In Sherry Hucklebridge & Max Nelson (eds.). 2018. *NELS 48: Proceedings of the Forty-Eighth Annual Meeting of the North East Linguistic Society: Vol. 3*, 207–220. Amherst, MA: GLSA.

Preference for locality is affected by the prefix/suffix asymmetry: Evidence from artificial language learning*

James White[1], René Kager[2], Tal Linzen[3,4], Giorgos Markopoulos[5], Alexander Martin[3,6], Andrew Nevins[1], Sharon Peperkamp[3], Krisztina Polgárdi[7], Nina Topintzi[5], & Ruben van de Vijver[8]

[1]University College London, [2]Utrecht University, [3]Laboratoire de Sciences Cognitives et Psycholinguistique (EHESS – DEC-ENS – CNRS – PSL), [4]Johns Hopkins University, [5]Aristotle University of Thessaloniki, [6]University of Edinburgh, [7]Hungarian Academy of Sciences, [8]Düsseldorf University

1. Introduction

Previous research indicates that learners have a strong bias in favor of locality when learning co-occurrence restrictions in an artificial language (Finley 2011, 2015; McMullin & Hansson 2014; McMullin 2016). This preference for locality has been found even for consonant harmony patterns, which are often nonlocal in natural languages (McMullin & Hansson 2014, Finley 2015, McMullin 2016). For instance, McMullin and Hansson (2014) found that learners in an artificial language experiment generalized a nonlocal sibilant harmony pattern to local contexts, but not *vice versa*.

However, it is not the case that learners *always* prefer locality when learning co-occurrence restrictions. Endress and Mehler (2010) found that adult learners were better at learning dependencies between two consonants, C_1 and C_2, when they were at the beginning and end of a word (**C_1**vccv**C_2**) than when they were adjacent to each other in

* We would like to thank Andrew Clark, Martin Rönsch, and Remco van der Veen for assistance in setting up and conducting the experiment. We also thank Sara Finley, Gunnar Hansson, Rachel Walker, and the audiences of NELS 48 and MFM 25 for helpful comments on this work. Funding for this work was provided by grants from the Deutsche Forschungsgemeinschaft and the British Academy/Leverhulme Trust.

word-medial position (cv**C_1C_2**vc). These results suggest that learners may preferentially code salient prosodic edges during learning, which could favor nonlocal dependencies when the relevant sounds are adjacent to salient edges (but cf. Lai 2015). In the current study, we investigate whether prosodic structure affects the preference for locality when learning novel vowel dependencies in an artificial language. We focus on two aspects of prosodic structure: affix type (suffix vs. prefix) and stress location.

Cross-linguistically, we find that suffixes are more likely than prefixes to participate in vowel harmony, particularly in affix-controlled harmony systems (Bakovic 2000, Finley & Badecker 2009, Nevins 2010). One potential explanation for this asymmetry could lie in the hierarchical prosodic structure of words. Researchers have independently argued for a prosodic structure in which suffixes are more closely integrated with roots than prefixes are (Nespor & Vogel 1986, Peperkamp 1997). Asymmetries between prefixes and suffixes are found for a variety of phonological phenomena beyond vowel harmony. For instance, in Samoan root+suffix forms mostly behave like monomorphemic words in terms of basic foot assignment and diphthong creation, but prefix+root forms behave as if there is an intervening prosodic boundary (Zuraw et al. 2014). If the root and suffixes form their own domain to the exclusion of prefixes, this smaller domain could act as a preferential domain for vowel harmony (see Hyman 2002).

There is also reason to believe that stress (or word prominence) could play a role in vowel harmony systems. Vowels in strong positions, such as the stressed syllable of the root, might act as preferential triggers for vowel harmony (Hyman 2002). For instance, height harmony spreads leftward from a stressed syllable in Pasiego Spanish (Hualde 1989, Kaisse 2016). Moreover, metaphony-style systems involve dependencies between a stressed vowel and a following vowel, which is often an affix (Walker 2005). In some metaphony systems, this restriction can even be nonlocal in nature (Walker 2004).

In the current study, we tested whether affix type (prefix or suffix) and stress location would influence the likelihood that learners would infer a local harmony pattern when presented with ambiguous input. Using the 'poverty of the stimulus' paradigm (Wilson 2006), we exposed learners to training data that were ambiguous between a local harmony pattern and a nonlocal dependency. Participants heard CVCV nonce roots paired with corresponding affixed forms (CV-CVCV or CVCV-CV). During exposure, roots always followed front/back harmony, and affix vowels alternated in backness depending on the root vowels. Because exposure roots were always fully harmonic, it was ambiguous whether the local or nonlocal root vowel was triggering the alternations in the affix vowels. Participants were later tested on disharmonic roots, where they were forced to choose between matching the affix vowel to the local root vowel or to the nonlocal root vowel. Their responses on disharmonic roots at test therefore allowed us to measure their preference for the local harmony pattern (compared to a nonlocal dependency) in each of our conditions.[1]

We manipulated two variables between-subjects, Affix Type (prefixes or suffixes) and Stress Location (local root vowel or nonlocal root vowel), for a total of four conditions. We focus on testing the following hypotheses:

[1] 'Local' vowel here assumes the vowels are adjacent on a vowel tier (e.g. Goldsmith 1979), and thus can be considered local though they have an intervening consonant.

