Open Library of Humanities

Impacts of direction and morphological locus of control on learnability of sibilant harmony

Jenna T. Conklin*, Linguistics Department, Carleton College, Northfield, MN, USA, jconklin@carleton.edu

Elliot Stork, Linguistics Department, University of Michigan, Ann Arbor, MI, USA

Ruken Bastimar, Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, MN, USA

Xueqi He, Independent researcher, USA

*Corresponding author.

A notable cross-linguistic gap exists in the typology of consonant harmony: Stem-controlled and suffix-controlled systems are known, but prefix-controlled consonant harmony remains unattested. To explore this gap, an artificial grammar learning (AGL) study was conducted, in which participants were passively trained and then tested on one of four possible sibilant harmony patterns, differing by direction and morphological locus of control, with English and Spanish native speakers. The effect of target-trigger distance was also tested by varying the stem length from one to four syllables. For both speaker groups, prefix-controlled harmony was no less learnable than the other harmonies tested. The effect of target-trigger distance was notably different across languages, with Spanish speakers showing no negative impact of target-trigger distance while English speakers' performance declined as target-trigger distance increased. Overall, the results indicate that the typological lack of prefix-controlled consonant harmony is not due to a lack of learnability and thus must be attributed to some other cause.

1. Introduction

In the realm of phonological theory, a great deal of ink has been spilled debating the best way to account for iterative, long-distance processes, such as consonant and vowel harmonies. Iterative processes, which apply to a series of targets in a repetitive fashion often resembling movement away from an initiating segment (the trigger), are notoriously difficult to explain under classical Optimality Theory (OT), which demands that all phonology be accounted for in a single step, with no repeated application permitted. The attempt to account for iterativity is one of the strongest driving forces behind the shift from classical OT to post-OT frameworks of various kinds, including stochastic, stratal, and serial models such as Stochastic Optimality Theory (Boersma, 1997), Noisy Harmonic Grammar (Boersma & Pater, 2008), Harmonic Serialism (McCarthy, 2009), Stratal OT (Bermúdez-Otero, 2003; Tolskaya, 2008), and various maximum entropy grammars (Goldwater & Johnson, 2003; Hayes & Wilson, 2008). These diverse approaches each rely on some combination of repeated application, probabilistic weighting of constraints or cells, or stratification of the analytical mechanism to model iterative harmony processes, and each argues for supremacy over its peers based on its ability to avoid predicting often wholly unnatural patterns, such as the well-known majority rules or sour grapes pathologies (Finley, 2009). Meanwhile, real-world typological distributions, including typological gaps, have received little attention, despite their potential to both shed light on the fundamental nature of harmony and aid in fine-tuning promising models to real-world data. In particular, a better understanding of the nature of unexpected typological gaps can provide insight into the reasons underlying the gap.

The most salient typological gap related to harmony involves prefix-controlled consonant harmony, which has not been documented in any natural language despite strong similarities to attested processes (Hansson, 2010). Prefix-initiated vowel harmonies are also exceedingly rare, although not unattested (McCollum & Essegbey, 2020); the distinction between prefix-initiated and prefix-controlled harmony is discussed in Section 1.2. In this paper, we will use the term "prefix-controlled" to refer to instances of harmony in which a prefix exercises harmonic control over a stem or other non-prefixal morpheme, setting aside instances of harmony in which both the target and trigger are prefixes. The studies presented in this paper utilize an artificial grammar learning (AGL) paradigm to demonstrate that the cross-linguistic lack of prefix-controlled consonant harmony is not due to a deficit of learnability, indicating that the typological gap must be accounted for via phonological theory, constraints on the evolution of language over time, or cognitive restrictions unrelated to learning.

1.1. Establishing causes of typological gaps

A central question in phonological typology is to what degree typological information represents an underlying cognitive reality and where it may contain misleading accidental gaps. Accidental gaps are aspects of grammar that remain unattested not because they are unattainable in

natural language, but simply because no currently documented languages happen to employ an otherwise possible piece of phonology (Crystal, 2003; Finley & Badecker, 2008; Gordon, 2007; McMullin, 2016). Artificial grammar learning (AGL) studies are a well-established paradigm for determining whether a particular linguistic pattern is learnable by naïve speakers. In an AGL study, participants are exposed to a simple artificial language created by researchers during a familiarization or training phase and are then tested on the knowledge they gained in a second phase of the study. This approach allows researchers to control many sources of variation that confound studies of natural languages by focusing learners' attention narrowly on a phenomenon of interest, thus accelerating the learning process. While the conclusions that can be drawn in this manner are necessarily limited in certain ways, the approach enjoys a fair degree of cognitive realism: The neurological mechanism for artificial language learning appears to be the same as that used in natural language acquisition (Finley, 2017; Friederici, Steinhauer, & Pfeifer, 2002; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012). Friederici et al. (2002) trained participants on an artificial language and found they had similar neurophysiological responses, as codified in event-related potential patterns, when processing the language they were trained on, as they had to their native language, while learners who were not trained on the artificial language did not process artificial language input in a manner similar to their native language. Morgan-Short et al. (2012) conducted a similar study, also testing the effects of explicit and implicit training styles in the artificial language learning session. Although subjects' performance did not differ by training style, there was a significant impact on the neurophysiological response measured between the implicit and explicit training groups, which they interpret as evidence that implicit training (learning the artificial language through meaningful, contextualized exposure rather than grammatical instruction) mirrors the acquisition process of natural language. Taken together, these results provide strong support for the use of AGL paradigms to test the learnability of possible linguistic patterns.

However, results from AGL studies should be interpreted with appropriate contextualization and recognition of their limitations. It is not impossible, for instance, that an AGL study might reflect biases rooted in participants' first language (L1), just as a second language study might. Similarly, the finding that participants did not learn a pattern in an artificial language is not as strong a piece of evidence as that they did learn it, for the simple reason that the amount of exposure needed to learn a pattern is not predetermined. Too much exposure may also induce participants to learn even unnatural patterns; well-designed AGL studies account for this possibility by testing the generalizability of subjects' knowledge or comparing the relative learnability of the phenomenon of interest and some other baseline (for instance, a variant widely acknowledged to be learnable). Despite these caveats, AGL work presents many opportunities for discovery not available under other experimental paradigms. Of particular interest in the present study, AGL studies allow researchers to measure gradience in learnability and to expose learners to patterns unattested in natural language. Thus, the results of AGL

work should parallel typological frequency insofar as typology reflects learnability, but they necessarily obscure typological realities that are predicated primarily on restrictions unrelated to learning.

1.2. Past AGL work on vowel harmony

A significant proportion of AGL studies on harmony to date have focused on vocalic harmony, as summarized in Finley (2017). Typologically unattested vowel harmony patterns have been found to suffer from a lack of learnability, demanding far greater amounts of time for naïve learners to acquire than linguistically common patterns or simply failing to be learned at all. One example of this is "majority rules" vowel harmony, in which the quality of a target vowel is determined not by linear order or feature values but by the sheer number of trigger vowels of that quality that are present (Finley & Badecker, 2008). Similarly, Wilson (2003) demonstrated that a natural assimilatory pattern was easier for subjects to learn than a phonetically unnatural cross-featural assimilation pattern (in which, for example, a segment selected the feature value for [nasal] based on the feature value of [dorsal] of the preceding segment). AGL investigations of long-distance vocalic processes have also demonstrated that typological trends reflect relative learnability with regard to the behavior of transparent and opaque vowels (Finley, 2015) and that learners will readily infer a local assimilation process from a non-local dependency, while failing to infer a long-distance dependency from a local one (Finley, 2011). Not all AGL investigations of vowel harmony have concluded that typologically dispreferred patterns suffer from a learning bias in a controlled setting, however: Studies comparing learning of harmony to disharmony have had contradictory results, with some finding no difference between treatment groups (Pycha, Nowak, Shin, & Shosted, 2003) and others finding a bias in favor of the naturally-occurring harmony pattern (Martin & Peperkamp, 2020; Martin & White, 2021).

When it comes to the directionality of vowel harmony processes cross-linguistically, root-controlled and regressive harmonies dominate the landscape (Hyman, 2002). Baković (2000) goes so far as to assert that the directionality of all attested vowel harmonies can be characterized either through morphological control (that is, that progressive harmony arises as a result of stem control in purely suffixing languages) or a dominant-recessive pattern (in which one feature value is dominant and triggers harmony). However, the existence of multiple [+ATR]-dominant harmonies exhibiting purely regressive directionality, such as those found in Karajá, Assamese, and Gua (Mahanta, 2007; Obiri-Yeboah & Rose, 2022; Ribeiro, 2002), suggests that this view does not encompass the full range of vowel harmony processes. Later work by Hyman (2002) proposes that the directionality of vowel harmony will generally either be related to prominence or be purely directional, but that if it is directional, it will be regressive. Under this view, stress-controlled, stem-controlled, and other prominence-related harmony types will be widespread, as will purely regressive harmony, but purely progressive harmony will occur but rarely, as in the limited Kinande prefix-controlled harmony presented by Hyman (2002).

More recent work has expanded the notion of prominence to include both metrical and edge prominence (McCollum & Essegbey, 2020). McCollum and Essegbey further introduced a previously undocumented progressive vowel harmony to the literature in their 2020 summary of Tutrugbu. In this language, labial harmony is controlled in a progressive manner by prefixes, making Tutrugbu labial harmony a rare instance of prefix-controlled vowel harmony. However, as McCollum and Essegbey make clear, Tutrugbu does not represent a case of purely progressive harmony, concurring with previous analyses that predict that purely progressive harmony will never occur. Instead, progressive directionality in Tutrugbu arises epiphenomenally as a consequence of the phonological prominence of the initial syllable in Tutrugbu, and non-initial prefixes are not capable of triggering harmony (McCollum & Essegbey, 2020). Thus, ample evidence exists that the directionality of vowel harmony is limited to purely regressive harmonies and systems whose direction is determined by some prominent trigger (i.e., morphological control, edge prominence, or metrical prominence). Under this schema, apparently prefix-controlled harmony is possible in cases where some form of prominence lends harmonic control to prefixes, but cases of pure prefix control – or any purely progressive harmony – remain unattested and unexpected.

Of particular interest to the current work are AGL investigations of purely progressive affix-controlled (that is, prefix-controlled) vowel harmony, which is only attested in natural languages in cases that are better characterized as prominence-controlled rather than purely progressive. Finley and Badecker (2009) investigated learning biases in vowel harmony dependent on the direction of harmonization and demonstrated that prefix-controlled vowel harmony was significantly less learnable than a suffix-controlled alternative. When it came to stem-controlled harmony, learners in their study assumed that stem-controlled harmony applied to prefixes even when they had only seen it in action with suffixal targets, and they similarly assumed that stem-controlled patterns acting on prefixal targets should generalize to suffixal targets. This finding suggests that despite the typological prevalence of right-to-left (regressive) vowel harmony, there is no cognitive default direction for stem-controlled vowel harmony. However, when it came to affix-controlled harmony, a significant bias emerged: Subjects in the suffix-control group acquired the harmony pattern and were able to generalize to prefix-controlled harmony, but those trained on a prefix-controlled pattern did not succeed in mastering or generalizing the pattern. Thus, in addition to being extremely rare in the real world, prefix-controlled vowel harmony appears to suffer a serious learning bias within an AGL paradigm. The present study intends to discover whether a similar lack of learnability characterizes prefix-controlled consonant harmony.

1.3. Consonant harmony

While the learnability of vowel harmony has been well explored with regard to directionality and locus of control, less thorough investigation has been conducted for consonant harmony. Given the fundamental differences in the phonetic bases underlying vowel and consonant harmonies, it is not reasonable to expect that the learning biases present for vowel harmony will necessarily

apply to consonant harmony as well. Vowel harmony is broadly agreed to arise from vowel-to-vowel coarticulation; although it appears to be a long-distance process when viewed through a segmental lens, at the gestural level it is fundamentally local (Majors, 2006; Ohala, 1994). By contrast, consonant harmony is often phonetically non-local (Heinz, 2010) and has its roots in speech planning rather than gestural overlap (Hansson, 2010). The formal mechanisms used to characterize them often differ as well, with vowel harmony and some consonant harmonies analyzed as a form of feature spreading while other forms of consonant harmony are identified as instances of agreement (Gafos, 2021). Furthermore, where vowel harmony sometimes involves consonants when it affects features that can apply to both consonants and vowels, consonant harmony typically does not involve vowels, as the minor place features (such as [sibilant]) that are commonly harmonized cannot be realized vocalically (Finley, 2017). This non-involvement of intervening segments is the critical difference between spreading and agreement.

