The role of long-distance phonological processes in spoken word recognition: a preliminary investigation

Phillip Burness, Kevin McMullin, and Tania Zamuner
University of Ottawa

Previous work has demonstrated that during spoken word recognition, listeners can use a variety of cues to anticipate an upcoming sound before the sound is encountered. However, this vein of research has largely focused on local phenomena that hold between adjacent sounds. In order to fill this gap, we combine the Visual World Paradigm with an Artificial Language Learning methodology to investigate whether knowledge of a long-distance pattern of sibilant harmony can be utilized during spoken word recognition. The hypothesis was that participants trained on sibilant harmony could more quickly identify a target word from among a set of competitors when that target contained a prefix which had undergone regressive sibilant harmony. Participants tended to behave as expected for the subset of items that they saw during training, but the effect did not reach statistical significance and did not extend to novel items. This suggests that participants did not learn the rule of sibilant harmony and may have been memorizing which base went with which alternant. Failure to learn the pattern may have been due to certain aspects of the design, which will be addressed in future iterations of the experiment.

1 Introduction

The overwhelming majority of phonological patterns in human language are local in the sense that they make reference to segments that are adjacent in a string. For example, the plural morpheme /-z/ in English is realized as the voiceless [-s] when it attaches to a noun that ends in a voiceless consonant (as in ‘cat ~ cats’). In this case, we have a local restriction against two adjacent segments that disagree in voicing. Not all phonological patterns, however, obey this criterion of locality; some languages contain long-distance patterns in which two segments interact in some way, but can be separated by an arbitrary number of other segments. One long-distance pattern is sibilant harmony (Rose and Walker 2004), and while not exceedingly common, it is attested in a variety of unrelated languages such as in Navajo (Na-Dene: Berkson 2013), in Aari (Omotic: Hayward 1990), and in Slovenian (Indo-European: Jurgec 2011). In a typical sibilant harmony system, all sibilants within a given word must agree in anteriority, meaning that a word-internal sequence of a [-ant] sibilant followed by a [+ant] sibilant (e.g., [f...s]), or vice versa, must be repaired so that both sibilants are [-ant] or both sibilants are [+ant]. Importantly, this holds no matter how many sounds intervene between the two sibilants. Take for instance the data in (1) from Navajo (Berkson 2013). The first-person possessive prefix is underlyingly /ʃi-/ and surfaces as such when it attaches to a word containing no sibilants (1a) or containing /ʃ/ (1b). When the prefix attaches to a word containing /s/, however, the prefix instead surfaces as [si-] like in (1c).1 In this example, the noun-internal /s/ is causing the prefix’s /ʃ/ to alternate, despite the two sibilants being separated by three other sounds.

1 While this process was historically productive, recent generations seem to prefer the [ʃi] allomorph in all cases (Berkson 2013). For possible reasons behind the disappearance of sibilant harmony in Navajo, see Berkson (2013).
Navajo sibilant harmony (Berkson 2013)

a. /ʃi-bid/ $\rightarrow$ [ʃi-bid] ‘my stomach’
b. /ʃi-giʃ/ $\rightarrow$ [ʃi-giʃ] ‘my cane’
c. /ʃi-bêː.so/ $\rightarrow$ [ʃi-bêː.so] ‘my money’

The long-distance nature of sibilant harmony and other types of consonant harmony is of particular interest when we consider its potential to facilitate language processing. A growing body of research has demonstrated that during spoken word recognition, listeners can use a variety of cues to anticipate an upcoming sound before it is realized (Dahan et al. 2001; Salverda et al. 2003, 2014; Gow and McMurray 2007; Beddor et al. 2013; Mahr et al. 2015; Blazej and Cohen-Goldberg 2015; Paquette-Smith et al. 2016; Zamuner et al. 2016, Desmeules-Trudel and Zamuner 2019). This literature, however, has focused on local dependencies between adjacent segments, as opposed to long-distance phenomena. Consider again the Navajo data above. One can imagine that a Navajo listener, upon hearing the [si-] alternant of the /ʃi-/ prefix, would expect a noun containing [s] to follow. Importantly, this would be a prediction based on a long-distance interaction (in the above example, across three other sounds). Navajo sibilant harmony is therefore an ideal candidate for extending the existing literature by testing whether long-distance phenomena can also cause listeners to anticipate future speech sounds during spoken word recognition.