(1) *Hypotheses*
  a. Learners have an overall preference for locality.
  b. The preference for locality is stronger between the root and a suffix than between the root and a prefix.
  c. The preference for locality is stronger if the local vowel is stressed than if the nonlocal vowel is stressed.

Hypothesis (a) follows from previous work showing that learners generally have a strong bias for locality when learning phonological patterns (e.g. McMullin & Hansson 2014, Finley 2015, McMullin 2016). Hypothesis (b) follows from the prefix/suffix asymmetry in prosodic structure; if the suffix is more closely integrated with the root than the prefix is, we expect learners to choose local harmony more frequently for suffixes than for prefixes. Hypothesis (c) follows from the prediction that the more prominent vowel (i.e. the stressed root vowel) will be considered a more likely trigger for the vowel alternations, all else being equal. Combining hypotheses (b) and (c), we expected the greatest locality preference for suffixes with stress on the local root vowel, and the weakest locality preference for prefixes with stress on the nonlocal root vowel.

This study also falls within our broader goal of exploring first language (L1) transfer and replicability in artificial grammar learning (AGL) experiments. AGL experiments have emerged as a dominant paradigm for exploring the biases active during language learning, both in phonology (e.g. Pycha et al. 2003; Wilson 2006; Peperkamp, Skoruppa, & Dupoux 2006; Baer-Henney & van de Vijver 2012; White 2014; White & Sundara 2014) and in syntax (e.g. Culbertson, Smolensky, & Legendre 2012). Unlike real language learning, AGL experiments allow researchers to have full control over the type and amount of linguistic input that learners encounter, and they allow us to test learning outcomes after very brief lab exposure to minimally different languages.

AGL experiments are typically used to explore cognitive biases assumed to be universal to all learners, but AGL experiments always carry with them the limitation that participants already speak an L1, which undoubtedly has some effect on learning in these tasks. However, it remains unclear to what extent, and in what manner, results in AGL studies vary across different L1 groups because AGL studies are generally conducted with participants who all speak the same L1. To truly explore this issue, it is necessary to conduct the same set of experiments with groups of participants with different linguistic backgrounds. In this paper, we report on our initial endeavor to address this issue by conducting the same experiment across labs in five countries with speakers of five L1s.

2. Method

2.1 Participants

A total of 277 participants completed the study across five countries. All participants were native speakers of the dominant language in the country where they were tested. (We will refer to this language as their L1.) An additional 78 participants were tested but did not complete the experiment because they never passed the verification phase (see Procedure section). Table (1) provides a breakdown of the number of participants for each L1 along with the testing location. No participants reported any hearing or language impairments. Participants received a small amount of money or course credit as compensation.

(2) *Number of participants according to L1/testing location*

L1	Testing location	Number of participants	Additional non-completers
Dutch	Utrecht University	77	3
English	University College London	33	32
French	Laboratoire de Sciences Cognitives et Psycholinguistique, Paris	38	10
German	Düsseldorf University	54	28
Greek	Aristotle University of Thessaloniki	75	5

2.2 Materials

The nonce words used in the experiment consisted of CVCV roots and corresponding affixed forms with a CV suffix or prefix (depending on condition). Consonants were limited to the set [p, g, z, n] for root-initial position and [p, t, d, g, z, m, n, l] for root-medial position. The vowels included two front vowels [i, e] and two back vowels [u, o]. There was one high-vowel affix [fi] ~ [fu] and one mid-vowel affix [be] ~ [bo]. We chose the segmental inventory such that it included only sounds existing as part of a phonemic contrast in all five of the L1s that we tested. This precaution was aimed at ensuring that all participants could accurately distinguish the segment categories. We decided to further restrict the initial consonants to avoid similarities with common function words in some of our languages (e.g. *les* [le], *mes* [me], and *tes* [te] in French). We also avoided similarities to real words or phrases in any of the languages wherever possible. Stress always appeared on the first or last syllable of the CVCV root, depending on condition; stress was not shifted in affixed forms.

Exposure phase. The exposure phase stimuli consisted of eight CVCV roots, one for each of the logically possible harmonic vowel sequences. Four roots contained front vowels ([gini], [pime], [zedi], [nege]) and four roots contained back vowels ([pulu], [nupo], [gozu], [zoto]). Each root was paired with the two affixes, [fi]/[fu] and [be]/[bo], either as suffixes or prefixes depending on condition. The vowel of the affix always matched the vowels of the root in terms of backness; for instance, in the suffix condition the front root [zedi] was paired with the affixed forms [zedi-fi] and [zedi-be] (or [fi-zedi] and [be-zedi] in the prefix condition). The back root [nupo] was paired with the forms [nupo-fu] and [nupo-bo] in the suffix condition. The affixed forms represented the plural and diminutive forms of the words; whether [fi]/[fu] or [be]/[bo] represented the plural or diminutive affix was counterbalanced across participants. For the version of the experiment where pictures were shown (see Procedure section below), each root was also paired with a picture of an everyday object or animal. The affixed forms were then paired with a plural picture or a small version of the picture depending on affix's meaning, either plural or diminutive respectively.