1.4. Typology of consonant harmony in natural languages

While consonant harmony is attested for a number of features, including [dorsal], [nasal], and [lateral], sibilant harmony, which typically involves a binary distinction between alveolar and postalveolar fricatives or affricates, is by far the most common, appearing in multiple unrelated language families around the globe (Hansson, 2010). Sibilant harmony often operates at a distance, with vowels and other consonants acting as transparent segments and numerous transparent segments separating target and trigger as a matter of course. The range and nature of harmonizing contrasts vary widely across languages, but many require a match in the posture and location of the tongue tip and blade posture (Hansson, 2010). A few salient typological generalizations on the directionality and locus of control of consonant harmony processes can be made.

Most consonant harmony processes either operate on a principle of stem control or are limited to root-internal co-occurrence; among the purely directional systems, only right-to-left processes are attested. The few potential left-to-right consonant harmonies have other confounding issues, like morpheme interaction or structure, although some argue that Tiene or Teralfene Flemish may constitute an example of pure left-to-right consonant harmony (Hansson, 2010). One limited form of prefix control can be observed in Navajo sibilant harmony; however, as the targets of this Navajo prefix control are themselves prefixes and not stems (Hansson, 2010, pp. 150–151), we do not consider this to constitute an effective counterexample to the broader typological generalization. As the only potential examples of prefix control can be explained by this constellation of properties, true progressive affix-controlled consonant harmony (i.e., prefix-controlled consonant harmony) remains unattested (Finley & Badecker, 2009; Hansson, 2010).

1.5. Effects of target-trigger distance in natural language

As mentioned above, consonant harmony is often a long-distance process with a substantial number of transparent segments between the trigger and the target phonemes (Finley, 2011, 2017; Heinz, 2010; K. J. McMullin, 2016). However, the application of consonant harmony is not as uniform as phonological descriptions often imply: In many languages, the likelihood of harmony applying decreases as the length of the transparent string, also known as target-trigger distance, increases. This is referred to as distance-based decay by Zymet (2014). Some debate has arisen as to the best way to quantify target-trigger distance: Zymet (2014) found that the number of transparent syllables is a more effective predictor of when distance-based decay will apply than the number of transparent segments for all processes he surveyed aside from retroflex assimilation. Distance-based decay has been documented for labial dissimilation in Malagasy, liquid dissimilation in Latin and English (Zymet, 2014), vowel harmony in Hungarian and Uyghur (Mayer, 2021), and labial vowel harmony in Tatar (Conklin, 2015), suggesting that distance-based decay is not a language-specific phenomenon, but a fixture of long-distance processes in general.

While distance-based decay is attested in many natural languages, little is known of the mechanism behind it. Specifically, does the increase in target-trigger distance induce greater variability in the implementation of harmony simply as a result of processing load or short-term memory limitations, or is the increased variability encoded in the speaker's linguistic grammar? Because the relevant data is largely limited to production, there is little foundation for approaching questions about the nature of distance-based decay. The present study aims to extend understanding of how learners acquire and implement long-distance dependencies over a range of target-trigger distances by testing the effect of target-trigger distance on grammaticality judgments in an AGL paradigm. By approaching the topic in an AGL setting rather than through production data, it is possible to forego analysis of production-based performance altogether and access learners' linguistic competence more directly. This provides a complementary view of distance-based decay relative to the work that has already been carried out, which has largely focused on distance-based decay in the production of speech rather than acceptability gradients in the perception grammar (see, e.g., Berkson, 2013; Gordon, 1999; Zymet, 2014).

1.6. Past AGL work on consonant harmony

A primary focus of past AGL work on consonant harmony has been the role of locality in learnability, where it has generally been found that non-local dependencies are readily generalized to local ones, while local dependencies do not imply non-local ones (see, e.g., Finley, 2011, 2012; McMullin, 2013). Other work has identified gaps in learnability between attested and unattested patterns, such as Lai's (2015) study of First-Last assimilation or Martin and

Peperkamp's (2020) study demonstrating the stark difference in learnability between harmony and disharmony. Several of these studies have incidentally illustrated basic facts about the learnability of consonant harmony weighed by direction and morphological locus of control: Progressive stem-controlled sibilant harmony was shown to be learnable by Finley (2011, 2012), while regressive affix-controlled consonant harmony (i.e., suffix-controlled consonant harmony) was learned by participants in McMullin's (2013) study. (As discussed in depth in Section 1.2, we do not view directionality or locus of control as theoretical primitives of naturally occurring harmony systems, but instead recognize with Hyman [2002], McCollum and Essegbey [2020] and many others the underlying complexity of directionality in harmony writ large. However, for simplicity's sake, in our experimental treatment of directionality we will refer to direction and locus of control as binary variables reflecting our experimental design.) No past studies have directly compared the relative learnability of all four logically possible combinations of Direction and Locus of Control, and, to our knowledge, no previous work has tested the learnability of prefix-controlled consonant harmony – the unattested process that we are most interested in. The aim of the present study is to directly compare the learnability of progressive affix-controlled (PAC), regressive affix-controlled (RAC), progressive stem-controlled (PSC), and regressive stem-controlled (RSC) sibilant harmonies to determine whether learnability can account for their relative typological prevalence and the cross-linguistic lack of prefix-controlled consonant harmonies.

2. Study 1 – English

2.1. Methods

2.1.1. Participants

One hundred and three English monolinguals were recruited via Prolific (https://www.prolific.co) and completed the study through Gorilla Experiment Builder (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020). One was excluded due to reported internet connectivity issues interrupting the task. Of the remaining 102 participants (42 male; 60 female; aged 20–69; M = 37.91; SD = 13.2), all were born and raised in the U.S. and reported having spent no significant time in a non-English-speaking country; none reported significant knowledge of a language other than English. Approximately 85% of participants self-identified as White, 6% as Asian, 6% as Black or African American, 2% as two or more races, and 1% as Native Hawaiian or other Pacific Islander. All participants but one had at least a high school diploma at the time of completion. Data collection for each participant took less than one hour. All study procedures were reviewed by the Institutional Review Board at Carleton College (Protocol 2020-21 1216). Prior to participation, all procedures were explained and participants provided informed consent electronically. Participants were compensated US$8.25 for their time.

2.1.2. Stimuli

Stimuli were constructed of CV syllables using the sounds /s, ʃ, p, t, k, b, d, g, n, l, i, e, a, o, u/. The unaffixed forms varied in length from one to four syllables, and all stems contained either /s/ or /ʃ/. Primary stress placement was always penultimate. Sample stimuli are provided in **Table 1**, with a full list available in the Appendix. Affixes took the forms /su/ (past) and /ʃi/ (future); in the affix-controlled groups, the affixes surfaced faithfully, while in stem-controlled groups, they alternated between the allomorphs [su~ʃu] and [si~ʃi] respectively, according to the consonant harmony rule in effect for the group in question. Stimuli were recorded by a female native speaker of American English familiar with the IPA but not involved in the design of the study. Recording was conducted in a sound-attenuated room with a Shure KSM32 cardioid condenser microphone and a Focusrite Red 8Pre audio interface using Adobe Audition, and recordings were digitized at 48 kHz. Recordings were intensity-normalized to 70 dB using a Praat script (Winn, 2013).

	Training sample	Testing sample
	verb, past tense, future tense	Which word **belongs** in the language?
Progressive Stem Control (PSC)	s<u>o</u>, ˈs<u>o</u>-su, ˈs<u>o</u>-si ˈl<u>osu</u>, lo'<u>su</u>-su, lo'<u>su</u>-si ˈl<u>oʃu</u>, lo'<u>ʃu</u>-ʃu, lo'<u>ʃu</u>-ʃi	s<u>o</u>: **s<u>o</u>su,** ˈs<u>o</u>ʃu ˈl<u>osu</u>: lo'<u>su</u>ʃu, **lo'<u>susu</u>** ˈl<u>oʃu</u>: **lo'<u>ʃuʃi</u>,** lo'<u>ʃu</u>si
Progressive Affix Control (PAC)	s<u>o</u>, ˈsu-s<u>o</u>, ˈʃi-<u>ʃo</u> ˈl<u>osu</u>, su-ˈl<u>osu</u>, ʃi-ˈl<u>oʃu</u> ˈl<u>oʃu</u>, su-ˈl<u>osu</u>, ʃi-ˈl<u>oʃu</u>	s<u>o</u>: **sus<u>o</u>,** ˈsus<u>o</u> ˈl<u>osu</u>: su'l<u>oʃu</u>, **su'l<u>osu</u>** ˈl<u>oʃu</u>: **ʃi'l<u>oʃu</u>,** si'l<u>oʃu</u>
Regressive Stem Control (RSC)	s<u>o</u>, ˈsu-s<u>o</u>, ˈsi-s<u>o</u> ˈl<u>osu</u>, su-ˈl<u>osu</u>, si-ˈl<u>osu</u> ˈl<u>oʃu</u>, ʃu-ˈl<u>oʃu</u>, ʃi-ˈl<u>oʃu</u>	s<u>o</u>: **sus<u>o</u>,** ˈʃus<u>o</u> ˈl<u>osu</u>: su'l<u>osu</u>, **ʃu'l<u>osu</u>** ˈl<u>oʃu</u>: **ʃi'l<u>oʃu</u>,** si'l<u>oʃu</u>
Regressive Affix Control (RAC)	s<u>o</u>, ˈs<u>o</u>-su, ˈ<u>ʃo</u>-ʃi ˈl<u>osu</u>, lo'<u>su</u>-su, lo'<u>ʃu</u>-ʃi ˈl<u>oʃu</u>, lo'<u>su</u>-su, lo'<u>ʃu</u>-ʃi	s<u>o</u>: **s<u>o</u>su,** ˈʃosu ˈl<u>osu</u>: lo'<u>ʃusu</u>, **lo'<u>susu</u>** ˈl<u>oʃu</u>: **lo'<u>ʃuʃi</u>,** lo'<u>su</u>si

Table 1: Sample stimuli. Underlining indicates the portion of the word portrayed as stem. Bolding indicates the correct answer in the Testing column. Primary stress assignment is marked with [ˈ] and is penultimate. A full list of stimuli is available in the Appendix.

2.1.3. Procedure

Participants were randomly assigned to one of four testing groups. Each group was exposed to a slightly different version of a sibilant harmony rule, namely either Regressive Affix Control

(RAC; n = 25), Progressive Affix Control (PAC; n = 26), Regressive Stem Control (RSC; n = 25), or Progressive Stem Control (PSC; n = 26). Regardless of group, participants completed a passive training task, a two-alternative forced-choice task with corrective feedback to evaluate their learning, and finally a short demographic questionnaire. Before and after the 2AFC task, participants completed a short (28 trial) judgment task; this task was intended as a pilot of its task design, and results are not reported here. Before beginning the study, participants completed a short task ensuring they were complying with instructions to wear headphones. The headphone check utilized antiphase tones to determine whether participants were using headphones or an external speaker, similar to the task described by Woods, Siegel, Traer, and McDermott (2017).

2.1.3.1. Training task

In the passive training task, participants were told they would be learning a new language and were asked to listen to sets of three words that included a verb stem and two affixed forms: A past tense and a future tense. They were instructed, "You do not need to memorize individual words; simply listen to the way words sound in this language." All instructions were shown in a black sans serif font on a solid white screen. Stimuli were presented auditorily only; at no point did participants encounter any orthographic representation of the artificial language. Periodic attention checks requiring participants to type a response to a simple question (unrelated to the substance of the training task) were used to ensure participants were actively partaking in the task as it proceeded.

The stimuli for the training task consisted of 12 minimal pairs (24 stems in all) with a CV syllable structure, varying in length from one to four syllables. Each stem contained exactly one sibilant (either [s] or [ʃ]). The training task consisted of four repetitions of 24 unique trials, presented in four blocks and randomized within each block; a trial consisted of a stem and two affixed forms, one with /su/ and the other with /ʃi/, with a 0.5 second pause between each form. This task included only correctly-affixed forms according to the harmony rule for that group. Participants completed an attention check and a thirty-second break after each block.

2.1.3.2. Judgment task

After the training session, subjects completed a brief judgment task designed to measure their understanding of the harmony pattern as well as their confidence in their ability to identify words belonging to the artificial language. The judgment task was presented to participants twice: once before and once after the 2AFC testing task. They were first asked to listen to a word and judge how well it fit into the language they have learned. Then they were asked to rate their confidence in their answer to the first question. Each judgment task had 28 trials. In this task, participants encountered both correctly and incorrectly affixed forms. This task was included as an exploratory task piloting its design, and the results are not reported here.