It is important to investigate long-distance phenomena for several reasons. First, increased distance between interacting segments is generally dispreferred in harmonic and dissimilatory processes cross-linguistically (Krämer 2003; Martin 2005; Hayes and Londe 2006; Hayes et al. 2009; Finley 2011, 2015, 2017; Kimper 2011; Jurgec 2011; McMullin and Hansson 2014; Zymet 2015; McMullin 2016). Second, experimental work has shown that distance can interfere with certain aspects of perception (Kimper 2017; Zellou and Pycha 2018). We can therefore ask whether predictions made during spoken-word recognition based on long-distance phonology will be weaker, or even completely absent. The answers to these questions, whatever they may be, will have consequences for theories of spoken-word recognition and may shed new light on the typology of long-distance phonological processes.

The remainder of this paper is structured as follows. Section 2 summarizes the major findings with regards to sound-based prediction in spoken word recognition and highlights the gap in the experimental literature that the current study is designed to fill. Section 3 presents the two experimental paradigms that will be used in the current study. Section 4 provides an in-depth description of the experimental methods. Section 5 presents and analyzes the experimental results. Finally, section 6 concludes and suggests ways in which the current experiment can be improved in future research.

2 Background

Listeners are capable of anticipating that the speech stream will contain a certain sound before they hear it. In some cases, this will allow them to identify a spoken word earlier than might otherwise be expected. One cue that listeners use to make such predictions is place assimilation. In English, underlying coronal nasal consonants often assimilate to the place feature of the following consonant (Gow and McMurray 2007). For example, the final [n] of ‘green’ in a word sequence like ‘green boat’ is likely to become [m] by assimilating to the labial place feature of the following [b]. Gow and McMurray (2007) conducted an eye-tracking study to investigate the effect that this process has on spoken word recognition. Participants heard word sequences like ‘green boat’ in which a word-final [n] either assimilated to the following consonant (e.g., [ɡiːm bʊt], or preserved its coronal articulation (e.g., [ɡiːn bʊt]). Results showed that participants fixated faster on the target image when the recording contained assimilated nasal consonants, suggesting that participants used their knowledge of this process to anticipate the identity of an upcoming consonant. Similar effects have been found for other cues as well, including vowel formant transitions (Dahan et al. 2001; Salverda et al. 2014; Mahr et al. 2015; Paquette-Smith et al. 2016), vowel nasalization (Beddor et al. 2013; Paquette-Smith et al. 2016; Zamuner et al. 2016, Desmeules-Trudel and Zamuner 2019), and segment duration (Salverda et al. 2003; Blazej and Cohen-Goldberg 2015).
The above-mentioned studies primarily investigated local phenomena that hold between adjacent elements or adjacent syllables, though there is at least one study that explored the predictive potential of a long-distance phonetic/phonological phenomenon. Carbary et al. (2015) investigated whether participants could make predictions based on anticipatory de-accentuation in English: a discourse-new noun may lack a pitch accent only if that same noun re-occurs later in the same sentence. Carbary et al. (2015) presented their participants with sentences like the pair in (2), where capitalization represents a pitch-accented word and where # represents a spliced in cough that masks the identity of a sound.

(2) Example sentence pair from Carbary et al. (2015)
   a. Regular: Drag the TRIANGLE with the HOUSE to the CIRCLE with the #OUSE.
   b. De-accented: Drag the TRIANGLE with the house to the CIRCLE with the #ouse.

The two target images were always the same shape (in example (2) this would be a circle), but superimposed on that shape were images of a minimal pair of nouns that share everything except their initial consonants (in example (2) the word pair would be ‘house’ and ‘mouse’). When the prepositional-phrase adjunct of the initial noun phrase was de-accented as in (2b), participants were more likely to select the same-attribute response (circle with the house) than the different-attribute response (circle with the mouse). Participants were also faster to answer in the de-accented condition, since de-accentuation places restrictions on what material can follow, namely that the de-accented noun (e.g., ‘house’ in 2b) must re-occur later in the sentence. Interestingly, in 18.6% of all trials, participants responded within 200ms of the cough offset, suggesting that they made their decision prior to the cough. These results suggest that participants used their knowledge of the rule of anticipatory de-accentuation to predict the re-occurrence of a discourse-new but unaccented noun, before actually hearing the re-occurrence of the noun.

Outside the single study above, the absence of studies investigating long-distance phenomena leaves open the question of whether such phenomena can also facilitate spoken word recognition by allowing listeners to anticipate upcoming sounds without the aid of coarticulatory cues from an adjacent segment. This is an important question considering how increased distance between interacting segments is generally dis-preferred in harmonic and dissimilatory processes cross-linguistically. Artificial language learning experiments have shown that even though participants are capable of learning long-distance processes that apply across one or more syllables, there is a preference for learning local versions of the same patterns (Finley 2011; McMullin and Hansson 2014; McMullin 2016). For example, McMullin and Hansson (2014) found that participants exposed to sibilant harmony applying across one vowel did not generalize harmony outward to more distant contexts, whereas participants exposed to sibilant harmony applying across two vowels generalized harmony both inward to more local contexts and outward to more distant contexts. Corroborating the learning results are descriptions of distance-based decay, which is the tendency for harmonic and dissimilatory processes to decrease in rate of application as more and more transparent segments intervene (Krämer 2003; Martin 2005; Hayes and Londe 2006; Hayes et al. 2009; Jurgec 2011; Kimper 2011; Zymet 2015). For example, front unrounded vowels are transparent to front/back harmony in Hungarian, but if a back vowel is followed by more transparent vowels, it is more likely to fail to trigger harmony, or only trigger harmony optionally (Hayes and Londe 2006).

