Learning consonant harmony in artificial languages: Locality

Kevin McMullin
UBC Department of Linguistics

This study uses an artificial language learning experiment to investigate consonant harmony from a learnability perspective. With respect to locality, and the distance between two interacting consonants, the typology of languages with consonant harmony includes two types of systems. First are those that apply harmony in local cases only (i.e. separated by at most one vowel). Second, there are languages in which harmony applies anywhere within some domain, irrespective of distance. Monolingual English speakers are able to learn both types of patterns in the lab, and furthermore, they generalize from limited exposure in ways that mirror the typology. This is taken as evidence that humans have certain learning biases that not only affect how they learn and generalize patterns, but that these biases, over time, can shape the typology of phonological patterns.

1 Introduction

1.1 Phonological patterns and learnability

A great deal of research in linguistics concentrates on identifying, describing, and analyzing phonological patterns. However, we still do not know how humans learn them. This paper focuses on the learning of one such pattern, consonant harmony, in which two consonants in a word are required to agree in some way. For example, Yaka, a language from central Africa, has a perfective suffix -ili, which attaches to a verb. However, if the verb contains a nasal consonant, then the [l] in the suffix must become a nasal [n]. Thus, the verb stem jan-a ‘cry out in pain’ is jan-ini in the perfective rather than *jan-ili (Hansson 2010; Hyman 1995). More than 130 languages are known to have some form of consonant harmony system, each with its own set of properties. Some of these properties are more common than others, but we do not know why. One possible explanation is that it is an issue of learnability; the rare patterns are simply harder to learn, so they are less likely to ever arise in a language, let alone persist over time. Current research in linguistics and cognitive psychology has lent support to this idea by showing that some patterns involving the interaction of non-adjacent sounds are indeed more difficult to learn than others (Creel et al. 2004; Newport and Aslin 2004), particularly with respect to the relative similarity between the sounds (Gebhart et al. 2009; Moreton 2012). This generalization is mirrored in the typology of consonant harmony, as two similar consonants like [s] and [ʃ], or [l] and [r] are much more likely to interact than two dissimilar consonants such as [m] and [k], or [s] and [b]. The present study goes beyond the issue of similarity and investigates the learnability of consonant harmony with respect to locality, hypothesizing that humans will learn these patterns and generalize them in a way that matches their distribution across the world’s languages.

1.2 The typology of locality in consonant harmony

This section provides a brief description of the typology of sibilant harmony patterns (see Hansson 2001, 2010a; Rose and Walker 2004 for comprehensive studies of the typology of consonant harmony), focusing especially on generalizations that can be made about locality (i.e. the relative distance between two interacting consonants). Languages that have a consonant harmony system require two interacting consonants to agree with respect to some feature. The relevant feature varies by language and can range from voicing, to nasality, to palatalization, and so on. The most common type of consonant

---

1 The research reported here was supported by Standard Research Grant 410-2009-2831 from the Social Sciences and Humanities Research Council of Canada, awarded to Gunnar Ólafur Hansson.
harmony, which is the one used in this study, is sibilant harmony, which affects sibilant consonants such as \{s,z,ts,dz,tʃ,dʒ\}. Languages with sibilant harmony often prohibit the co-occurrence of a sibilant from each of the following sets \{s,z,ts,dz\} and \{tʃ,dʒ\}. This can be a purely phonotactic constraint, such that no word would ever have a possibility of violating it, or it can be seen in morphophonological alternations.

When considering the issue of locality, it seems plausible that languages would be more likely to apply consonant harmony to two consonants that are relatively close together in a word, because the pattern should be easier for the learner to detect. With this line of thinking, consonant harmony patterns that apply over longer distances would become increasingly more rare. Interestingly, however, this is only partially the case. Languages apply harmony either to local consonant pairs only (separated by at most one vowel) or to all cases within a word, both local and non-local, no matter the distance between them. This is an important typological split from a learnability perspective, as the range of possible consonant harmony patterns seems to be restricted by something other than purely computational limitations. One could easily imagine a language that has a dependency that holds between two consonants when they are separated by two vowels, but not by three vowels. No such system is known to exist.