2.1.3.3. 2AFC task

The primary means of assessing subjects' learning was a two-alternative forced-choice (2AFC) task. The 2AFC task included all 24 training stems as well as 24 novel stems; each subject heard each stem twice, once with the harmonic form as the first alternative and once with the harmonic form as the second alternative, for a total of 96 trials. Subjects completed an attention check similar in format to a 2AFC trial and a thirty-second break every 32 trials. Before beginning the 2AFC task, participants completed an ungraded three-trial practice task to ensure they were familiar with the format of the task and comfortable using the specified keyboard keys to respond.

Each testing trial consisted of three forms: A verb stem and two possible affixed forms, one harmonic and one disharmonic, as illustrated in **Table 1**. Participants were instructed to select the word they believed was correct based on the stem that was provided. If they did not respond within five seconds, the experiment moved on to the next trial. Subjects received corrective feedback after each trial; feedback consisted of text reading, "Your answer is correct/incorrect!" with an accompanying red cross mark or green check as a visual aid.

2.1.4. Analysis

Responses to the 2AFC task were coded as either correct or incorrect. Due to the binary nature of the response variable, a binomial generalized logistic regression mixed model was used to analyze the effects of Group (treatment coded), Trial, and Number of Intervening Syllables. Models included a random effect for Subject, as well as some interaction terms and effects of the biphone and/or phoneme probabilities of each test item. Models were carried out in *lme4* (Bates, Mächler, Bolker, & Walker, 2015) in R version 4.3.1 (R Core Team, 2022). P-values were generated using *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2015).

2.2. Study 1 Results

The central question of this study was whether naïve subjects would learn a sibilant harmony pattern in an artificial language more successfully depending on the type of pattern (progressive affix-controlled, regressive affix-controlled, progressive stem-controlled, or regressive stem-controlled). Secondarily, it considers whether the distance between target and trigger sibilants, parametrized as the number of intervening syllables (0, 1, 2, or 3), impacted subjects' performance. An initial examination of the data revealed that the progressive stem-controlled group had the highest proportion of correct responses, and that the two progressive groups had more successful learners than the two regressive groups, as shown in **Figure 1**. To determine the number of successful learners in each group, the binomial distribution with an alpha-level of 0.01 was used to compute the percent of responses that must be correct in order to differ significantly from 50% given the total number of trials (n = 96): 60%. Disparities among groups were driven partly, but not wholly, by differences in the number of successful learners in each group; this is immediately

visible in **Figure 1**, which shows that a greater number of participants performed at or near chance in the two regressive groups than in the two progressive ones. Furthermore, when only successful learners were examined, the progressive stem-controlled group continued to exhibit the best performance and the two regressive groups the worst. However, this simple division does not provide a full picture of the findings of this study.
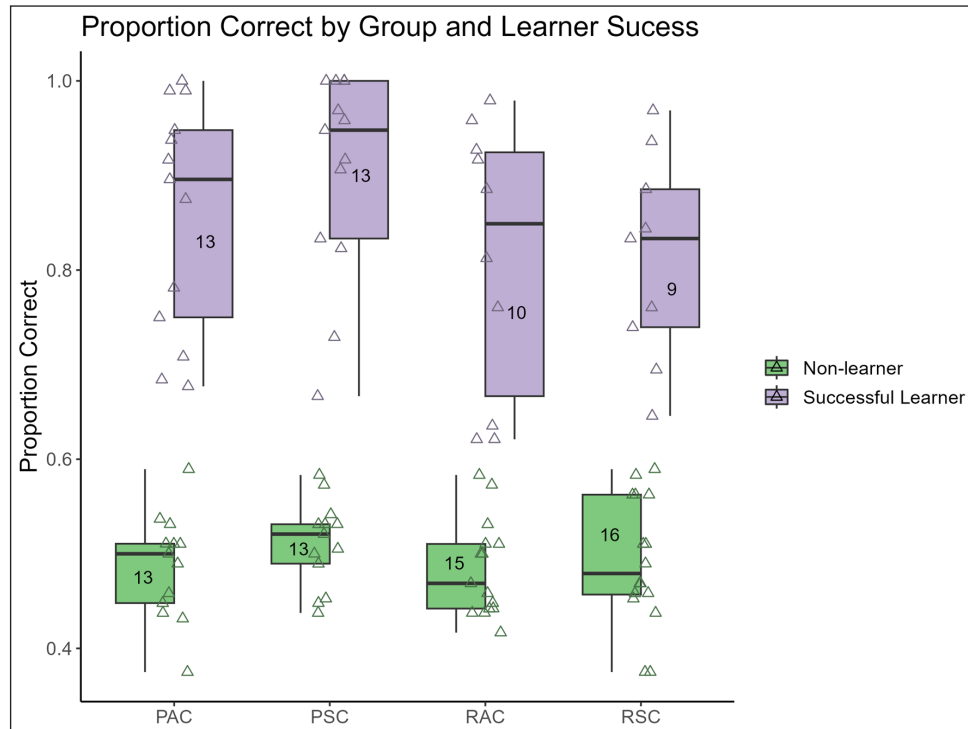

Figure 1: Proportion correct by group, with learners divided by ultimate success (greater or less than 60%) for English speakers. Annotations display the number of learners in each box. Each triangle indicates the proportion of trials answered correctly by a single speaker.

To test whether the rate of correct responses was higher in the prefix-controlled (PAC) group than for the other three groups and whether performance differed over time across groups, a binomial logistic regression model was fitted in R (R Core Team, 2022) using the *lme4* package (Bates et al., 2015). The model evaluated the log odds of subjects selecting the correct answer with regard to the factors Group, Trial, Number of Intervening Transparent Syllables, and Biphone Probability. Interactions of Group and Trial and of Number of Intervening Transparent Syllables and Trial were included in the model. Additionally, a random slope for Subject with regard to Trial and Number of Intervening Transparent Syllables was included. The variable Group had four levels – PAC, PSC, RAC, and RSC – with the reference level set to PAC, the group which bears the focus of the study. The variables Trial and Number of Intervening Transparent Syllables were centered on their mean values to

simplify interpretation. **Table 2** displays the log odds of selecting the right answer for each factor and the interaction of the fixed factors, and **Figure 2** displays the average proportion correct over time for each group.

In **Figure 2**, it appears that the performance of the two groups trained on a stem-controlled pattern remained fairly consistent over time, while the two affix-controlled groups showed notable improvement over the course of the testing task. Ultimate achievement was greatest for the Progressive Stem Control group, followed closely by Progressive Affix Control, then Regressive Affix Control, and finally, Regressive Stem Control. Thus, upon visual inspection of the results, progressive groups outperformed regressive ones, and affix-controlled patterns took longer to learn than stem-controlled ones. This is indicated by the continued improvement of affix-controlled groups in **Figure 2** compared to stem-controlled groups, whose performance underwent only slight improvements over the course of the 2AFC task with feedback. However, it is important to consider whether these apparent differences are supported by the regression model before interpreting the results.
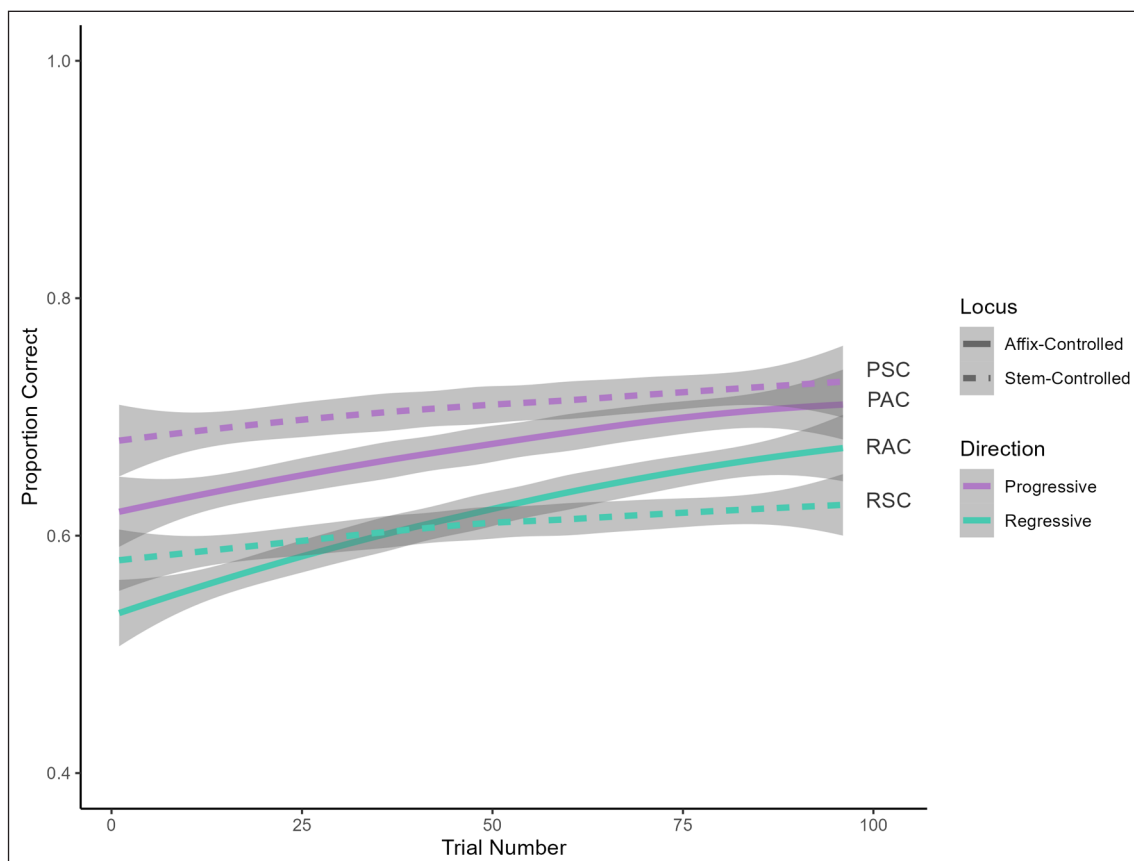

Figure 2: Proportion correct over time for English speakers (smoothed model-predicted values).

	Term	Estimate	Standard Error	z Statistic	p
$\beta 0$	(Intercept)	1.152127	0.284037	4.056	***
$\beta 1$	NITS	−0.140528	0.034264	−4.101	***
$\beta 2$	Trial	0.290761	0.08089	3.595	***
$\beta 3$	Group (PSC)	0.215575	0.397074	0.543	0.587192
$\beta 4$	Group (RAC)	−0.493848	0.389264	−1.269	0.204559
$\beta 5$	Group (RSC)	−0.360262	0.399694	−0.901	0.367405
$\beta 6$	Biphone Probability	0.00628	0.255787	0.025	0.980413
$\beta 7$	NITS: Trial	−0.04611	0.023761	−1.941	0.052316
$\beta 8$	Trial: Group (PSC)	−0.058339	0.111642	−0.523	0.601285
$\beta 9$	Trial: Group (RAC)	−0.006527	0.108189	−0.06	0.951894
$\beta 10$	Trial: Group (RSC)	−0.09441	0.108632	−0.869	0.384802

Table 2: Results of GLMM estimating the log odds of selecting the correctly-harmonized form, according to the training group (Group), Trial, where Trial indicates change over time, Number of Intervening Syllables (NITS), and Biphone Probability for English-speaking subjects. Trial and NITS are centered on their means; * indicates $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

In the initial inspection of **Figure 2**, it was evident that the hypothesis that the Progressive Affix Control group would perform much more poorly than the other groups was not supported by the data, and the model reported in **Table 2** confirms this. The model evaluates the simple effects of Group, Trial, and Number of Intervening Transparent Syllables. With regard to Group, we can observe that the log odds of selecting a correct answer at the midpoint of the study (i.e., on trial number 48, on which Trial was centered) for an (obviously theoretical) stimulus with 0.75 intervening transparent syllables and a biphone probability of 0.269 were 1.152 for those in the Progressive Affix Control group, corresponding to a 75.99% probability of a correct answer (β0, p < 0.001). Furthermore, at the midpoints of Trial, Number of Intervening Transparent Syllables, and Biphone Probability, there was no significant difference between the performance of the PAC group and that of the other three groups (β3, β4, β5). Thus, the Progressive Affix Control harmony process was far from unlearnable.