Another reason for asking whether long-distance phenomena can facilitate spoken word recognition is that previous work has shown that distance can interfere with certain aspects of perception (Kimper 2017; Zellou and Pycha 2018). Vowel harmony, for example, is hypothesized to be a means of enhancing hard-to-perceive phonetic features by increasing the temporal extent of the phonetic feature (Suomi 1983; Kaun 1995). Kimper (2017) conducted a series of experiments to test whether this perceptual benefit can hold across transparent vowels. In his experiments, participants were presented with nonsense words followed by an isolated vowel and were asked whether the isolated vowel occurred in the preceding word. Participants were consistently faster and more accurate when the nonsense words exhibited vowel harmony along a relevant feature dimension with the isolated vowel, but this effect was diminished when transparent vowels intervened. Zellou and Pycha (2018) tested Ohala’s (1993) suggestion that acoustic changes are more susceptible to misattribution the further they are from their source. In one experiment, participants
were presented with recordings of nonsense words of the shape [gɪC] and asked to estimate the height of the speaker. The reason for this is that differences in the formants of [ɪ] can have two causes: increased speaker height is correlated with lower formant values, and coda consonants cause the second formant of [ɪ] to slope downwards towards the end of the vowel. As the temporal extent of the coda-induced formant transitions increased, participants were more likely to judge the speaker as being taller, meaning that as the onset of F2 lowering became more and more distant from its source, participants were more and more likely to “over-hypothesize” and identify the F2 lowering as being due to increased speaker height in addition to being caused by the coda vowel, when in fact it was only due to the latter.

The cross-linguistic dispreference for long-distance phenomena and the potential for such patterns to interfere with certain aspects of perception could lead us to expect that long-distance phenomena may not factor into spoken word recognition, or else participate only weakly in this process. Results from an eye-tracking experiment by Tobin et al. (2010) seem to support this prediction. They took minimal pairs of monosyllabic words where one had a front vowel and the other had a back vowel (e.g., ‘pail/pole’ and ‘taste/toast’) and recorded them in the frame sentence “Pick up a X”. Due to anticipatory coarticulation, the vowels in the unstressed words ‘up’ and ‘a’ are slightly more front when the target word has a front vowel (e.g., “pick up a pail”) and are slightly more back prior to a back vowel (e.g., “pick up a pole”). In one condition, the pre-target portion of the sentence contained appropriate regressive vowel-to-vowel coarticulation (e.g., the target ‘pail’ was preceded by the words “pick up a” spliced out of a recording of “pick up a pail”). In the other condition, the pre-target portion of the sentence contained inappropriate regressive vowel-to-vowel coarticulation (e.g., the target “pail” was preceded by the words “pick up a” spliced out of a recording of “pick up a pole”). The proportion of target fixations did not deviate from the proportion of fixations to the competitor until about 200ms after the word’s onset, suggesting that the fronting/retracting of the pre-target unstressed vowels did not cause anticipation of the target word prior to its onset. When the coarticulation was inappropriate, however, the rate at which the proportion of fixations to the target grew was significantly slower. Tobin et al. (2010) interpreted these results to mean that the long-range fronting/retracting of the unstressed vowels is perceived, but either the processing system does not make use of this cue, or its impact is weak and therefore appears as a delayed effect.

Consider again the Navajo data from example (1), where the distance between trigger and target can be arbitrarily large. Listeners may not actively pay attention to sibilant harmony during spoken word recognition since its source is likely to be misinterpreted. Alternatively, or perhaps additionally, the processing benefit provided by sibilant harmony may decay over time and would therefore not be as useful as other cues, in which case sibilant harmony would be ignored in favour of allocating attentional resources to identifying more stable and helpful cues. Finally, in light of the results from Tobin et al. (2010), we could expect that any potential effects of sibilant harmony on spoken-word recognition may be delayed relative to the effect of other more local cues. Investigating sibilant harmony experimentally will therefore refine our understanding of the limits on the language processing system’s ability to anticipate information and may shed light on the nature of long-distance processes more generally.