Any theory of phonology that we have should account for why languages make this distinction between local and non-local dependencies, but nothing in between, as well as why no language has a non-local dependency that does not hold at local distances. In section 4, I will argue that this typological asymmetry is the result of a learning bias that allows humans to discover only these two types of patterns. A summary and evaluation of a formal learning algorithm that can potentially account for this (a precedence learner from Heinz 2010) will also be discussed.

1.3 Artificial language learning experiments

The traditional way to study any linguistic issues related to learning is to observe infants as they acquire the patterns in their native language. However, as many of these patterns are relatively rare, access to children learning the language is not always feasible. As a result, it is becoming increasingly common for researchers to construct artificial languages for experimental studies (Finley and Badecker 2009; Moreton 2012; Moreton and Pater 2011; Nevins 2010). In a typical experiment, subjects complete a training phase in which they are exposed words from the language, followed by a testing phase. Both of these phases can take many different forms. In the testing phase, for example, some experiments have a 2 alternative forced choice task in which the subject must decide which of two words is “correct” or “sounds better”. Alternatively, subjects may be asked to produce certain words, and their responses as well as the errors they make can be taken as evidence for what they have or have not learned (e.g. Rose and King 2007).

This methodology creates an accessible way to study any type of phonological pattern, whether it is relatively frequent, relatively rare, or completely unattested across the world’s languages. An additional benefit of this method is that it allows the researcher to have control over the input that the learner gets. These are the obvious advantages for those interested in questions about how humans learn, but the methodology is not without criticisms (see Moreton and Pater 2011 for an overview of the findings and criticisms of these experiments). For example, we may not know what biases the learners are bringing in from their own native language, or their language experience as a whole. Furthermore, we still do not know about the relationship between how children and adults learn the patterns in a new language. These are ongoing, unresolved debates that deserve a full investigation of their own, but in the meantime, having a well-constructed control condition can help us to understand what biases a learner might come in with, so that we can build them into any statistical models.

1.4 Previous studies

Of particular interest here are studies that have investigated the learning of consonant harmony at different levels of locality. Finley (2011) examined how adults learn local, transvocalic dependencies in an artificial language with sibilant harmony, as compared with nonlocal dependencies (across two vowels). In the training phase of experiment 1, subjects heard pairs of words. The first word was a two syllable stem of the form evSv, where S is a sibilant [s] or [ʃ]. The second was the same stem with a
suffix -Su, where the sibilant in the suffix was identical to the sibilant in the stem. Thus, all suffixed training items showed evidence of “first-order” harmony. In the testing phase, subjects chose which of two suffixed forms was correct, one of which had matching sibilants (harmony) while the other had mismatched sibilants. Experiment 2 was similar, but learners were trained only on a “second-order” pattern, in which the stems were of the form Svcv. The results indicated an asymmetry that reflects the typology described above. That is, subjects in experiment 1 learned the first-order pattern for both familiar and unfamiliar first-order test items, and did not generalize this to the second-order contexts, in which the sibilants were further apart. Subjects in experiment 2 were successful in learning the second-order pattern and they generalized the pattern inward to the first-order context. The author used these results as evidence for first-order patterns having a “privileged status”.

Using methods similar to those of Finley (2011), Finley (2012) expanded her study by presenting the subjects with longer stems to further investigate the role of distance between two interacting consonants. Subjects were trained on pairs of words consisting of a three syllable stem with one sibilant followed by the suffixed form, in which the sibilant in the suffix -Su displayed harmony with the stem. This study makes two important conclusions. First, that learners are able to learn this type of sibilant harmony pattern even when the sibilants are up to five segments away (i.e. Svcvcv-Su); the further away the dependency, the harder it is to learn. Second, when learners are trained on a “second-order” pattern, with cvSvcv-Su words, they generalize to even longer distances: Svcvcv-Su. Based on these results, Finley argues that when learners are tasked with discovering a long-distance interaction, they do so without reference to intervening distance, and so they apply the pattern to all contexts. Crucially, this generalization excludes the local-only patterns that apply harmony when the distance is at most transvocalic.