The model also provides information about how the four training groups differed in the way their performance evolved over time. The log odds of selecting a correct answer did improve over time for the PAC learners at the means of Number of Intervening Transparent Syllables

and Biphone Probability ($\beta 2$, $p < 0.001$). The results in $\beta 8 - \beta 10$ indicate that there was no significant difference in the slope of performance over time between the three other groups and the PAC group, indicating that while all four groups showed improvement over time, the differences in rate of improvement between groups (visualized in **Figure 2**) were not significant. To ensure that this result generalized over time, the best-fitting model was rerun with each of the remaining three groups (PSC, RSC, RAC) as the reference level. No notable changes to model results were obtained through this investigation; results from the rerun models are available in the Appendix.

2.2.1. Biphone probabilities

One possible source of variation in the data presented in this paper lies in the content of the nonword items themselves. As the stimuli are nonwords, calculating lexical frequency or neighborhood density was not possible, but a similar metric of phonotactic probability was calculated. Phonotactic probability refers to "the frequency with which legal phonological segments and sequences of segments occur in a given language" (Vitevitch & Luce, 2004, p. 481) and has been shown to affect speech segmentation (Gaygen, 1998; Pitt & McQueen, 1998), children's acquisition of novel words (e.g., Storkel, 2001, 2003), and recognition of spoken words (Vitevitch & Luce, 1999). To determine whether phonotactic probability influenced participants' responses, we used Vitevitch and Luce's (2004) phonotactic probability calculator to calculate both the phonemic and the biphone probabilities for each non-word in the task. Phonemic probability refers to the sum of the position-specific probability of each phoneme appearing where it does in the word, while biphone probability refers to the sum of the position-specific probabilities that each sequence of two phonemes appears in its given location. Both phoneme probability and biphone probability were included as possible predictors when fitting the model; however, for Study 1, the best-fitting model only included biphone probability, not phoneme probability. The effect of biphone probability was not significant ($\beta 0$, $p = 0.98$), indicating that there was no significant impact of biphone probability on the performance of the PAC group at the means of Trial and Number of Intervening Syllables.

2.2.2. Target-trigger distance

A secondary goal of this study was to determine whether the number of transparent syllables intervening between target and trigger impacted participants' performance on the AGL task, under the assumption that greater target-trigger distances may not only lead to a decrease in the incidence of harmony observed in spoken language, but may also decrease listeners' awareness of harmonicity. Model results reveal that as target-trigger distance decreased, performance improved for the PAC group at the means of Trial and Biphone Probability ($\beta 1$, $p < 0.001$). Furthermore, the interaction of Number of Intervening Transparent Syllables and Trial was not

significant (β7, $p = 0.052$), indicating that the improvement of the PAC group's performance over time was not mediated by target-trigger distance. The impact of target-trigger distance is visualized in **Figure 3**, which displays the proportion of correct responses given by each group by target-trigger distance (parametrized in the variable Number of Intervening Transparent Syllables [NITS]). It should be noted that target-trigger distance, of necessity, partially reflects word length: Words with a target-trigger distance of three must be at least five syllables long, as shown in (1a), while those with a target-trigger distance of 0 may be of any length, as shown in (1b). Bolding notes target and trigger segments.

(1)   a.   **s**ugatepa**s**e
        b.   **ʃ**iʃo; **ʃ**iʃibane; **s**u**s**api

As can be observed in **Figure 3**, the proportion of correct responses given declined gradually as target-trigger distance increased across all groups but the Regressive Stem Control group. The implications of this finding will be discussed further in Section 4.2.
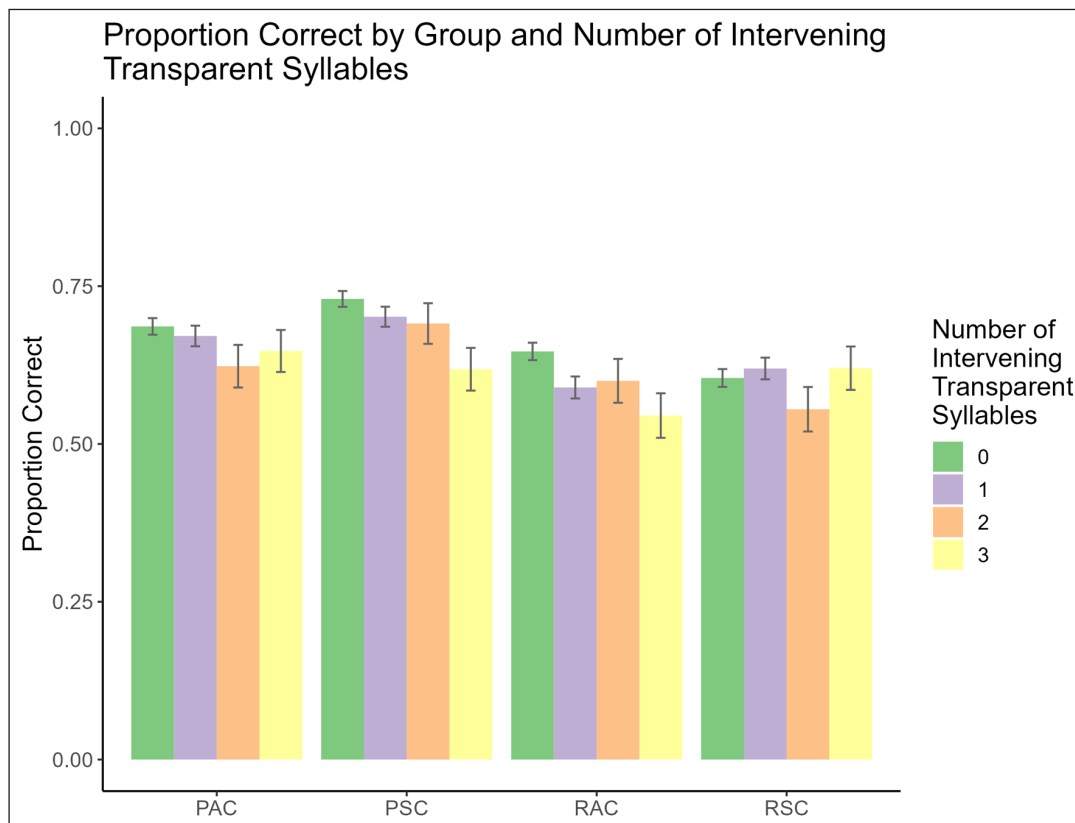

Figure 3: Proportion of correct responses by group and number of intervening transparent syllables for English-speaking subjects.

2.2.3. Summary of Study 1 results

While a cursory inspection of the data suggests that differences in performance existed based on the type of harmony pattern subjects were exposed to in training, these apparent differences were not substantiated by the regression analysis. No significant difference in performance was found between groups, and no significant differences emerged between groups with regard to the rate of learning (improvement over time). Thus, affix-controlled consonant harmony – including prefix-controlled harmony – was ultimately not less learnable than stem-controlled harmony. All groups did improve over time, suggesting that the corrective feedback provided throughout the 2AFC task continued to buttress subjects' understanding of the harmony pattern as the task proceeded. Performance was also impacted by target-trigger distance, with stimuli where the target was closer to the trigger generally eliciting a higher proportion of correct responses than those where the target and trigger were separated by a greater number of intervening transparent syllables.

2.3. Discussion of Study 1 results

The core question of this study is whether the typological lack of prefix-controlled sibilant harmony can be attributed to a relative lack of learnability when held against suffix-controlled and stem-controlled comparison groups in an artificial grammar learning study. In this section, several key findings in the relative learnability of these types of sibilant harmony emerged: No statistically significant differences in performance were found between learners exposed to progressive harmony and those exposed to regressive harmony, although the learners exposed to progressive harmony experienced slightly more success in acquiring the pattern than those exposed to regressive harmony. Similarly, no statistically significant differences in the speed of learning emerged between groups, although the rate of learning modeled in **Figure 2** depicts those learning a stem-controlled pattern generally reaching their ultimate level of achievement sooner than those in affix-controlled groups, whose performance continued to improve over a longer window of time. These minor differences were not sufficiently distinct to emerge as significant in the statistical analysis, suggesting that the trend was not robust or reliable. There was, however, a statistically significant effect of Trial (performance improved over time) and of target-trigger distance (performance worsened with greater target-trigger distances). Taken together, these results suggest that the typological lack of prefix-controlled consonant harmony is not driven by learnability, since the prefix-controlled pattern is indeed learnable.

The lack of a distinction between progressive and regressive patterns in the current results is particularly surprising in light of typological trends: As discussed in Section 1.4, past work on consonant harmony has found a cross-linguistic bias toward right-to-left directionality (Hansson, 2010). The learners in Study 1 were much more adept at acquiring progressive harmony patterns, including the prefix-controlled pattern, than expected – a notable departure from typological trends that demands some kind of explanation. The simplest explanation lies in the nature of

the learners recruited for the present study: They were all monolingual speakers of American English. English is notable for its strong commitment to left-edge prominence, a bias that could have influenced the results, leading the progressive groups to perform comparably to regressive ones. Past work that shows the influence of the left-edge bias in English includes Tyler and Cutler (2009), in which English speakers benefited from left-aligned but not right-aligned pitch movement cues in a speech segmentation task, and various studies have shown that English stress is strongly aligned with the left edge of the word and acts as a key speech segmentation tool (see, e.g., Cutler & Carter, 1987; Mattys, White, & Melhorn, 2005). This English left-edge bias may have skewed the present results in a way that is not in line with typological trends, which are not subject to influence by learners' previous linguistic background. To determine the extent to which the English language background of the subjects may have given rise to this unexpected progressive advantage, Study 2 repeats the experimental tasks with subjects from a distinct linguistic background – Spanish speakers.

3. Study 2 – Spanish

3.1. Background

Upon the conclusion of Study 1, a key question was whether the English preference for the left edge of the word could explain the unexpectedly successful performance of the progressive groups. To test this, a second study using identical stimuli and task design was conducted with speakers from a language background that did not favor the left edge. For this study, the ideal linguistic background would be one that elevates the right edge of the word across multiple levels of language. In English, for instance, certain syntactic processes, such as subject pronoun drop, are argued to occur in order to strengthen the left edge of prosodic phrases (Weir, 2012). Furthermore, English stress is trochaic – a property that may be construed as left-alignment within feet – and placement of primary stress and phrase-level prominence can fall on the initial (or non-final) foot in longer words and phrases, subject to a variety of phonologically- and morphologically-informed conditions (Hayes, 1995; Zsiga, 2013). Due to the complexity of the English stress assignment system, it would be an oversimplification to claim that word-level or phrase-level stress in English universally favors either edge of the word. Thus, the prominence of the left edge is reinforced by processes operating at various levels of language in English.

Spanish provides a striking contrast to the left-edge prominence found in English. Spanish stress, although also trochaic, is right-aligned at the word level and the phrase level (Prieto, 2006; Roca, 1991). Spanish prosodic phrases are determined in part by a prosodic constraint that is right-aligned (ALIGN-XP,R; see Prieto, 2006). Furthermore, noun + noun conjuncts append the modifying noun to the right of the head, unlike English, which attaches adjuncts to the left (compare *perro policía* to 'police dog' [Varela, 2012]). While these are only a subset of the processes found in each language, the prosodic differences are substantial. It is largely due to this difference that we selected Spanish as the language background for our second study.

3.2. Methods

The second study replicated the first, but with native speakers of Spanish recruited via Prolific (https://www.prolific.co). A total of 107 native Spanish speakers were recruited online to complete the series of tasks described in Section 2.1.3, administered with Spanish directions, and six were excluded due to failed attention checks, internet connectivity issues interrupting their experimental session, or selecting the first (or second) member of the two-alternative forced choice task for more than 70% of trials, suggesting they were not completing the task attentively. Thus, the total number of speakers whose data was included in the analysis was 101 (43M; 57F; 1 gender not reported; M = 29.9, SD = 9.06): 26 in the PSC group and 25 each in the PAC, RAC, and RSC groups. Fifty-six speakers reported residing in Mexico, 14 in Chile, 29 in Spain, and one each in the United States (born in Mexico) and United Kingdom (born in Venezuela). All participants reported that they had completed at least a high school diploma at the time of the study.

In addition to Spanish, 100 speakers reported some degree of multilingualism. Of those, 100 reported some knowledge of English, 18 French, six Italian, five German, four Portuguese, three Japanese, and one each Arabic, Basque, Catalan, Galician, Hindi, Polish, and Valencian. Sixty-one reported knowledge of two languages (Spanish and English), 35 had knowledge of three languages, and four reported knowledge of four languages.

The stimuli,[1] task design, and all other elements of the study were identical to those described in Study 1. All study procedures were reviewed by the Institutional Review Board at Carleton College (Protocol 2022-23 1562).

