyang (Eds.), *Proceedings of the 32nd West*

Coast Conference on Formal Linguistics (pp. 41-53). Somerville, MA: Cascadilla Proceedings Project.

Pierrehumbert, Janet. 2000. Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, J. & P. Hopper (eds.), *Frequency effects and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Prince, Alan, and Paul Smolensky. 1993. *Optimality Theory: constraint interaction in generative grammar*. Ms, Rutgers University & University of Colorado, Boulder. Published 2004, Malden, Mass. & Oxford: Blackwell.

Pye, Clifton. 1983. Mayan telegraphese: Intonational determinants of inflectional development in Quiché Mayan. *Language* 59.583-604.

Steriade, Donca. 1999. Alternatives to the syllabic interpretation of consonantal phonotactics. In *Proceedings of the 1998 Linguistics and Phonetics Conference,* ed. by Osamu Fujimura, Brian Joseph, and Bohumil Palek, 205-242. Prague: The Karolinum Press.

Sundara, Megha, Katherine Demuth, and 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.

Vaux, Bert. 2003. Syllabification in Armenian, Universal Grammar, and the Lexicon. *Linguistic Inquiry*, 34(1).91-125.

Warkentin, Viola, and Ruth Brend. 1974. Chol Phonology. *Linguistics* 12(132).87–102.

Wax Cavallaro, Maya. 2021. [SG] and final consonant allophony in Tz'utujil. In R. Bennett, R. Bibbs, M.L. Brinkerhoff, M.J. Kaplan, S. Rich, N. Van Handel & M. Wax Cavallaro (eds.), *Proceedings of the 2020 Annual Meeting on Phonology*. Washington, DC: Linguistic Society of America.

Wax Cavallaro, Maya. 2022. Cross-linguistic patterns of sonorant devoicing as evidence for [spread glottis] and [voice] in sonorants. Santa Cruz: University of California, Santa Cruz qualifying paper.

Westbury, John R., and Patricia A. Keating. 1986. On the naturalness of stop consonant voicing. *JL* 22.145-166.

Wetzels, W. Leo, and Joan Mascaró. 2001. The Typology of Voicing and Devoicing. *Language* 77.207–244.

White, James. 2013. Evidence for a learning bias against saltatory phonological alternations. *Cognition* 130.96–115.

Wilson, Colin. 2003. Experimental investigation of phonological naturalness. *WCCFL* 22.533–546.

Wilson, Colin. 2006. Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science* 30. 945–982.

Zec, Draga, and Elizabeth Zsiga. 2022. Tone and stress as agents of cross-dialectal variation: The case of Serbian. In Kubozono, H., Ito, J & A. Mester (eds), *Prosody and Prosodic Interfaces*.

Zehr, Jérémy., and Florian Schwarz. 2018. *PennController for Internet Based Experiments (IBEX).*