1.5 The present study

This paper presents the results of another artificial language learning experiment that replicates the findings of Finley (2011; 2012), while some key methodological differences, described in section 2.5, allow for a further investigation of unresolved issues.

2 Methodology

2.1 Subjects

Thirty-six adults aged 18-40 participated in the study (25 female, 8 male, 3 unspecified; mean age of 24). All were native speakers of North American English, reported no speech or hearing disorders and had no experience with a language containing a harmony system. Subjects were compensated $10 for participating in the experiment, which took about 45 minutes to complete.

2.2 Stimuli

Stimuli were recorded by a phonetically trained, male native English speaker. While he knew that the stimuli would be used for an artificial language experiment, he was unaware of the exact topic of study and the hypothesis. Most of the stimuli consisted of three-syllable “verbs” which took the form cvcvvcv. There were also six two-syllable verbs of the form cvcv for the practice phase, as described below. Additionally, the same speaker recorded two corresponding suffixed verbs for each root, one of each -su and -ʃi. Consonants in the stems were chosen from sibilants [s, f] as well as stop consonants [p, t, k, b, d, g] and vowels were chosen from [i, e, a, o, u]. The stimuli set was carefully constructed such that there was no other predictable pattern in the data. Crucially, the same number of each vowel and consonant appeared in each position for the verb roots containing [s] or [ʃ].
2.3 Design and procedure

Subjects were told that their task was to practice pronouncing words from a new language and to learn how to say verbs from the language in the past and future tense. Each subject worked through three phases of the experiment.

In the practice phase, subjects were told how to conjugate verbs in the language. They then heard, over a set of headphones, six pairs of words for the past tense (a verb root followed by a suffixed form with -su), and six word pairs for the future tense (this time with the suffix -fi). The six verb roots in this portion were just two syllables and contained no sibilants. As a result, there was no input with any evidence of an alternation in the practice phase.

In the training phase, subjects heard triplets of words, the first of which was always a three-syllable verb root. Half of these contained one sibilant (one quarter with [s], one quarter with [ʃ]) and the other half contained only stop consonants. This was done in part to test how learners perform without 100% of the input containing meaningful information, and in part to allow the other half of the input to be manipulated in future studies. The two subsequent words in each triplet were the four-syllable suffixed forms. Since both suffixes begin with a sibilant, if the verb root also contained a sibilant, one of the suffixed forms would display consonant harmony triggered by the suffix. For example, given the verb bugaso, the suffixed forms would be bugaso-su and bugaʃo-ʃi. Each triplet was presented twice, with the suffixed forms counterbalanced for order and the order of triplets was randomized for each subject. The subjects were asked to repeat each word into a head-mounted microphone that recorded their productions. This was done to allow for possible further analysis, as well as to reinforce the training process. In total, each subject heard and repeated 120 triplets twice each for a total of 720 productions. The words in both the practice phase and the training phase were presented over the headphones only, and the subjects got no information about any semantic content of the words, except that they were verbs that can be conjugated in these two tenses. Subjects were assigned to one of three training conditions, described below in section 2.4, that differed in the types of words presented in their training phase.

In the testing phase, subjects heard 30 new verbs, and then completed a forced choice task, in which they were asked to choose the correct conjugated form of the verb. Testing items included 30 suffixed harmonic/non-harmonic verb pairs of the form cvcvcv-Sv. These pairs consisted of ten each with the sibilant in the three different consonantal positions in the root. There were an equal number of -su and -fi forms, and the order of presentation was randomized for each subject. It is important to note that all testing items required the subject to choose an alternation in the root in order to identify the suffixed form with harmony (i.e. if the testing root contained a [s], then the corresponding suffixed forms would both have the suffix -fi). It would be ideal to have many filler items in the testing phase, as well as harmonic responses that do not require an alternation. However, a confound of that approach is that any items presented in the testing phase effectively become a part of the training, so it is necessary to both limit the testing items and maximize the amount of relevant data from each subject. As a result, subjects were tested only on cases that required the choice of an unfaithful root alternation in favour of harmony with the sibilant in the suffix. Subjects were given 3 seconds after the onset of the final word in the triplet to register a response, and were presented with their response time.