3.3. Results of Study 2

The goal of Study 2 was to discover whether the pattern of performance found in Study 1, in which no significant difference was present between the Progressive Affix Control group and the other three groups (although in absolute terms the performance of the two progressive groups exceeded that of the two regressive groups) would hold true for subjects with a different linguistic background. An initial glance at the data, as shown in **Figure 4** and **Figure 5**, demonstrates that the results from the Spanish speakers who completed Study 2 did not imitate the performance of the English speakers in Study 1. **Figure 4** shows the mean percent correct for all learners and for those who successfully learned the pattern, and **Figure 5** shows each group's

---

[1] A minority of forms appearing in the nonword dataset may have offered Spanish learners some advantage due to their similarity to existing Spanish words. A sensitivity review of the nonword set conducted by a native Spanish speaker flagged 11 of the 336 nonwords appearing in the study as possible sources of lexical interference. Out of these 11 forms, only four were real words (*Xuxa,* a name; *celu* 'cell phone'*; tuso* 'bald' or 'heartbroken'; and *siza* 'sleeveless'), while the remaining seven were regionally possible variations of real words pronouncing as /ʃ/ phonemes more typically produced as /j/ or /t͡ʃ/. Given that these items represented only 3.2% of the set of nonwords, any impact of these items on the overall results would be minimal and it was not found necessary to replace the items with true nonwords. Instead, the nonword stimulus set was kept in its original form in order to maintain consistency with Study 1 in study design as fully as possible.

performance modeled over time. (See Section 2.2 for an explanation of how successful learners were identified.) In Study 2, the Progressive Stem Control (PSC) group performed most poorly, with Regressive Stem Control (RSC) performing next best and Progressive and Regressive Affix Control (PAC/RAC) closely aligned for the best performance. Unlike in Study 1, the number of successful learners per group cannot cleanly explain overall group performance, as can be observed by comparing **Figure 4** and **Figure 5**. Rather, both the number of successful learners and the relative success levels of successful and unsuccessful learners visibly influenced the final group performance to a greater degree than in Study 1. Furthermore, the two progressive groups of Spanish-speaking subjects exhibited a clear division into highly successful learners and unsuccessful or only marginally successful learners, clustering below 68% and above 86% correct on the 2AFC task. Similar clustering appears in the RSC group, with a gap between 79% and 94%. This division is visible in **Figure 4** when examining the triangles indicating percent correct by speaker, with two speakers in PAC and two in PSC falling in the lower cluster of subjects, despite their inclusion in the successful learner category. Furthermore, it is worth noting the difference in the distribution of by-speaker means between the PAC and RAC groups in **Figure 4**. Despite having very similar group means (71.5% and 71.3%, respectively), PAC successful learners were
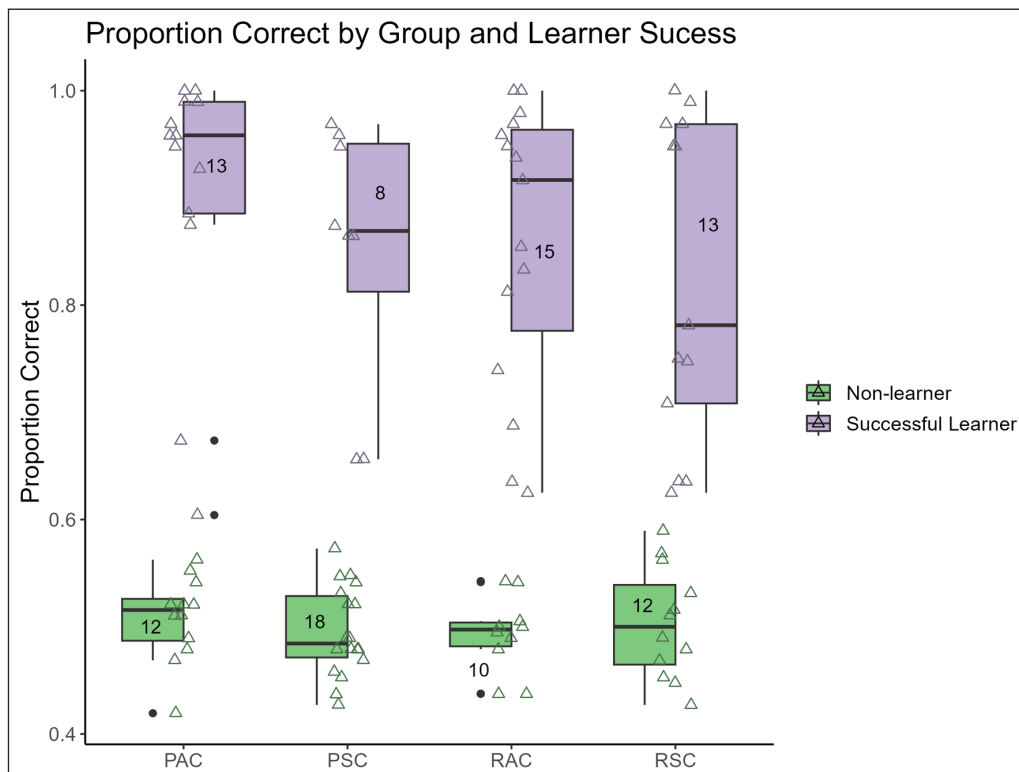

Figure 4: Proportion correct by group, with learners divided by ultimate success (greater or less than 60%) for Spanish speakers. Annotations display the number of learners in each box. Each triangle indicates the proportion of trials answered correctly by a single speaker.

generally much more successful than PAC unsuccessful learners, while RAC learners exhibited a more even distribution of performance across speakers.

A binomial logistic regression model was fitted in R (R Core Team, 2022) using the *lme4* package (Bates et al., 2015) to ascertain which group displayed the highest rate of correct responses and how the performance of each group changed over the course of the task. The model included fixed effects of Group, Trial, Number of Intervening Transparent Syllables, Biphone Probability, and Phoneme Probability, with Trial and Number of Intervening Transparent Syllables centered on their means for ease of interpretation. Interactions were included for Trial and Group and for Number of Intervening Transparent Syllables and Trial, as well as a random slope for Subject in relation to Trial. Results of this model are summarized in **Table 3**.

As was the case in Study 1, it is evident from **Figure 5** that the Progressive Affix Control group did not suffer a systematic disadvantage in learning for the Spanish speakers any more than it did for the English speakers, and the model summarized in **Table 3** confirms this. The first line of **Table 3** delivers the log odds of a PAC learner selecting the correct answer on the

	Term	Estimate	Standard Error	z Statistic	p
$\beta 0$	(Intercept)	1.753151	0.343617	5.102	***
$\beta 1$	NITS	0.004022	0.039851	0.101	0.9196
$\beta 2$	Trial	0.175155	0.10806	1.621	0.105
$\beta 3$	Group (PSC)	−0.923724	0.443704	−2.082	*
$\beta 4$	Group (RAC)	−0.064027	0.45125	−0.142	0.8872
$\beta 5$	Group (RSC)	−0.445119	0.449611	−0.99	0.3222
$\beta 6$	Biphone Probability	−0.725059	0.412679	−1.757	0.0789
$\beta 7$	Phoneme Probability	4.647844	2.600521	1.787	0.0739
$\beta 8$	NITS: Trial	−0.03219	0.024697	−1.303	0.1924
$\beta 9$	Trial: Group (PSC)	−0.020303	0.141357	−0.144	0.8858
$\beta 10$	Trial: Group (RAC)	0.209226	0.145512	1.438	0.1505
$\beta 11$	Trial: Group (RSC)	0.014375	0.143405	0.1	0.9202

Table 3: Results of GLM (for Study 2) estimating the log odds of selecting the correctly-harmonized form, according to the training group (Group), Trial (centered), Number of Intervening Syllables (NITS – centered), Biphone Probability, and Phoneme Probability; * indicates $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

48th trial (the midpoint on which the variable Trial was centered) at the reference level (means) of the remaining predictors: 1.753, or an 85.2% probability (β0, p < 0.001). In fact, at this midpoint, PAC significantly outperforms the PSC group (β3, p < 0.05) and has performance indistinguishable from the two regressive groups (RAC: β4, p = 0.887; RSC: β5, p = 0.322). Thus, it is evident that prefix-controlled harmony suffered no deficit of learnability in Study 2 relative to the other training groups.

The simple effect of Trial and the interaction between Trial and Group provide information about how the performance of the four groups changed over time. The estimate for Trial gives the slope of modeled improvement in performance for the PAC group at the means of Number of Intervening Transparent Syllables (0.75), Biphone Probability, and Phoneme Probability, which is not significant (β5, p = 0.321417), indicating that no significant improvement over time occurred under the reference conditions. Furthermore, the results given in β9 – β11 of **Table 3** model the difference in the slope of performance over time between the three other groups
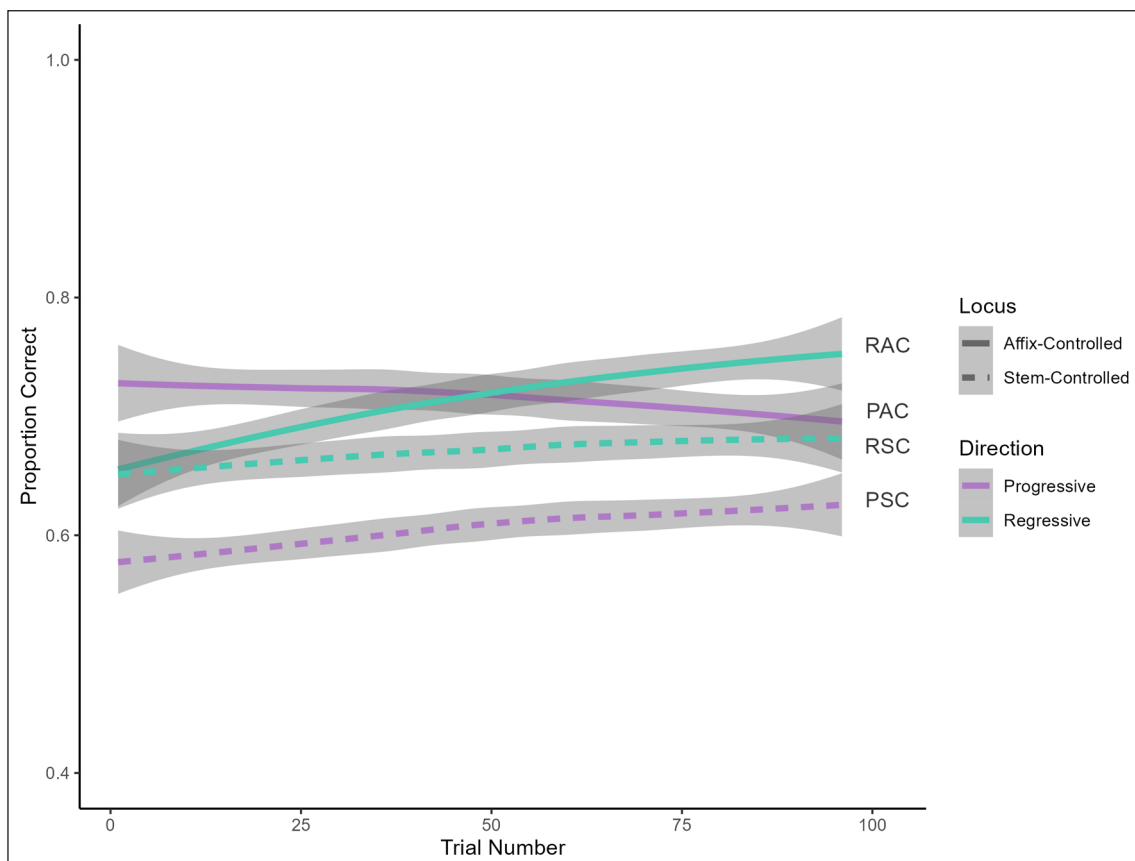

Figure 5: Proportion correct by group over time for Spanish speakers (smoothed model-predicted values).

and the PAC group. They indicate that there was no significant difference in improvement over time between the PAC group and the other three groups. To ensure no important simple effects were overlooked, particularly with regard to Trial and its interactions, the best-fitting model was rerun with each of the remaining three groups (PSC, RSC, RAC) as the reference level. Full results from these models are available in the Appendix. The only notable effect that emerged from this analysis appeared when the reference level was RAC: In this model, the effect of Trial was significant ($\beta = 0.384$, SE $= 0.106$, $z = 3.616$, $p < 0.001$). This simple effect indicates that for the RAC group at the mean of the remaining predictors, the log odds of a subject selecting the correct answer improved by 0.384 for each additional trial. In short, the RAC group continued to improve over the course of the experiment, while the other three groups showed no significant change in performance due to Trial at reference level.