Verification phase. Eight new harmonic roots, along with corresponding affixed forms, were created for the verification phase following the same method described for the exposure phase. There were again four roots containing front vowels ([zipi], [gile], [peti], [neze]) and four roots containing back vowels ([nugu], [zuno], [pomu], [godo]). We were careful to avoid accidental regularities across the exposure and verification phases that could have allowed participants to succeed without learning the vowel co-occurrences (e.g. by ensuring that specific consonants did not regularly occur with front or back vowels).

Generalization phase. The generalization phase included harmonic roots as well as disharmonic roots. There were 8 new harmonic roots, one for each of the logically possible harmonic vowel combinations, and 32 disharmonic roots, 4 for each of the logically possible disharmonic vowel combinations. Each of the roots was created following the same restrictions used in previous phases. Affixed forms were created by adding [fi]/[fu] and [be]/[bo] before or after each root, depending on condition. Because participants were always choosing between the front and back affix allomorphs, affixed forms with both versions of the affix (front and back) were created and recorded.

Stimulus recording and testing apparatus. The stimuli were recorded in a sound-attenuated booth by a phonetically trained, native speaker of Hebrew (the third author). Hebrew is unrelated to any of our participants' L1s, but it contains all of the phonemes that we used in the experiment. The sounds in the stimuli were pronounced as they would be in Hebrew, except that a small amount of aspiration was added to [p] and [t] to make the contrast between [t] and [d] more clear for our English-speaking participants. (English has an aspirated/unaspirated distinction for word-initial stops rather than a true voiced/voiceless distinction.) The stimuli were equalized for intensity.

The study was conducted on a computer in a quiet room within each of the respective labs. The experiment was coded in Python, using Pygame. The same experiment script was used across all labs, with the instructions modified according to participant L1.

2.3 Procedure

There were three phases in the experiment: exposure, verification, and generalization. Participants were instructed that they would be learning words in a foreign language, and that they would later be tested on what they had learned. English and German participants had a slightly different procedure compared to Dutch, French, and Greek participants.

In each exposure trial, English and German participants saw a singular picture on the left side of the screen. After a 500 ms pause, the root word corresponding to that picture was played through headphones, after which the picture disappeared. Next, the appropriate picture for the affixed form (either a plural picture or a small version of the singular picture) appeared on the right side of the screen, and the affixed form (plural or diminutive) was played through headphones after a 500 ms pause. The vowel of the affix always matched the backness of the root vowels during exposure. After hearing both words, participants pressed the spacebar to move on to the next trial. The exposure phase included a total of 16 trials (8 roots, each presented with 2 affixes). The order of trials was randomized anew for each participant.

After exposure, participants completed a verification phase in which they were tested using a two-alternative forced-choice task. A verification trial was identical to an exposure trial except that participants heard two options for the affixed form, one with the (correct) harmonic affix vowel and one with the (incorrect) disharmonic affix vowel (e.g. [gile]…[gile-fi]…[gile-fu]). Participants were instructed to select the correct option for the affixed form by pressing a key marked '1' for the first option that they heard or a key marked '2' for the second option. There were 16 randomized trials in the verification phase (8 roots, each presented with 2 affixes). Throughout the experiment, participants were instructed to follow their intuition if they were unsure of the correct answer.

Participants were required to answer at least 13 out of 16 trials (81%) correctly to proceed to the generalization phase. This criterion is comparable to the 80% criterion used in White 2014. The probability of participants reaching the criterion due to random chance was .01. The purpose of training to criterion is to ensure that participants have learned some pattern from the harmonic roots before testing their generalization of that pattern. Participants who answered fewer than 13 trials correctly were required to cycle through the exposure and verification phases until they reached the criterion. The same trials appeared in each cycle, but in a different randomized order. A screen appeared after each verification phase showing participants their percentage correct. Participants who were unable to reach the criterion after 5 cycles were taken to the end of the experiment without completing the generalization phase; their data were excluded from all analyses.

In the generalization phase, participants were tested in novel harmonic and disharmonic roots. The trial procedure was identical to that of verification trials. Participants were told that they would be tested on new words from the same language, but were not made aware of any differences between the phases. There were a total of 80 randomized trials in the generalization phase (8 harmonic roots and 32 disharmonic roots, each paired with 2 affixes). After every 20 trials, participants were required to take a break of at least 20 seconds, during which a screen reminded them of the instructions.