2.4 Training conditions

The subjects were divided into three conditions: Local, Nonlocal, and Control. Stimuli for these three groups differed only in the training phase, with the testing items being the same for all groups. For the portion of the training items that contained sibilants, the Local group was trained on verb roots that contained sibilants only in the last consonant position (i.e. cvcvs), while the Nonlocal group was trained on cvScvc roots. The Control group was trained only on verbs that contained no sibilants. This was done to reveal any potential pre-existing biases that English speakers may have toward choosing harmony, or any biases introduced by the experimental procedure.
2.5 Some important differences

The experimental design outlined above contains some unique methodological features that will be pointed out in this section. First, in the conditions where subjects are expected to learn a consonant harmony pattern (the local and nonlocal conditions) only half of the triplets in the training phase contain any evidence of an alternation. The other half can be thought of as filler items, which give no information about sibilant harmony to the learners. This could not, for example, have been done in Finley’s studies (described in section 1.4), since having roots with no sibilants would require the choice of a default suffix. The default option would appear more frequently in training than the alternative, and potentially bias the subjects towards the default suffix in the testing phase. The design of the present experiment avoids this problem by having two completely different suffixes (one for past tense and one for future tense), each with a different sibilant, any root that did not contain a sibilant would still have two suffixed forms, one with each sibilant.

Second, the control condition for this study is quite different than in most artificial language learning experiments. If the study is of a purely phonotactic nature, then subjects in the control condition often skip the training phase completely and proceed directly to the testing phase to determine whether they have a preference for or against words adhering to the phonotactic pattern of interest. For studies that present a morphophonological alternation, subjects in the control condition are usually exposed only to the roots, and then are tested on whether they have a preference for one affixed form or another. One confound of these approaches is that the subjects in the control condition have not actually completed the same procedure as those in other conditions. In the present study, all subjects, whether in the control condition or not, are first trained on 720 items before proceeding to the testing phase. This was done in light of the fact that hearing and repeating hundreds of nonsense words can often be quite boring. Having a control condition that has gone through the same procedure allows for a more fair comparison across groups. An additional advantage to this design is that if the subjects in the control condition come in with no biases towards sibilant harmony (and no propensity to make mistakes), then they are expected to never pick a suffixed form that is unfaithful to the root. Rather, the training phase should have taught them simply to add a suffix, and otherwise remain faithful. This is a departure from many other artificial language learning experiments where subjects in the control condition are expected to perform near chance, instead of near 0%.

3 Results and analysis

As reported in section 2.1, thirty-six subjects participated in this study. However, the data from six subjects was dropped (4 Control, 1 Local, 1 Nonlocal) as the result of one of the following: failure to follow instructions (2 subjects), choosing either the first or second test item on all recorded trials (2 subjects), equipment failure (2 subjects). Of the remaining 30 subjects, 827 of a possible 900 responses were registered within the allotted three-second timeframe.

3.1 An overview of responses

The first step in analyzing the results is to get a clear picture of what subjects’ choices were in each of the training conditions and for each type of testing item. To do this, I will present the results in this section as the proportion of testing items in which subjects chose to apply consonant harmony rather than stay faithful to the root. Keeping in mind that subjects in the control condition saw no evidence of an alternation in their input, their scores should be close to 0% if they are in no way biased towards harmony. Figure 1 below presents the proportion of responses where the subject chose a non-faithful alternation in the root in order to have harmony with the sibilant in the suffix. The results here are intended to give an overall picture of what the subjects in each condition chose, but will not be used to test for statistical significance.
The above results can be summarized as follows. Subjects in the control group are most likely to choose an alternation that results in harmony for sibilants in the second syllable of the root, and are least likely to choose an alternation for word-initial sibilants. Compared to the control group, subjects who were trained locally are more likely to choose an alternation in all positions, though this is most evident in the local cases, followed by the cases with a sibilant in the second syllable, with minimal distinction in the word initial position (at distance 3). Subjects in the nonlocal training condition are also more likely to choose an alternation at all distances compared to the control group. Compared to the local group, they are less likely to do so at distance 1, but more likely at distances 2 and 3.