3.3.1. Biphone probabilities

As was done for Study 1 (see Section 2.2.1), the phoneme and biphone probabilities for each of the nonwords was calculated using Vitevitch and Luce's (2004) Spanish-language phonemic probability calculator (available at https://calculator.ku.edu/phonotactic/Spanish/words as of this writing). Unlike in Study 1, the best-fitting model contained terms for both Biphone Probability and Phoneme Probability; however, neither term was significant, indicating that these properties of individual items did not serve to predict subjects' performance on those items ($\beta6$, $p = 0.0789$; $\beta7$, $p = 0.0739$).

3.3.2. Target-trigger distance

As discussed for Study 1 in Section 2.2.2, a secondary point of interest in this study was to investigate the impact of target-trigger distance on performance in the AGL task. **Figure 6** visualizes the proportion of correct responses delivered by each of the four groups at each target-trigger distance tested. Target-trigger distance was parametrized as Number of Intervening Transparent Syllables (NITS). In Study 1, performance decreased for each additional intervening transparent syllable for all groups except the RSC group, whose performance across target-trigger distances was relatively consistent. In Study 2, however, increased target-trigger distance did not correspond to a general worsening in performance. This is evident in the model, in which the effect of NITS was not significant ($\beta1$, $p = 0.9196$) and the interaction of NITS and Trial was not significant, indicating no change in the effect of NITS over time ($\beta8$, $p = 0.1924$). Instead, target-trigger distance was relatively steady across all four levels of NITS for all four groups, as can be observed in **Figure 6**. This result will be discussed further in Section 4.2.
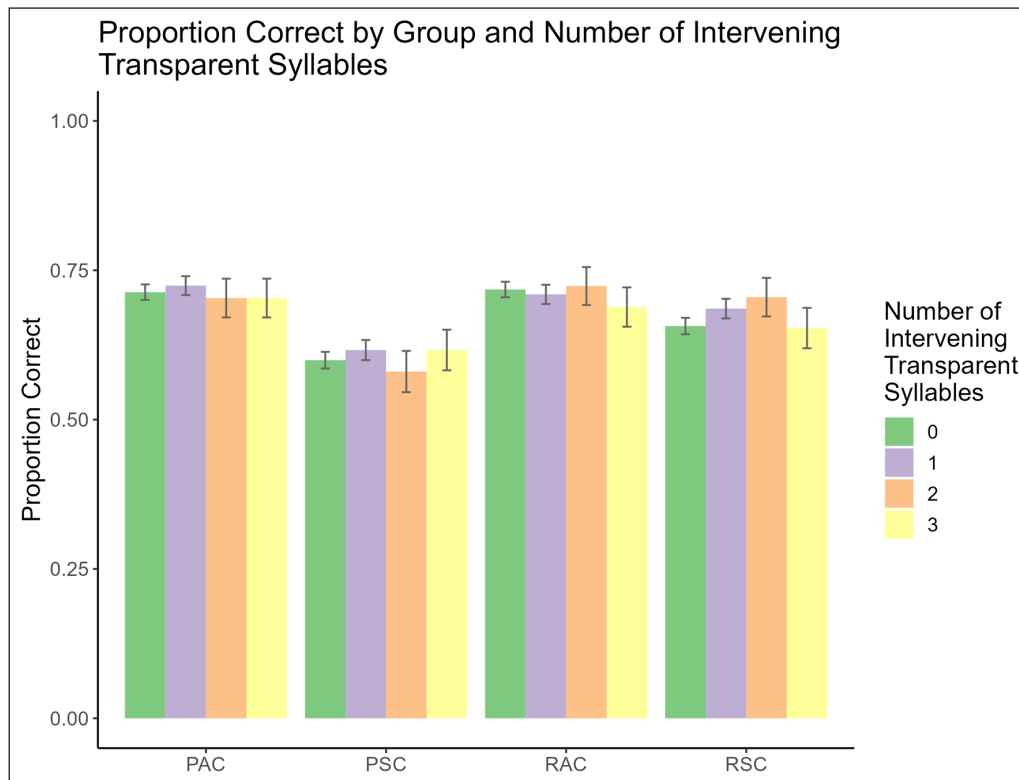
Figure 6: Proportion correct responses given by Spanish speakers by Group and Number of Intervening Syllables.

3.3.3. Summary of Study 2 results

Like in Study 1, the Progressive Affix Control group suffered no disadvantage in learnability compared to the other three groups. Unlike in Study 1, the PAC group actually outperformed the PSC group in a statistically significant manner. Progressive groups differed from regressive groups in the clustering of individual learners' performance: The distribution of progressive subjects' overall performance was bimodal, while regressive learners were more evenly distributed over the entire range. Also unlike in Study 1, there was no significant effect of Trial or NITS, indicating that subjects' performance did not improve over time and a larger target-trigger distance did not negatively impact performance. These findings are explored further in Section 4.

3.4. Discussion of Study 2 results

The results of Study 2 demonstrated that Spanish speakers, like English speakers, do not have a systematic learning bias against prefix-controlled sibilant harmony. In absolute terms, the PAC group, together with the RAC group, exhibited the strongest performance of the four harmony types for Spanish speakers, and statistical modeling confirmed that PAC performance equaled

or exceeded that of every other group. This finding makes clear that the surprising absence of progressive bias found in Study 1 was not limited only to English speakers, although absolute differences in progressive and regressive performance were clearer and more pronounced in the English data than the Spanish data, judging by the fact that the Progressive Stem Control group exhibited the worst performance of the four groups in Study 2. Additionally, progressive groups in Study 2 (as well as, to a lesser degree, the RSC group) were marked by a bimodal distribution of subjects into highly successful learners and unsuccessful or only moderately successful learners. We find it notable that the only group not to exhibit this bimodal distribution of mean performance by subject was the Regressive Affix Control group – the group with the steepest apparent slope in **Figure 5**. Taken together, these two aspects of the data suggest that the RAC group may still have been learning the target pattern throughout the 2AFC task, while subjects from the other three groups either acquired or failed to acquire the target pattern earlier (either during the training task or early in the 2AFC task). If this is the case, it may be a reflection of a higher level of complexity or less intuitively apparent pattern in the suffix-controlled harmony; however, this is mere speculation as the analysis design of the current study is not equipped to offer any kind of quantitative commentary on this point. What is clear is that the prefix-controlled harmony did not suffer a deficit of learnability relative to the other types of harmony tested.

3.4.1. The role of /s/ and /ʃ/ in Spanish

Because /ʃ/ is not canonically phonemic in Spanish, it may not be an obvious choice for a study of sibilant harmony, as sibilant harmony hinges on the ability to reliably distinguish between /s/ and /ʃ/. However, although /ʃ/ is not a phoneme in most Spanish dialects, it is reasonable to expect that Spanish speakers will easily distinguish it from /s/, for several reasons. First, it is commonly used paralinguistically to request quiet,[2] much in the same manner as in English. Additionally, it is a phonological component of /t͡ʃ/ (<ch>), which is phonemic in most Spanish dialects (Varela, 2012). Phonologically, the articulation and thereby also recognition of the components of a contour sound are expected as prerequisites of using that contour sound. Furthermore, some regional dialects in northern Mexico, Chile, and southern Spain, among other places, reduce the /t͡ʃ/ affricate to /ʃ/ as a matter of course (Varela, 2012). Given that 100 out of 101 participants recruited for this study spoke Mexican, Chilean, or Peninsular Spanish, it is not unreasonable to expect them to have some familiarity with these dialects. Finally, the overall success learners had in the 2AFC task indicates that the /ʃ/ itself was likely not an obstacle – indeed, there were more successful learners of the sibilant harmony pattern in the Spanish study (49) than the English study (45). For all these reasons, it is unlikely that the lack of phonemic /ʃ/ in Spanish had a substantial impact on the generalizability of the results.

---

[2] Personal communication, AnaMaria Molina Castrillón, June 7, 2023.

4. General discussion

This AGL-based exploration of the learnability of prefix-controlled sibilant harmony demonstrated that native speakers of both English and Spanish are capable of learning a prefix-controlled pattern after only a short training session. This striking result makes clear that traditional assumptions about the impossibility of prefix-controlled consonant harmony must be reevaluated. Furthermore, not only was the prefix-controlled pattern as learnable as widely-attested patterns such as regressive affix control and progressive stem control, but PAC learners outperformed subjects exposed to a progressive stem control harmony in the Spanish study. In this discussion, we consider the implications of these results and reflect on how language background may have led to an unexpected lack of progressive bias, whether the current results should prompt a reexamination of the assumption that prefix-controlled harmony is dispreferred, and how aspects of the study design may have shaped the current results.

4.1. Unexpected lack of regressive advantage

In the world outside the lab, regressive harmony offers languages a distinct communicative advantage that progressive harmony lacks: It provides a preview of upcoming phonological or even lexical content, potentially making the speech perception process more streamlined and efficient. By contrast, the primary boons progressive harmony can lend to speech perception efficiency are to offer a secondary cue to the location of a word boundary, or, especially for acoustically difficult-to-distinguish features, to provide additional opportunities for listeners to recognize the feature value present over a harmonic span (Kaun, 2004).[3] However, as this second advantage is primarily tied to harmony processes involving perceptually difficult features (such as rounding) and is strengthened by multiple repetitions of the harmonizing feature, it likely does not play a strong role in the emergence or maintenance of sibilant harmony.

If the communicative advantage enjoyed by regressive harmony is to blame for its typological over-representation, it stands to reason it would exert no influence in the AGL context, which by its very nature tests only the learnability of the competing processes. By this logic, it is reasonable to propose that progressive sibilant harmony may be equally if not more learnable than similar regressive processes. The results of Study 1 indicate that, for English speakers, progressive harmonies were indeed as learnable as regressive harmonies, while Study 2 found that progressive affix-controlled harmony and regressive harmonies enjoyed similar learnability greater than that present for progressive stem-controlled sibilant harmony. This mismatch between progressive learnability and regressive typological dominance serves as an excellent reminder that AGL results only reflect a limited subset of the pressures that shape typological success. In Section 4.3, we consider the role of factors beyond learnability in accounting for typological gaps in harmony systems.

---

[3] Thanks are due to an anonymous reviewer for reminding us of the relevance of Kaun's work.

4.2. Differences in performance between languages

One of the most notable differences between the performance of the English monolinguals in Study 1 and the Spanish-speaking subjects in Study 2 lay in the differing impacts of target-trigger distance and change over time on correct identification of harmonic forms. English-speaking subjects' performance worsened as target-trigger distance increased, while the performance of the Spanish-speaking subjects was consistent across all four target-trigger distances tested. Similarly, English-speaking subjects' performance improved over time across all four groups, while Spanish speakers only showed a significant effect of Trial for the RAC group, which may suggest that the Spanish PAC, PSC, and RSC groups achieved their ultimate proficiency earlier than the RAC group and that the English-speaking subjects did so during the training task. These discrepancies may be linked to the multilingualism of the Spanish-speaking subject pool. Unlike the English-speaking participants, the subjects in Study 2 were not monolingual: All but one reported at least some knowledge of English, and 39 of the 101 participants reported knowledge of three or more languages. Past work has found evidence of a bilingual advantage in L3 learning, particularly in the area of vocabulary (see, e.g., Keshavarz & Astaneh, 2004; Salomé, Casalis, & Commissaire, 2022), although this advantage can also be limited to changes in processing but not performance in the L3 (Rutgers & Evans, 2017) and has also been documented for phonetic learning (Antoniou, Liang, Ettlinger, & Wong, 2015). Given the similarity between learning an artificial grammar and learning an L2, it seems reasonable to assume that this bilingual advantage could extend to multilingual AGL participants such as those in Study 2. Added facility with the learning and processing of novel language material could account for the disparity in the effect of target-trigger distance observed between Study 1 and Study 2, as well as the apparently earlier plateaus in learning observed in the Spanish-speaking group.