Dutch, French, and Greek participants followed a similar procedure with a couple of modifications. First, the pictures were removed from the experiment completely; instead, a fixation point (+) appeared in the middle of the screen to mark the beginning of each trial. Second, during the exposure phase, trials were blocked by affix height such that participants heard all of the [fi]/[fu] affixed forms first, or all of the [be]/[bo] affixed forms first (counterbalanced). To compensate for the lack of pictures, participants were instructed that they would be hearing singular/plural pairs of words or plain/diminutive pairs of words before the relevant block of exposure trials (along with a brief explanation of the terms 'plural' and 'diminutive'). We counterbalanced which affix meant plural or diminutive across participants. The forms were blocked by affix height only during the exposure phase, not during the verification and generalization phases.[2]

[2] We first ran the experiment with English and German participants, but the attrition rate due to participants failing to pass the verification phase was very high (49% for English; 34% for German). Before running the experiment with the other language groups, we decided to modify the procedure with the aim of reducing the amount of attrition. We do not expect these changes (removing the pictures or blocking the exposure phase by affix height) to have a drastic effect on the results given that participants were trained to criterion before moving on to the generalization phase; nonetheless, we are currently

3. Results

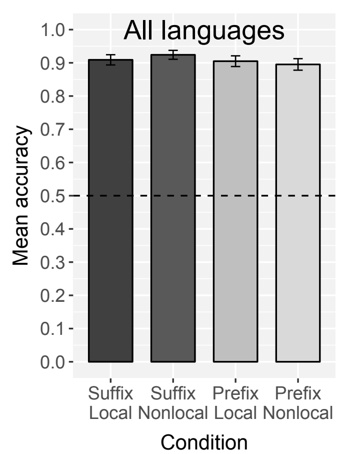
We analyzed the results with mixed effects logistic regression models (see Jaegar 2008), implemented in R (R Core Team 2016) using the *lme4* package (Bates et al. 2015). The initial models included fixed effects for Affix Type (suffix or prefix), Stress Location (local or nonlocal root vowel), and an Affix Type by Stress Location interaction. The model included random intercepts for Subject and Item; as the experimental design is fully between subjects, random slopes were not included.

We used backwards stepwise comparison to reach the final model, removing fixed effects one at a time if they did not significantly improve the model fit according to a likelihood ratio test (using the *anova()* function in R, see Baayen 2008). If the interaction significantly improved model fit, we retained all of the component main effects as well, even if they were non-significant (i.e. we did not reduce the model at all). Otherwise, we removed the interaction effect first, followed by the main effect that provided the least improvement to the model (which turned out to always be Stress Location).

3.1 Harmonic roots

Figure (3) shows the overall mean accuracy for harmonic roots in the generalization phase, according to Affix Type and Stress Location. Because participants were trained to criterion on harmonic roots during the exposure and verification phases, we expected accuracy to be high for these items across the board. As seen in Figure (3), the overall mean accuracy for harmonic roots was indeed very high (91% overall), and it was comparably high across the four conditions. The final model for harmonic roots contained only a significant intercept ($\beta = 3.07$, $z = 21.59$, $p < .01$) reflecting an overall accuracy significantly greater than chance. The effects for Affix Type, Stress Location, and their interaction all failed to significantly improve the model (all $p > .4$).

(3) *Mean accuracy for harmonic roots in the generalization phase. The dotted line shows 50% chance level.*



running the modified version of the experiment with English and German participants to provide an equal comparison across all language groups.

Accuracy on harmonic roots was comparably high across all five L1 groups (ranging from 88% to 92% overall). As the harmonic roots were not our primary interest, we do not analyze them further.

3.2 Disharmonic roots

Figure (4) shows the results for disharmonic roots in the generalization phase, according to Affix Type and Stress Location. There was no 'correct' response for these items; instead, we report the proportion of trials in which participants chose to match the affix vowel to the local root vowel (as opposed to the nonlocal root vowel). Results are shown for all of the L1 groups combined (first panel), as well as for each L1 group individually.