The above results are interesting, but it is not yet clear which of these differences is meaningful. For example, if we determine that those trained in the local condition have indeed learned the local pattern, can we conclude whether or not they have generalized to the nonlocal items at distance 2? Similarly, those in the nonlocal condition appear to have learned the pattern at the distance they were trained on (distance 2), and generalized to the local contexts, but do we have evidence that they have generalized to the word initial positions at distance 3?

### 3.2 A mixed-effects logistic regression model

An alternative way to analyze data sets comprised of categorical responses, rather than using the average proportions for each subject, is with a logistic regression model (see Jaeger 2008 for a description of these models intended for linguists). In a logistic regression, the model finds the best fit for the log odds of choosing one response or another – in this case harmony vs. no harmony – based on a set of predictor variables. The predictor variables here, which are also categorical values, are the training conditions (Control, Local, Nonlocal), as well as the distance between sibilants for each testing item (1, 2,
or 3). Note that the hypothesis predicts different results for each group at each distance, so it is important to include interactions between group and distance in the model as well. Other variables (including trial number, whether the test item with harmony was presented first or second, and whether the triggering suffix was -ji or -su) were included in other models, but showed no significant effects and the models including these variables did not give a significantly better fit to the data, so they are not included below in Table 1. Additionally, with the use of a mixed-effects logistic regression, the model can account for tendencies for individual subjects. Barr et al. (in press) argues for a maximally rich random-effects structure in regression, so long as it is justified by the design. In the model presented below, a random intercept as well as a random slope for each distance was included for each subject. The model below was created with the glmer function from the lme4 package (Bates et al. 2012) implemented in R (R Core Team 2012).

**Table 1**

A summary of the fixed portion of mixed-effects logistic regression model

| Coefficient           | Estimate | SE  | Pr(>|z|) |
|-----------------------|----------|-----|---------|
| Intercept             | -1.623   | 0.363 | <0.001* |
| Distance 2            | 0.703    | 0.393 | 0.074   |
| Distance 3            | -1.171   | 0.536 | 0.029*  |
| Local Group           | 1.802    | 0.481 | <0.001* |
| Nonlocal Group        | 1.359    | 0.483 | 0.005*  |
| Distance 2 x Local Group | -1.444 | 0.517 | 0.005*  |
| Distance 3 x Local Group | -1.110 | 0.692 | 0.109   |
| Distance 2 x Nonlocal Group | -0.6336 | 0.519 | 0.222   |
| Distance 3 x Nonlocal Group | 0.5212 | 0.658 | 0.428   |

In the table above, the estimate for the Intercept indicates the model’s guess for the likelihood (in terms of log odds) that an average subject in the control group will choose harmony when presented with a distance 1 testing item. The negative estimate of -1.623 indicates that the model predicts the odds of choosing harmony to be \( e^{-1.623} = 0.197 \), which, in terms of probability, translates to a 16.5% chance. The estimates for the subsequent main effects of predictor variables indicate whether there is an increase or decrease in the likelihood of choosing harmony, as compared to this baseline.

With respect to the main effects, the estimate for distance 2 is positive, indicating an increase in the likelihood of control subjects choosing harmony, while the estimate for distance 3 is negative, indicating a decrease. Note that the effect of distance 3 reaches a significance level of <0.05. For distance 2, it is approaching significance with \( p=0.074 \). For the main effects of group, both the Local and Nonlocal groups have positive estimates that are significant, and so they are much more likely to choose harmony at distance 1. This indicates that they have learned a consonant harmony pattern, or at least have begun to learn it. For the local group, this is the pattern they were trained on, but for the Nonlocal group, this effect demonstrates that they are choosing harmony at distance 1, even though they were trained at distance 2.