4.3. Accidental typological gaps

The results of this study suggest that the typological bias against prefix-controlled sibilant harmony cannot be accounted for through differences in learnability. This leaves two possible explanations: The apparent cross-linguistic lack of prefix-controlled consonant harmonies may represent an accidental gap, or it may suffer a practical or theoretical disadvantage not related to learnability. Accidental gaps occur when theoretically possible phenomena remain unattested without any linguistic or practical explanation beyond the fact that no languages that happen to have been documented in the academic literature contain the feature or process in question.[4] Is there any evidence that the lack of documented, fully-developed prefix-controlled consonant harmony systems represents an accidental typological gap rather than a true cross-linguistic generalization? A thoroughgoing analysis is well beyond the scope of this paper, but a brief

---

[4] See Gordon (2002) for a discussion of accidental gaps in the typology of stress systems.

sketch is feasible. The most complete list of consonant harmony systems appears in Hansson (2010) in the form of a database of 178 consonant harmony systems across 134 languages, of which 70 systems targeted sibilants in some way. Of these 178 consonant harmony systems, only one (Navajo) exhibited any evidence of prefix control, and in this case, only a single trigger prefix existed, which exclusively targeted other prefixes, meaning that the system did not constitute evidence of a prefix triggering harmony in a stem. While it would be hasty to assume that this database necessarily constitutes a representative sample of all possible human languages, it at the very least suggests that the incidence of prefix-controlled harmony is extraordinarily low, and correspondingly that the likelihood that the lack of prefix-controlled consonant harmony represents an accidental gap is also quite low.

Given the lack of direct evidence, it is worth considering what cognitive and theoretical restrictions support the notion that the lack of prefix-controlled consonant harmony represents a true typological gap rather than an accidental one – and, contrarily, what arguments buttress the idea of an accidental gap. Both theoretical and experimental reasons exist that argue in favor of the gap being accidental. The data found in the current work demonstrates experimentally that prefix-controlled consonant harmony is, at the very least, learnable. Learnability is the first criterion: All possible linguistic patterns must, at the very least, be learnable, although some learnable patterns may still not be theoretically predicted (Alderete, 2008). The question then becomes whether theoretical accounts of consonant harmony predict the existence of prefix-controlled processes.

Recent theoretical work on consonant harmony relies on the theory of Agreement by Correspondence ([ABC]; Rose & Walker, 2004) to account for patterns of long-distance agreement such as sibilant harmony (Bennett, DelBusso, & Iacoponi, 2016; Hansson, 2010). ABC is an optimality-theoretic approach to long-distance agreement that posits a family of Correspondence constraints (CORR) that enforce agreement between segments in a domain on the basis of phonological similarity (Rose & Walker, 2004). The CORR constraint interacts with IDENT constraints and surface identity constraints to force an alternation in corresponding segments resulting in agreement. However, ABC in and of itself does not make any specific predictions about directionality or prefix control: Past work has used ABC to generate contradictory predictions about the existence of prefix-controlled harmony. For instance, Iacoponi's (2015) factorial typology predicts only regressive and dominant harmonies will exist, but it does so as a consequence of intentionally modeling Baković's set of possible harmony types and not as an inherent property of ABC. By contrast, McCollum and Essegbey's (2020) thorough exploration of prefix-controlled vowel harmony in Tutrugbu demonstrates a viable ABC account of progressive affix-controlled harmony, equally applicable to [sibilant] as to [round]. While there are solid arguments to be made for the need for separate theoretical accounts for spreading-based harmony such as rounding harmony and agreement-based processes like sibilant harmony, McCollum and

Essegbey's work at a minimum demonstrates that ABC can be used to model prefix-controlled harmony. This comparison, as well as common sense, should make clear that any answers to the question of whether prefix-controlled harmony is possible will not fall out from the theory itself, but the assumptions underlying the theory. A thoroughgoing reevaluation of these assumptions is a matter for future research, but we will briefly consider what other evidence might support the notion that prefix-controlled harmony is or is not possible.

In addition to a lack of evidence of prefix-controlled consonant harmony arising naturally and a failure to motivate it as an innate consequence of the theoretical representation of long-distance agreement, there are compelling reasons to believe that prefix-controlled harmony is dispreferred. It has been argued that the relationship between root and suffix is more closely prosodically bound than the root and prefix (Nespor & Vogel, 2007), and that the root + suffix conjunct is the preferred domain for vowel harmony compared to the root + prefix domain (Hyman, 2002; White, Kager, Linzen, Markopoulos, Martin, Nevins, Peperkamp, Polgárdi, Topintzi, & van de Vijver, 2018). Given this close relationship, it is easy to imagine that the root + suffix conjunct may also be a preferred domain for consonant harmony.

Alternatively, prefix-controlled harmony may be dispreferred because it fails to offer a competitive advantage in communication: As discussed in Section 4.1, regressive harmony aids in speech perception by foreshadowing the properties of upcoming segments and thus narrowing the number of possible morphemes competing for precedence through top-down processing. However, progressive harmony primarily aids in speech perception by providing an additional cue to the placement of word boundaries and redundant cues to the harmonizing feature value, neither of which offers much advantage in the case of sibilant harmony, which does not affect enough segments to provide either effective word boundary cues or helpful redundancy of an acoustically difficult feature. It is also possible that the preference for regressive consonant harmony may be due to consonant harmony arising as an aspect of speech planning, which is typically more prone to anticipatory effects (Hansson, 2010). Any or all of these reasons may account for progressive affix-controlled consonant harmony remaining unattested despite its apparent learnability. Given the existence of prefix-controlled vowel harmony, however rare, we believe the question of whether prefix-controlled consonant harmony can emerge in natural language hinges on the distinction between vowel harmony and consonant harmony. Does the root-prefix boundary pose a greater obstacle to processes of spreading (i.e., vowel harmony) than it does to agreement (most consonant harmonies)? Do the preference for anticipation in speech planning and the rooting of consonant harmony in speech planning conspire to prevent the emergence of prefix-controlled consonant harmony in a manner not applicable to its vocalic counterpart? Future consideration of the status of prefix-controlled consonant harmony must begin with these questions and continue with a thorough consideration of the assumptions underlying our theoretical accounts of harmony.

4.4. Differences from vowel harmony

It is worth noting that the central finding of this study – that learners are capable of acquiring a prefix-controlled consonant harmony pattern despite its unattested nature – is fundamentally at odds with past AGL work on vowel harmony, which has shown that learners have success in acquiring suffix-controlled patterns, but not prefix-controlled ones (Finley & Badecker, 2009). While it is possible that differences of study design led to these divergent findings across vowel and consonant harmony, it is not implausible to suppose that one of the several true differences between vowel and consonant harmony is responsible for the divergent result. Prefix-controlled harmony is unattested for consonants and extremely rare for vowels, but it does not necessarily follow that these trends can be attributed to the same underlying cause: Vowel harmony may suffer from a lack of learnability in the prefix-controlled condition, while consonant harmony fails to occur in a prefix-controlled form not because it is unlearnable, but because it offers no advantage or has no plausible historical basis from which to emerge. While it would be premature to consider these suppositions reliable based only on the contrast between two studies that have notable differences in design (it is possible that the different amounts of training in the two studies led to the differing results), our initial findings, combined with those of Finley and Badecker (2009), certainly suggest that there may be an underlying difference in the reasons for the rarity of prefix-controlled harmony across the vowel and consonant domains. This is not surprising, given that these two types of harmony, despite sharing a name, are attributed to fundamentally different mechanisms. Vowel harmony is generally recognized as an example of feature spreading, with diachronic roots in coarticulation, while certain consonant harmonies, including sibilant harmony, are better characterized as a form of long-distance agreement, with roots in the speech planning mechanism (Gafos, 2021; Hansson, 2010). Considering the nature of the connection between the underlying mechanism and the learnability of prefix control, it is not unreasonable to posit that the two mechanisms may interact with the prefix-stem boundary in fundamentally different manners, leading to these divergent findings.

4.5. Shortcomings of the current paradigm and directions for future research

One factor that may have influenced the results of this study is differences in stimuli design across the progressive and regressive groups. In addition to differing by direction and locus of control, the four learner groups in this study necessarily differed in a second meaningful way: Changes to the locus of control introduced changes to the stimuli design. As illustrated in the Appendix, the affixes shown to stem-controlled learner groups alternated between two allomorphs, leaving four allomorphs [su, ʃu, si, ʃi] for two morphemes /Su, Si/, while the affix-controlled groups saw only one allomorph per morpheme (/su/ and /ʃi/). This difference in design may also be construed as a difference in complexity; however, the design of the study attempted to compensate for this

difference by presenting triplets for each trial, as discussed in Section 2.1.2. Thus, we believe that this difference has been minimized as far as the innate differences underlying affix control and stem control will allow.

In an open-ended question at the end of the study, we asked participants to characterize the pattern they recognized in the data, if they had been conscious of any: Responses suggest that a small minority of participants in the affix-control groups may have framed the pattern as matching the correct vowel to the correct sibilant, while some framed the pattern as a question of sibilant matching, and others failed to identify any pattern at all. However, respondents in the stem-controlled groups gave similar responses, with at least one mistakenly believing that the key to answering correctly was learning which vowel belonged to which "tense." Given that cross-featural implications (such as [α sibilant] → [α back]) are unattested in natural language for any feature, we are skeptical that these responses reflect the subjects' unconscious learning, particularly given that learners in the affix-control groups were more likely to learn the pattern implicitly (i.e., without recognizing the nature of the rule) than those in the stem control groups.

Another aspect of the study design that could have influenced the results relates to the placement of stress in the stimuli. In addition to the dissimilarities in stimulus construction observed across the groups, it is worth noting that stress placement interacted with stems differently in the suffixing groups than the prefixing groups. Since stress was always assigned to the penultimate syllable, prefixing groups (i.e., RSC/PAC) consistently heard the same CV syllable stressed each time a given nonword appeared, while suffixing groups (RAC/PSC) encountered stress on different syllables. To demonstrate, consider the stem [gateˈpase]: A valid affixed token for the prefixing groups might be [sugateˈpase], whereas for the suffixing groups, the unaffixed form undergoes a stress-placement alternation when the suffix is added, from [gateˈpase] to [gatepaˈsesu]. The stress shifts by a single syllable, potentially accentuating the presence of the suffix. If this were an advantage for the suffixing groups, one would anticipate both the regressive affix control and progressive stem control groups to exhibit superior performance to the two remaining groups. However, given that the PSC and RAC groups were not uniformly disadvantaged across the two studies, despite the relatively poor PSC performance in Study 2, it seems likely that the shift in stress placement did not meaningfully influence the participants' performance.

Finally, the simplistic division of sibilant harmony patterns into four clear-cut categories based on direction and locus of control used in this study fails to fully capture the complexity of harmony systems present in natural language. Some harmonies – such as Chumash sibilant harmony – are purely regressive, lacking any morphological locus of control (Hansson, 2010, p. 45). Others are bidirectional, while still others follow a dominant-recessive pattern in which one feature value overrides the other (Baković, 2000; Hansson, 2010). The four-way distinction laid out here necessarily elides these complexities in an attempt to make generalizations about the

relative learnability of harmony based upon direction and locus of control, but in so doing fails to fully capture the range of natural harmony systems. Thus, our finding that prefix-controlled harmony is not less learnable than attested consonant harmonies should not be taken to imply that all consonant harmony systems fall neatly into the four-way division employed here.

Future research building on the current findings should consider whether other consonant harmonies share similar learnability profiles to sibilant harmony, what role stress plays in shaping the learning of consonant harmony, and whether the results presented here generalize across a wider range of participant language backgrounds. If this future work reinforces the current finding that prefix-controlled consonant harmony is learnable, phonologists must seek explanations for the typological gap in other domains, such as phonological theory, the diachronic emergence of harmony, or cognitive restrictions unrelated to learnability. Most importantly, future work should consider the implications of the learnability of prefix-controlled consonant harmony, weighed against its typological absence, when developing theoretical accounts of consonant harmony.

Additional files

Supplementary File 1: Appendix. Stimuli set & secondary model results from the experiments. DOI: https://doi.org/10.16995/labphon.9727.s1

Supplementary File 2: Data. Deidentified data from the experiments. DOI: https://doi.org/10.16995/labphon.9727.s2

Supplementary File 3: Code. Statistical code from the experiments. DOI: https://doi.org/10.16995/labphon.9727.s3

Competing interests

The authors have no competing interests to declare.