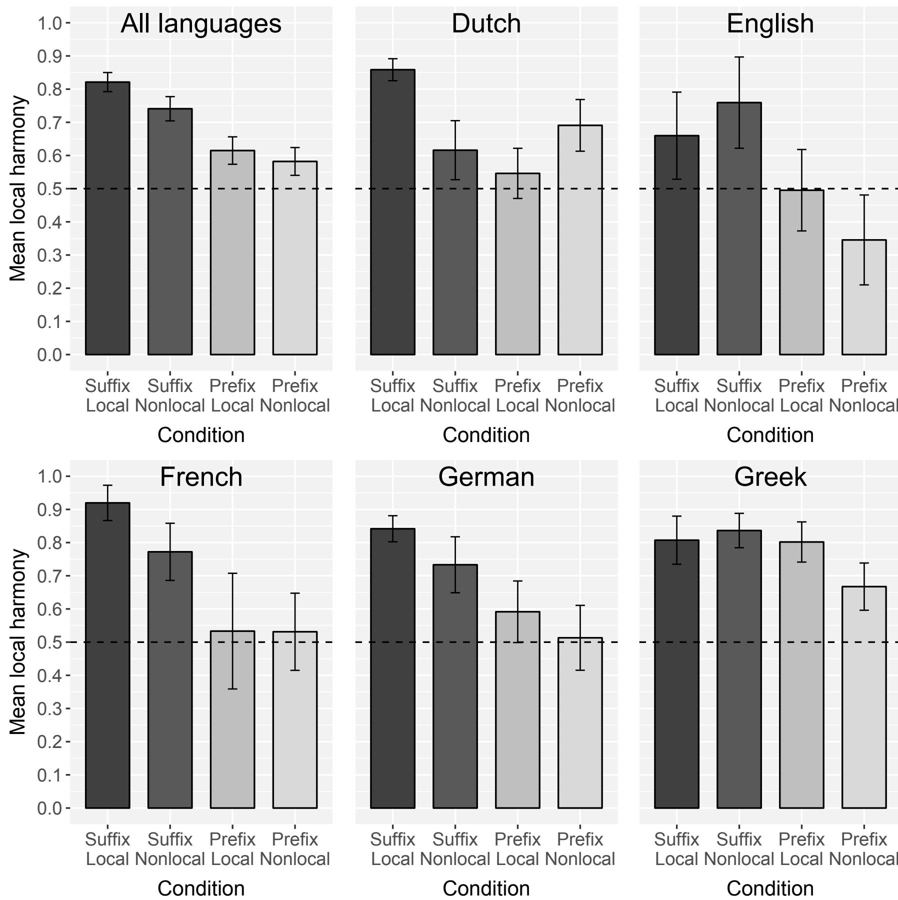
Looking at Figure (4), we see some broad patterns in the results. Overall, there is a greater preference for local harmony in the suffix conditions (dark bars) than in the prefix conditions (light bars). In the suffix conditions, we see a robust preference for locality across the L1 groups, reflected by the fact that the dark bars are virtually always well above the 50% chance line (which would mark no preference). Preference for locality is overall much weaker in the prefix conditions. Turning to Stress Location, we see a slightly greater preference for local harmony overall when the local vowel is stressed, but the difference is small and not consistent across L1 groups.

For the statistical analysis, we ran separate models for the overall data (all L1 groups combined) and for each L1 group individually, following the process described in section 3.1. An overview of the statistical results is provided in Table (5). When considering all of the data together, only the main effect of Affix Type is significant. Neither the main effect of Stress Location nor the interaction effect resulted in a significant improvement to the model. Three of the L1 groups—English, French, and German—follow the same general pattern found in the overall results. For Dutch, only the interaction effect reaches significance, indicating that there is a significantly greater preference for locality only when there is both a suffix and local stress. For Greek, there are no significant differences between the conditions. Though the Greek data numerically suggest a potential interaction effect (i.e. a lower preference for locality if there is both a prefix and nonlocal stress), the interaction does not reach significance in the model (p = .12).[3]

Looking at individual results, we see different distributions for the participants in the suffix conditions and those in the prefix conditions in terms of how often they chose agreement with the local vowel, as shown in the histograms in Figure (6). In the suffix conditions, we see a fairly unimodal distribution around 1, indicating that most participants chose the local vowel consistently, near 100% of the time. In the prefix conditions, by contrast, we see a more trimodal distribution. Though a large group of participants consistently chose the local vowel, another sizeable group of participants consistently chose the nonlocal vowel, and a third group (bunched around the 50% chance line) showed no consistent preference. These distributions further support the conclusion that there was a strong preference for local harmony for suffixes, but the preference was much weaker for prefixes.

[3] This is true even if we change which condition is coded as the baseline in the model.

(4) *Mean local harmony chosen for disharmonic roots in the Generalization phase. The dotted line shows 50% chance level.*

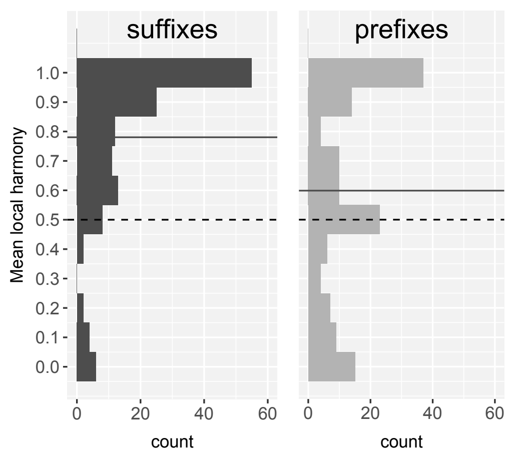


(5) *Summary of statistical analysis, overall and by individual L1 group.*[4] *Non-significant effects are shaded.*