For the interactions of distance and group, the estimates indicate whether the likelihood of choosing harmony increases or decreases after already taking into account the main effects of group and distance, as well as the intercept. For the Local group, there is a significant decrease in the likelihood of choosing harmony at distance 2 and a decrease that is approaching significance at distance 3. This

---

2 The results for individual subjects are reported in Appendix 1. They are quite interesting, as it seems that some subjects learn the pattern quite well, while others perform at a lower level than the control group. The reasons for and consequences of having different types of human learners have not been discussed in the literature, but this topic warrants further research.

3 One problem with logistic regression models is that it can be difficult to assess significance in models that include interactions of the predictor variables. For this reason, the reader should not be concerned that some effects do not quite reach the arbitrary significance level, and instead focus on the effect size and whether it is positive or negative.
indicates that they have not generalized the local pattern that they learned to either of the nonlocal distances. In contrast, the interactions of the Nonlocal group with both distances 2 and 3 did not result in a significant increase or decrease in the likelihood of choosing harmony indicating that the pattern they have learned applies to all distances.

These estimates can be difficult to interpret because of their complexity. For interactions, the estimates do not represent a final probability, but a number that needs to first be combined with several. However, it is possible to calculate the final probabilities for each group at each distance. As an example, the overall odds of an average subject in the local group selecting harmony at distance 2 can be calculated as \( e^{1.623+0.793+1.802-1.444} = 0.570 \), which translates to a 36.3% chance. The fitted probabilities for each group at each distance are presented in Table 2 below. They correspond roughly to the average proportions of harmony responses that were presented above in Figure 1, but are actually the resulting probabilities for the best fit from the logistic regression model after building in the random effects.

### Table 2
Fitted probabilities of choosing harmony, based on the mixed-effects logistic regression model from Table 1

<table>
<thead>
<tr>
<th>Group</th>
<th>Distance</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td></td>
<td>16.4%</td>
<td>28.5%</td>
<td>5.8%</td>
</tr>
<tr>
<td>Local</td>
<td></td>
<td>54.4%</td>
<td>36.3%</td>
<td>10.9%</td>
</tr>
<tr>
<td>Nonlocal</td>
<td></td>
<td>43.4%</td>
<td>45.1%</td>
<td>28.6%</td>
</tr>
</tbody>
</table>

### 4 Discussion

#### 4.1 Comparing results to typology

The results presented above provide evidence in support of the hypothesis that humans will learn new phonological patterns and generalize them in a way that matches up to the typology of the world’s languages. As described in section 1, the typology of consonant harmony reveals just two types of languages with respect to locality. The first type consists of the languages that apply harmony only when the two dependent consonants are separated by at most one vowel. The second type is the set of languages that apply harmony whenever the relevant consonants co-occur within some domain, as defined without reference to any phonological representation, including transvocalic contexts as well as longer distances. In the results of this experiment, we see that humans learning a new artificial language can learn both types of languages. However, it is how the learners tend to generalize the pattern that makes this result most interesting.

Consider the learners in the local group. Their relevant training items were restricted to instances of harmony applying in transvocalic contexts. Not only were they more likely than the control group to choose a harmonic response in the same types of items that they were trained on, but they did not (or at least were less likely to) generalize this to nonlocal distances. Once the learners in the local group discover the pattern, they do not extend it to nonlocal distances, even though such a pattern is attested in natural languages, and they saw no evidence to the contrary. That is, the subjects seem to be making the most direct application of their training as is possible – harmony occurs in local contexts, and nowhere else.

We now turn to the nonlocal group, whose relevant training items were restricted to instances of nonlocal harmony applying over two syllables (cvScv-Sv). Subjects in this group were more likely than the control group to apply harmony in the same types of words that they were trained on. Furthermore, they generalized this pattern to types of words that were not encountered in their training data – to roots that had a sibilant in the final syllable (a shorter distance 1), and roots that had a sibilant in the first syllable (a longer distance 3). So in this case, subjects discovered a pattern that applied harmony at a nonlocal distance. Additionally, they correctly rule out the possibility of it being the first type of language that applies harmony only at local distances, and so they settle on generalizing it to all contexts within the
word, in line with the pattern we see in the second, and only other type of attested consonant harmony language.