References

Alderete, J. (2008). Using learnability as a filter on factorial typology: A new approach to Anderson and Browne's generalization. *Lingua, 118*(8), 1177–1220. DOI: https://doi.org/10.1016/j.lingua.2008.03.002

Antoniou, M., Liang, E., Ettlinger, M., & Wong, P. C. M. (2015). The bilingual advantage in phonetic learning. *Bilingualism: Language and Cognition, 18*(4), 683–695. DOI: https://doi.org/10.1017/S1366728914000777

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388–407. DOI: https://doi.org/10.3758/s13428-019-01237-x

Baković, E. (2000). *Harmony, dominance and control* (Doctoral dissertation, Rutgers University, Newark, NJ). DOI: https://doi.org/10.7282/T3TQ60BJ

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Bennett, W. G., DelBusso, N., & Iacoponi, L. (2016). Formally mapping the typologies of interacting ABCD systems. *Proceedings of the Annual Meetings on Phonology, 3*. DOI: https://doi.org/10.3765/amp.v3i0.3697

Berkson, K. (2013). Optionality and locality: Evidence from Navajo sibilant harmony. *Laboratory Phonology*, *4*(2), 287–337. DOI: https://doi.org/10.1515/lp-2013-0011

Bermúdez-Otero, R. (2003). The acquisition of phonological opacity. *Variation within Optimality Theory: Proceedings of the Stockholm Workshop onVariation within Optimality Theory*, 25–36. http://roa.rutgers.edu/files/593-0403/593-BERMUDEZ-OTERO-0-0.PDF

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

Boersma, P., & Pater, J. (2008). Convergence properties of a gradual learner in Harmonic Grammar. In J. McCarthy & J. Pater (Eds.), *Harmonic Grammar and Harmonic Serialism* (pp. 389–434). Sheffield: Equinox.

Conklin, J. (2015). *The interaction of gradient and categorical processes of long-distance vowel-to-vowel assimilation in Kazan Tatar* (Unpublished MA thesis, Purdue University, West Lafayette, IN). Retrieved from https://docs.lib.purdue.edu/open_access_theses/565/

Crystal, D. (2003). *A dictionary of linguistics & phonetics* (5th ed). Malden, MA: Blackwell.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, *2*(3), 133–142. DOI: https://doi.org/10.1016/0885-2308(87)90004-0

Finley, S. (2009). *Formal and cognitive restrictions on vowel harmony*. (Doctoral dissertation: Johns Hopkins University, Baltimore, MD) Retrieved from https://doi.org/doi:10.7282/T3K64H05

Finley, S. (2011). The privileged status of locality in consonant harmony. *Journal of Memory and Language*, *65*(1), 74–83. DOI: https://doi.org/10.1016/j.jml.2011.02.006

Finley, S. (2012). Testing the limits of long-distance learning: Learning beyond a three-segment window. *Cognitive Science*, *36*(4), 740–756. DOI: https://doi.org/10.1111/j.1551-6709.2011.01227.x

Finley, S. (2015). Learning nonadjacent dependencies in phonology: Transparent vowels in vowel harmony. *Language*, *91*(1), 48–72. DOI: https://doi.org/10.1353/lan.2015.0010

Finley, S. (2017). Locality and harmony: Perspectives from artificial grammar learning. *Language and Linguistics Compass*, *11*(1), e12233. DOI: https://doi.org/10.1111/lnc3.12233

Finley, S., & Badecker, W. (2008). Analytic biases for vowel harmony languages. *Proceedings of the 27th West Coast Conference on Formal Linguistics, 27*, 168–176.

Finley, S., & Badecker, W. (2009). Right-to-left biases for vowel harmony: Evidence from artificial grammar. *Proceedings of the 38th North East Linguistic Society Annual Meeting, 1*, 269–282.

Friederici, A., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences*, *99*(1), 529–534. DOI: https://doi.org/10.1073/pnas.012611199

Gafos, A. (2021). Consonant harmony, disharmony, memory and time scales. *Proceedings of the Society for Computation in Linguistics*, *4*(1), 188–205.

Gaygen, D. (1998). *Effects of phonotactic probability on the recognition of words in continuous speech*. (Doctoral dissertation, State University of New York, Buffalo, NY) Retrieved from https://www.

proquest.com/openview/633b57dffeba65d2b2bef477632b1a6f/1?pq-origsite=gscholar&cbl=18750&diss=y (Accession No. 9918207).

Goldwater, S., & Johnson, M. (2003). Learning OT constraint rankings using a maximum entropy model. *Proceedings of the Workshop on Variation within Optimality Theory*, 111–120.

Gordon, M. K. (1999). The "neutral" vowels of Finnish: How neutral are they. *Linguistica Uralica*, *35*(1), 17–21. DOI: https://doi.org/10.3176/lu.1999.1.03

Gordon, M. (2002). A factorial typology of quantity-insensitive stress. *Natural Language & Linguistic Theory*, *20*(3), 491–552. DOI: https://doi.org/10.1023/A:1015810531699

Gordon, M. (2007). Typology in Optimality Theory: *Language and Linguistics Compass*, *1*(6), 750–769. DOI: https://doi.org/10.1111/j.1749-818X.2007.00036.x

Hansson, G. Ó. (2010). *Consonant harmony: Long-distance interaction in phonology*. Berkeley: University of California Press.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, *39*(3), 379–440. DOI: https://doi.org/10.1162/ling.2008.39.3.379

Heinz, J. (2010). Learning long-distance phonotactics. *Linguistic Inquiry*, *41*(4), 623–661. DOI: https://doi.org/10.1162/LING_a_00015

Hyman, L. (2002). Is there a right-to-left bias in vowel harmony? *Proceedings of the 9th International Phonology Meeting*.

Iacoponi, L. (2015). *Directionality in consonant harmony and headed agreement by correspondence*. ROA 1244.

Kaun, A. R. (2004). The typology of rounding harmony. In B. Hayes, R. Kirchner & D. Steriade (Eds.), *Phonetically Based Phonology* (1st ed., pp. 87–116). Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511486401.004

Keshavarz, M. H., & Astaneh, H. (2004). The impact of binguality on the learning of English vocabulary as a foreign language (L3). *International Journal of Bilingual Education and Bilingualism*, *7*(4), 295–302. DOI: https://doi.org/10.1080/13670050408667814

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). *lmerTest: Tests in linear mixed effects models* [R package version 3.1.3]. https://cran.r-project.org/web/packages/lmerTest/index.html

Lai, R. (2015). Learnable vs. Unlearnable harmony patterns. *Linguistic Inquiry*, *46*(3), 425–451. DOI: https://doi.org/10.1162/LING_a_00188

Mahanta, S. (2007). *Directionality and locality in vowel harmony: With special reference to vowel harmony in Assamese* (Doctoral dissertation, Institution, Utrecht). Retrieved from https://hdl.handle.net/11858/00-001M-0000-0012-8F8D-E

Majors, T. (2006). The development of stress-dependent harmony. *Southwest Journal of Linguistics*, *25*(1), 59–83.

Martin, A., & Peperkamp, S. (2020). Phonetically natural rules benefit from a learning bias: A re-examination of vowel harmony and disharmony. *Phonology, 37*(1), 65–90. DOI: https://doi.org/10.1017/S0952675720000044

Martin, A., & White, J. (2021). Vowel harmony and disharmony are not equivalent in learning. *Linguistic Inquiry, 52*(1), 227–239. DOI: https://doi.org/10.1162/ling_a_00375

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*(4), 477–500. DOI: https://doi.org/10.1037/0096-3445.134.4.477

Mayer, C. (2021). Capturing gradience in long-distance phonology using probabilistic tier-based strictly local grammars. *Society for Computation in Linguistics, 4*(1), 39–50. DOI: https://doi.org/10.7275/231Q-P480

McCarthy, J. (2009). Harmony in harmonic serialism. *Linguistics Department Faculty Publication Series*.

McCollum, A. G., & Essegbey, J. (2020). Initial prominence and progressive vowel harmony in Tutrugbu. *Phonological Data and Analysis, 2*(3), 1–37. DOI: https://doi.org/10.3765/pda.v2art3.14

McMullin, K. J. (2016). *Tier-based locality in long-distance phonotactics: Learnability and typology* (Doctoral dissertation, University of British Columbia, Vancouver). https://doi.org/10.14288/1.0228114

McMullin, K. (2013). Learning consonant harmony in artificial languages. *Actes Du Congrès de l'ACL*.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience, 24*(4), 933–9477. DOI: https://doi.org/10.1162/jocn_a_00119

Nespor, M., & Vogel, I. (2007). *Prosodic phonology*. Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110977790

Obiri-Yeboah, M., & Rose, S. (2022). Vowel harmony and phonological phrasing in Gua. *Natural Language & Linguistic Theory, 40*(1), 159–193. DOI: https://doi.org/10.1007/s11049-021-09509-y

Ohala, J. J. (1994). Towards a universal, phonetically based theory of vowel harmony. *ICSLP 3*, 491–494. DOI: https://doi.org/10.21437/ICSLP.1994-113

Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language, 39*(3), 347–370. DOI: https://doi.org/10.1006/jmla.1998.2571

Prieto, P. (2006). Phonological phrasing in Spanish. In F. Martínez-Gil & S. Colina (Eds.), *Optimality-theoretic studies in Spanish phonology* (pp. 39–61). Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/la.99.03pri

Pycha, A., Nowak, P. M., Shin, E., & Shosted, R. (2003). Phonological rule-learning and its implications for a theory of vowel harmony. In M. Tsujimura & G. Garding (Eds.), *WCCFL 22 Proceedings* (pp. 101–114). Somerville, MA: Cascadilla Press.

R Core Team. (2022). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Ribeiro, E. R. (2002). Directionality in vowel harmony: The case of Karajá (Macro-Jê). *Annual Meeting of the Berkeley Linguistics Society*, *28*(1), Article 1. DOI: https://doi.org/10.3765/bls.v28i1.3859

Roca, I. (1991). Stress and syllables in Spanish. In H. Campos & F. Martínez-Gil (Eds.), *Current Studies in Spanish Linguistics* (pp. 599–635). Washington, DC: Georgetown University Press.

Rose, S., & Walker, R. (2004). A typology of consonant agreement as correspondence. *Language*, *80*(3), 475–531. DOI: https://doi.org/10.1353/lan.2004.0144

Rutgers, D., & Evans, M. (2017). Bilingual education and L3 learning: Metalinguistic advantage or not? *International Journal of Bilingual Education and Bilingualism, 20*(7), 788–806. DOI: https://doi.org/10.1080/13670050.2015.1103698

Salomé, F., Casalis, S., & Commissaire, E. (2022). Bilingual advantage in L3 vocabulary acquisition: Evidence of a generalized learning benefit among classroom-immersion children. *Bilingualism: Language and Cognition, 25*(2), 242–255. DOI: https://doi.org/10.1017/S1366728921000687

Tolskaya, I. K. (2008). *Oroch vowel harmony* (Master's thesis, Universitetet i Tromsø, Tromsø). Retrieved from https://munin.uit.no/handle/10037/1404

Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America, 126*(1), 367–376. DOI: https://doi.org/10.1121/1.3129127

Varela, S. (2012). Derivation and compounding. In J. I. Hualde, A. Olarrea & E. O'Rourke (Eds.), *The handbook of Hispanic linguistics* (pp. 209–226). Malden, MA: Wiley-Blackwell. DOI: https://doi.org/10.1002/9781118228098.ch11

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40*(3), 374–408. DOI: https://doi.org/10.1006/jmla.1998.2618

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers, 36*(3), 481–487. DOI: https://doi.org/10.3758/BF03195594

Weir, A. (2012). Left-edge deletion in English and subject omission in diaries. *English Language & Linguistics, 16*(1), 105–129. DOI: https://doi.org/10.1017/S136067431100030X

White, J., Kager, R., Linzen, T., Markopoulos, G., Martin, A., Nevins, A., Peperkamp, S., Polgárdi, K., Topintzi, N., & van de Vijver, R. (2018). Preference for locality is affected by the prefix/suffix asymmetry: Evidence from artificial language learning. In S. Hucklebridge & M. Nelson (Eds.), *NELS 48: Proceedings of the Forty-Eighth Annual Meeting of the North East Linguistic Society: Vol. 3,* 207–220. Amherst, MA: GLSA. Retrieved from https://dspace.library.uu.nl/bitstream/handle/1874/421973/White_Kager_Linzen_Markopoulos_Martin_Nevins_Peperkamp_Polgardi_Topintzi_van_de_Vijver_2018_NELS_48.pdf?sequence=1

Wilson, C. (2003). Experimental investigation of phonological naturalness. In G. Garding & M. Tsujimura (Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, 533–546.

Winn, M. (2013). *Scale intensity* [Praat Script]. http://www.mattwinn.com/praat/Scale_intensity_check_maxima_v3.txt

Woods, K. J. P., Siegel, M., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception & Psychophysics*, *79*(7), 2064–2072. DOI: https://doi.org/10.3758/s13414-017-1361-2

Zsiga, E. (2013). *The sounds of language: An introduction to phonetics and phonology*. Hoboken: Wiley-Blackwell. DOI: https://doi.org/10.1002/9781394260980

Zymet, J. (2014). *Distance-based decay in long-distance phonological processes: A probabilistic model for Malagasy, Latin, English, and Hungarian* (Master's thesis, University of California, Los Angeles, CA). https://escholarship.org/uc/item/20n7n3gj