L1 group	Affix Type = *Suffix*	Stress = *Local*	Interaction (*Suffix & Local*)
All languages	$z = 4.46, p < .01$	$\chi^2 = 1.80, p = .18$	$\chi^2 = .15, p = .70$
Dutch	$z = -1.06, p = .29$	$z = -1.43, p = .15$	$z = 2.76, p < .01$
English	$z = 2.17, p = .03$	$\chi^2 = .45, p = .50$	$\chi^2 = 1.38, p = .24$
French	$z = 2.53, p = .01$	$\chi^2 = .41, p = .52$	$\chi^2 = 1.30, p = .25$
German	$z = 3.27, p < .01$	$\chi^2 = .59, p = .44$	$\chi^2 = .05, p = .82$
Greek	$\chi^2 = 1.87, p = .17$	$\chi^2 = .64, p = .42$	$\chi^2 = 2.36, p = .12$

[4] The prefix-nonlocal condition was coded as the baseline for all models. The numbers reported in this table were generated as follows. If the interaction effect was significant (as in Dutch), we report the Wald z and p-value for all effects in that model. Otherwise, if an effect was removed from the final model, we report the likelihood ratio test values (χ^2 and p-value) that resulted in its removal; if an effect remained significant in the final model, we report its Wald z and p-value in the final model.

(6) *Histograms showing proportion of local harmony chosen for participants exposed to suffixes (left) and prefixes (right). Dotted lines show the 50% chance level and solid lines show the aggregate means.*



4. Discussion

To summarize the results, we found that participants exposed to a pattern that was ambiguous between a local harmony pattern and a nonlocal vowel dependency pattern had an overall preference for the local pattern. This bias towards locality is consistent with previous studies (e.g. Finley 2011, 2015; McMullin & Hansson 2014; McMullin 2016). However, the preference for locality varied in our study depending on the type of affix involved. We found a greater preference for locality for suffixes than for prefixes, a difference that was robust across the L1 groups that we tested (with the possible exception of Greek). In terms of stress, we found no significant differences in any of our L1 groups based on the location of stress, except for Dutch where the combination of a suffix and local stress resulted in a greater locality preference (an interaction effect).

The asymmetry that we found between prefixes and suffixes is consistent with the view that suffixes are more closely integrated with roots than prefixes are (Nespor & Vogel 1986, Peperkamp 1997), and that the root+suffix domain (excluding the prefix) is a preferred domain for vowel harmony (Hyman 2002). Under such an account, participants in the suffix conditions were strongly biased to interpret the exposure pattern as a case of local harmony, which in turn would cause them to choose the local vowel as the trigger in disharmonic roots at test. On the other hand, participants in the prefix conditions were not as strongly biased towards local harmony due to the greater prosodic boundary between a prefix and a root, so they were more likely to settle on the alternative nonlocal dependency, or even an analysis with no clear trigger at all (hence the large number of participants in the prefix conditions with no preference). It is not the case that participants in the prefix conditions failed to learn anything—we trained them to criterion, and they were as accurate on harmonic roots as participants in the suffix conditions. Many of them just showed no preference when it came to disharmonic roots.

A potential motivation for the nonlocal dependency analysis is that it involves the first and last vowels of the words. Endress and Mehler (2010) argued that word edges serve as particularly salient cues when learners are acquiring novel phonological dependencies. In their study, learners found it easier to learn a nonlocal dependency between two consonants at the word edges ($\mathbf{C_1}$vccv$\mathbf{C_2}$) than a local dependency in word-medial position (cv$\mathbf{C_1C_2}$vc). If locality and word-edge salience are competing pressures during phonological dependency learning, it may be that participants in the prefix conditions were more evenly influenced by the two alternatives, whereas in the suffix conditions, the (stronger) locality bias outweighed the salience of the word edges.

Typologically, first-last phonological dependencies appear to be unattested (or at least rare) in the world's languages, despite involving word edges. In a recent study, Lai (2015) found that learners in an AGL experiment were unable to learn a first-last sibilant dependency (i.e. initial and final sibilants must agree in anteriority, regardless of intervening segments), but they could easily learn a sibilant harmony pattern (i.e. subsequent sibilants must agree in anteriority). The sibilant harmony pattern is local if autosegmental tiers are assumed, but the first-last pattern is not. In a follow-up study using a different task, Avcu (2018) found that the first-last sibilant dependency was learned, but not as well as the sibilant harmony pattern. Avcu raises the possibility that locality (defined more precisely as the Subregular Hypothesis; Heinz 2010) holds as a strong bias specific to language learning, but learners can also rely on the salience of word edges as a domain-general learning mechanism. Along these lines, a possible interpretation of our results could be that the strong linguistic bias for locality is triggered particularly within the root+suffix domain, but outside of this domain learners are more likely to rely on other learning strategies, such as using word edges as salient anchors.

Another point worth noting is that the current experiment involved a root-controlled harmony pattern, meaning that the root vowels determined the form of the affix vowels. For affix-controlled harmony (which our study did not test), the typological asymmetry between prefixes and suffixes is fairly clear: harmony triggered by prefixes is uncommon, and prefixes triggering harmony asymmetrically implies that suffixes will also trigger harmony (Bakovic 2000, Finley & Badecker 2009, Nevins 2010). The typology for root-controlled harmony is less clear, in part because many vowel harmony languages are predominantly suffixing. However, the results of the current study provide experimental support for the view that an asymmetry exists between prefixes and suffixes not only for affix-controlled harmony, but also for root-controlled harmony.