4.2 A learning bias

An increasingly common explanation for why some phonological patterns are more common than others is that the more frequent patterns are more likely to survive over time because of cognitive learning biases. This type of bias, which facilitates the learning of some patterns, prevents the learning of others, and helps determine how a pattern is generalized, is labeled as an analytic bias. Over time, a propensity of humans to learn a certain pattern would result in the pattern having a high survival rate (Moreton 2008). This is different from a more traditional approach to phonological change that is based on phonetic naturalness, or channel bias – the tendency for some sounds to be either misproduced or misperceived in certain contexts. The result discussed above suggests that humans have a learning bias that affects how humans learn the properties of locality for long-distance interactions. The subjects, who had no experience with a language containing harmony, were able to learn the pattern and they generalized in a way that matched the typology. This is taken as evidence that it is a learning bias that has resulted in only two types of consonant harmony patterns in the world’s languages.

4.3 Formal models and other possible outcomes

The results of this experiment support the hypothesis that humans will generalize a phonological pattern in a way that mirrors the typology of that pattern. However, note that the subjects in this experiment spoke no languages other than English, and presumably know nothing about the typology of consonant harmony systems. This section then, will ignore the typological facts and consider the possible outcomes from a learner that has no preconceived notion of how consonant harmony patterns should apply.

First, consider a learner that is trained only on items that exhibit local harmony, and some viable strategies that the learner might use to discover a consonant harmony pattern. One possible type of learner is one that relies on brute force, looking for a dependency between any two consonants in a word, no matter how far apart they are and without keeping track of distances. This learner, though trained only on harmony in local contexts, will learn only that there is a dependency between [s] and [ʃ]. When presented with testing items, it will incorrectly apply this pattern to both local and nonlocal sibilant pairs, since it does not record the distance. In order to prohibit it from overextending the pattern, there are at least two possible modifications that can be made. First is to allow the learner to also keep track of the distance between the two consonants. In the training phase, the learner will discover that a dependency exists between sibilants that are separated by a vowel. In the testing phase, it will look for any such instances and apply harmony. A second, similar option is to have the learner restrict its maximum search space to the largest distance between sibilants seen in the training data. It will have completed the training phase having never seen any instances of two sibilants that are separated by more than one vowel. During the testing phase, it will not even look for sibilants in a nonlocal relationship, and so it will apply harmony in the local cases only. Either of the latter two learners would give a result analogous to what we observed in the experiment for the local group.

With this in mind, however, a problem arises when considering the task of learning the pattern with input restricted to harmony at distance 2. Only the first of the three learners presented above, which incorrectly generalized a local pattern to nonlocal contexts, is capable of learning and generalizing in the same way as the nonlocal group. In this case, the modifications that solved this problem for the case of local harmony would result in it learning to apply harmony either at distance 2 only, or to apply it at any distance up to 2. This paradox emphasizes the fact that even though humans learners tend to generalize consonant harmony patterns in a way that directly relates to the typology, finding one learning algorithm that can simultaneously learn both types of languages is not trivial.
4.4 The precedence model of learning long-distance dependencies

In an attempt to explain how humans might learn long-distance dependencies, including consonant harmony, Heinz (2010) proposes a precedence model of learning long-distance phonotactics. In principle, it is a very simple model, and in practice it makes strong predictions about what the typology of consonant harmony should look like if language learners use the proposed strategy when learning the pattern.

In the precedence model of learning, a learner takes into account *precedence relationships* in addition to keeping track of bigrams. For example, in a word like “pants” /pænts/, the learner keeps track of the bigrams \{p,æ\}, \{æ,n\}, \{n,t\}, and \{t,s\}, which are all adjacent phonemes. Additionally, the learner tracks the precedence relationships \{p,æ\}, \{p,n\}, \{p,t\}, \{p,s\}, \{æ,n\}, \{æ,t\}, \{æ,s\}, \{n,t\}, \{n,s\} and \{t,s\} without reference to distance, just that the first sound precedes the second somewhere in the word. Therefore, such a learner makes a distinction between co-occurrence restrictions between adjacent segments, and dependencies that hold between any two segments. For a case of nonlocal sibilant harmony (that is purely phonotactic), the precedence relations \{s,f\} and \{f,s\} would never be encountered in the language, so the learner would recognize that two different sibilants should no co-occur within a word. Note that this model does not keep track of any segments that occur in between the segments of each precedence relationship, and thus no intervening segment can have any influence on the nature of the precedence relationship. Such a property makes the strong prediction that there is no possibility of an intervening segment blocking the interaction between non-adjacent phonemes, or at least that such a property would never be learned. Heinz argues that this property is desirable since there are no discovered instances of blocking in the typology of consonant harmony. However, several languages, including some Berber dialects (Elmedlaoui 1995; Hansson 2010b) and Rwanda (Walker and Mpiranya 2005), may exhibit the blocking of consonant harmony, so a reevaluation of this aspect of the precedence learning model is in order.

There is another prediction that the precedence learning model makes by adding the notion of a consonant tier. By doing so, the learner can also keep track of bigram relationships between consonants alone, therefore making a distinction between consonants that are adjacent on the consonant tier, and consonants that are in precedence relations, but ignoring any further distinction regarding distance. This would give rise to the typological split mentioned above between languages that have only local dependencies, and languages in which the dependency holds at all distances. One downside to this model is that it does not account for the effects of similarity. The learner is equally capable of picking out dependencies among sounds that are relatively dissimilar, which is reflected neither in the typology of consonant harmony, nor in the experimental research reported in the literature. However, this algorithm is not incompatible with other models of learning that are biased towards picking out similar segments (e.g. Hayes and Wilson 2008), and such a combination could provide a more comprehensive model of how humans are learning phonological patterns.

4.5 Directions for future research

While this study has presented evidence that humans learn and generalize patterns in a way that reflects the patterns that occur in natural languages, we do not know the limits on human learning for unattested patterns. For example, is it possible to learn a distinction between sibilant harmony that occurs between sibilants across one vowel or two vowels, but not at further distances? Is it possible to learn a non-local sibilant harmony pattern that does not apply at local distances as well? Answers to these types of questions will help us to better understand the restrictions on human learning and whether there is a direct relationship with the typological distribution of phonological patterns.

Another question that merits investigation is whether evidence of a local dependency in the input would actually help a learner to discover the dependency at a greater distance. That is, would the learner first discover the (more typologically common) local dependency and then check for the dependency at the longer distances? This line of research would give us insight into the strategies that learners use when they are trying to discover the phonological patterns of a language.
5 Summary and conclusions

The goal of this study was to give an explanation for why the typology of consonant harmony, with respect to locality, includes only two types of languages – those with local harmony, and those with harmony that applies across all distances. A logistic regression mixed model showed that learners trained on sibilant harmony only at a local (transvocalic) distance learned the pattern, but did not generalize the pattern to sibilants at nonlocal distances (two or three syllables away from the triggering suffix). Learners trained on sibilant harmony only at the nonlocal distance (two syllables away) learned this pattern, and generalized not only to the shorter, local distance, but also to the word initial sibilants that were three syllables away from the triggering suffix. This experiment, with several differences in methodology, has replicated findings from Finley (2011, 2012). The results are taken as evidence that humans have a learning bias that affects how we learn and generalize patterns, and that this learning bias is responsible for the typological shape of consonant harmony systems.

References


Kevin McMullin  
kevinmcm@alumni.ubc.ca
Appendix 1 – Individual subject responses

Control

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion harmony</td>
<td>0.0</td>
<td>0.2</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Observed probabilities | Fitted probabilities

Local

<table>
<thead>
<tr>
<th></th>
<th>107</th>
<th>108</th>
<th>109</th>
<th>110</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion harmony</td>
<td>0.0</td>
<td>0.2</td>
<td>0.4</td>
<td>0.6</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Observed probabilities | Fitted probabilities
Nonlocal Distance
Proportion harmony
0.0
0.2
0.4
0.6
0.8
1.0

Observed probabilities
Fitted probabilities

Distance

Proportion harmony