At first glance, our results appear to diverge from the findings of Finley and Badecker (2009), who concluded that there was a bias against prefixes as harmony triggers, but not as harmony targets. Their study tested both affix-controlled and root-controlled harmony patterns in an artificial language experiment. In the case of affix-controlled harmony, participants found it more difficult to learn that prefixes triggered harmony compared to suffixes (e.g. a pattern such as [beme]…[mu-bomo] was harder to learn than one such as [beme]…[bomo-mu]). For root-controlled harmony, Finley and Badecker trained participants with suffixes or prefixes and then tested whether they generalized the pattern to the other type of affix. They found no difference in how often participants generalized from prefixes to suffixes or *vice versa* in the case of root-controlled harmony.

However, Finley and Badecker's (2009) experiment was unable to distinguish between whether participants learned a local harmony pattern or a nonlocal vowel dependency. Their training data were very similar to the training data presented to our participants: pairs of words such as [beme] followed by [mi-beme] or [beme-mi] depending on condition. Our results suggest that many participants will in fact learn a nonlocal dependency from such training data, particularly in the prefix condition. At test, participants in Finley and Badecker's study completed a forced-choice test (e.g. [beme-gi] vs. [beme-gu]) that included the untrained affix type. Crucially, however, participants were tested only on harmonic roots. With harmonic roots, a local harmony pattern and a nonlocal dependency both predict the same outcome, a harmonic affix. Only by testing disharmonic roots can we disambiguate which type of pattern was learned. Thus, even though Finley and Badecker found similar performance for harmonic roots in the prefix and suffix conditions (as did we), our findings suggest that many participants in the prefix condition would have in fact settled on a different (nonlocal) analysis.

Turning to the stress component of our study, we found no clear effect of stress on learners' preference for locality. At one level, this result is surprising given that stress plays a role in a number of vowel-vowel dependencies in natural languages. Though stressed vowels can serve as triggers or targets for harmony patterns, it may be that a learning preference only arises for stressed vowels as potential targets, rather than as potential triggers (the latter being the case in the current study). Under licensing accounts (Walker 2004, 2005), we would indeed expect stressed vowels to serve as preferred targets for harmony, though not necessarily as preferred triggers. Further work comparing stressed vowels as potential targets and triggers of harmony would be needed to test this possibility. Another possibility is that stress is an area where the L1 of our participants had a notable effect. Our five L1s differ in terms of the role that stress plays in their phonology. We might expect the largest effect of stress for participants with an L1 where word-level stress plays the largest role (Dutch, German), and the smallest effect (perhaps no effect) where word-level stress plays the smallest role (French). We plan to look at this factor in detail once we have collected more data from all L1 groups.

A limitation of the work at present is that all of our L1s are predominantly suffixing languages. It remains a possibility that the prefix/suffix asymmetry observed in the experiment was due to the general suffixing preference of our participants' L1s. If the L1 independently motivates a word structure where suffixes are more closely linked to roots, then our results could stem from L1 transfer rather than a universal asymmetry between suffixes and prefixes. To address this issue, we would like to conduct the same experiment with speakers of a predominantly prefixing language, or at least a language where prefixes and suffixes are more evenly distributed. We also plan to conduct the experiment with speakers of a vowel harmony language (e.g. Hungarian or Turkish) to explore how learners who already have a vowel harmony pattern in their L1 will differ from those who do not.

To conclude, we have shown that learners have an overall bias in favor of locality when acquiring novel vowel-vowel dependencies, but the preference for locality is much stronger across a suffix boundary than across a prefix boundary. The results support the view that the root and suffix(es) form a preferential domain for (local) vowel harmony to the exclusion of prefixes, and this distinction could play a role in explaining typological asymmetries between prefixes and suffixes in vowel harmony languages.

References

Avcu, Enes. 2018. Experimental investigation of the Subregular Hypothesis. In *Proceedings of the 35th West Coast Conference on Formal Linguistics*, ed. by Wm. G. Bennett, Lindsay Hracs, and Dennis Ryan Storoshenko, 77–86. Somerville, MA: Cascadilla.

Baayen, R. H. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baer-Henney, Dinah, & Ruben van de Vijver. 2012. On the role of substance, locality, and amount of exposure in the acquisition of morphophonemic alternations. *Laboratory Phonology* 3:221–250.

Bates, Douglas, Martin Mächler, Ben Bolker, & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using {lme4}. *Journal of Statistical Software* 67:1–48.

Baković, Eric. 2000. *Harmony, dominance and control.* Doctoral dissertation, Rutgers, The State University of New Jersey.

Culbertson, Jennifer, Paul Smolensky, & Géraldine Legendre. 2012. Learning biases predict a word order universal. *Cognition* 122:306–329.

Endress, Ansgar D., & Jacques Mehler. 2010. Perceptual constraints in phonotactic learning. *Journal of Experimental Psychology: Human Perception and Performance* 36: 235–250.

Finley, Sara. 2011. The privileged status of locality in consonant harmony. *Journal of Memory and Language* 65:74–83.

Finley, Sara. 2015. Learning non-adjacent dependencies in phonology: Transparent vowels in vowel harmony. *Language* 91:48–72.

Finley, Sara, & William Badecker. 2009. Right-to-left biases for vowel harmony: Evidence from artificial grammar. In *Proceedings of the 38th Annual Meeting of the North East Linguistic Society, Vol. 1*, ed. by Anisa Schardi, Martin Walkow, and Muhammad Abdurrahman, 269–282. Amherst: University of Massachusetts, Graduate Linguistic Student Association.

Goldsmith, John. 1979. *Autosegmental phonology*. New York: Garland.

Heinz, Jeffrey. 2010. Learning long-distance phonotactics. *Linguistic Inquiry* 41:623–661.

Hualde, José I. 1989. Autosegmental and metrical spreading in the vowel-harmony systems of northwestern Spain. *Linguistics* 27:773–805.

Hyman, Larry. (2002). Is there a right-to-left bias in vowel harmony? Ms., University of California, Berkeley.

Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59:434–446.

Kaisse, M. Ellen. 2016. The syllabic position of glides in Spanish: insights from Pasiego vowel harmony. In *The syllable and stress: Studies in honor of James W. Harris*, ed. by Rafael A. Núñez-Cedeño, 29–50. De Gruyter Mouton.

Lai, Regine. 2015. Learnable vs. unlearnable harmony patterns. *Linguistic Inquiry* 46:425–451.

McMullin, Kevin. 2016. Tier-based locality in long-distance phonotactics: Learnability and typology. Doctoral dissertation, University of British Columbia, Vancouver.

McMullin, Kevin, & Gunnar Ólafur Hansson. 2014. Locality in long-distance phonotactics: evidence for modular learning. In *Proceedings of the 44th Annual Meeting of the North East Linguistic Society, Vol. 2*, ed. by Jyoti Iyer and Leland Kusmer, 1–14. Amherst: University of Massachusetts, Graduate Linguistic Student Association.

Nespor, Marina, & Irene Vogel. 1986. *Prosodic phonology.* Dordrecht: Foris.

Nevins, Andrew. 2010. *Locality in vowel harmony.* Cambridge, MA: MIT Press.

Peperkamp, Sharon. 1997. *Prosodic words.* HIL dissertations 34. The Hague: Holland Academic Graphics.

Peperkamp, Sharon, Katrin Skoruppa, & Emmanuel Dupoux. 2006. The role of phonetic naturalness in phonological rule acquisition. In *Proceedings of the 30th Boston University Conference on Language Development*, ed. by David Bamman, Tatiana Magnitskaia, and Colleen Zaller, 464–475. Somerville, MA: Cascadilla.

Pycha, Anne, Pawel Nowak, Eurie Shin, & Ryan Shosted. 2003. Phonological rule-learning and its implications for a theory of vowel harmony. In *WCCFL 22: Proceedings of the 22nd West Coast Conference on Formal Linguistics*, ed. by Gina Garding and Mimu Tsujimura, 423–435. Somerville, MA: Cascadilla.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from http://www.r-project.org/.

Walker, Rachel. 2004. Vowel features licensing at a distance: Evidence from Northern Spanish language varieties. In *WCCFL 23: Proceedings of the 23rd West Coast Conference on Formal Linguistics*, ed. by Vineeta Chand, Ann Kelleher, Angelo J. Rodríguez, and Benjamin Schmesier, 773–786. Somerville, MA: Cascadilla.

Walker, Rachel. 2005. Weak triggers in vowel harmony. *Natural Language & Linguistic Theory* 23:917–989.

White, James. 2014. Evidence for a learning bias against saltatory phonological alternations. *Cognition* 130:96–115.

White, James, & Megha Sundara. 2014. Biased generalization of newly learned phonological alternations by 12-month0old infants. *Cognition* 133:85–90.

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

Zuraw, Kie, Kristine Yu, & Robyn Orfitelli. 2014. Word-level prosody of Samoan. *Phonology* 31:271–327.

James White, René Kager, Tal Linzen, Giorgos Markopoulos, Alexander Martin, Andrew Nevins, Sharon Peperkamp, Krisztina Polgárdi, Nina Topintzi, Ruben van de Vijver
j.c.white@ucl.ac.uk
R.W.J.Kager@uu.nl
tal.linzen@jhu.edu
gmarkop@lit.auth.gr
alexander.martin@ed.ac.uk
a.nevins@ucl.ac.uk
Sharon.Peperkamp@ens.fr
polgardi@nytud.hu
topintzi@enl.auth.gr
Ruben.Vijver@uni-duesseldorf.de