### 3 Experimental framework

To investigate the processing of long-distance phenomena during spoken word recognition, the current study combines two widely-used paradigms into a single experiment. In order to single out the effects of sibilant harmony on language comprehension, the current study takes advantage of the high level of control afforded by the artificial language learning (ALL) paradigm. Additionally, since we are interested in observing the language comprehension process, which is both rapid and dynamic, we employ the visual world paradigm (VWP), which offers fine-grained time resolution. We present an overview of both paradigms below, before providing an in-depth breakdown of the current experiment’s methodology.

Quantitative research on sibilant harmony has largely been conducted using the ALL paradigm. In an ALL experiment, participants are first exposed to a constructed language in a learning phase and later tested to determine what they learned as well as whether/how they generalize what they learned. In her review of the literature on long-distance phonology, Finley (2017) discusses three major advantages of the ALL
paradigm. First, researchers have precise control over the languages in their experiments: they can manipulate the size and shape of the lexicon, the language’s phoneme inventory, the language’s morphology, and even the frequency of individual words or patterns within the language, thus allowing them to compare participants’ performance on minimally different languages in order to test precise hypotheses (Finley 2017). For example, Hudson Kam and Newport (2005, 2009) adjusted the number of exceptions to the rules of their language in order to test whether, when, and how learners regularize inconsistent input. The second major advantage of the ALL paradigm is that researchers can manipulate the details of how participants are exposed to the language, generally by simulating a poverty of the stimulus situation in which participants only see a subset of the data (Finley 2017). Take for instance Wilson (2006), who exposed one group of participants to instances of palatalization before high front vowels (/ki/ → [tʃi]) and another group to instances of palatalization before mid front vowels (/ke/ → [tʃe]). Despite both data sets being compatible with a language that palatalizes before both high and mid vowels, Wilson (2006) found that only the mid vowel group generalized palatalization to both high and mid vowel contexts. The final advantage of the ALL paradigm is that we can use it to test the psychological reality of typologically rare or unattested patterns (Finley 2017). For example, Moreton (2008) presented participants with three patterns: one in which vowels agreed in height, one in which consonants agreed in voicing, and one in which the voicing of a consonant depended on the height of the vowel next to it. The last of these is typologically suspect (i.e., Moreton (2008) found only three documented cases of this pattern, all of which were unproductive) and participants performed more poorly on it than on the former two patterns.

A typical ALL experiment measures only response accuracy and reaction times, and so the ALL paradigm on its own is not well-suited to giving us insight into how participants are processing the artificial language online. As a means of addressing this disadvantage, we follow Farris-Trimble and McMurray (2018) by augmenting the ALL paradigm to include methods from the visual world paradigm (VWP). The VWP has been extensively used in psycholinguistic research, and overviews of the major results obtained in this line of work are available in Huettig et al. (2011) and Kaplan et al. (2016). In a typical VWP experiment, participants are seated in front of a display showing a target image, and one or more competitor images that are related to the target in some pre-defined way (for example, they begin with the same consonant as the target). Participants are provided with an auditory stimulus, and their eye-movements are tracked relative to the screen as this stimulus unfolds. As Huettig et al. (2011) point out, a major advantage of the VWP is that, because the paradigm relies on the listeners’ tendency to look at relevant parts of the display as they are mentioned, the listeners do not have to perform any meta-linguistic judgments, which would potentially affect the way in which participants process the linguistic stimulus. The logic that links eye-tracking data to the cognitive function of language processing is that eye gaze reliably indicates a participant’s attention within a visual field (Liversedge and Findlay 2001), and that attention is often directed to objects that are relevant to the task at hand (Trueswell 2008). When studying lexical activation dynamics as will be done in the current study, the assumption is that a participant is more likely to look towards an image of an object as their mental representation of that object and its linguistic label receives more activation from the linguistic input (Alloppenna et al. 1998; Tanenhaus et al. 2000).

It should be noted that in the current experiment, the lexical entries that will assumedly be receiving activation from the auditory stimulus are not part of the participants’ native language, but rather belong to an artificial language that they will learn in the lab. Thusly, one could be concerned that the behavioural data collected in the current experiment will not reflect the true nature of language processing outside of the lab environment. We do not believe that this is necessarily an issue for two reasons. First, several studies have found that grammaticality in artificial languages evokes the same neurological reactions as grammaticality in natural languages (Friederici et al. 2002; Morgan-Short et al. 2010, 2012; Silva et al. 2016; Moore-Cantwell et al. 2018). Second, several previous VWP experiments have used artificial lexicons (Magnuson et al. 2003; Creel et al. 2006; Shatzman and McQueen 2006; Sulpizio and McQueen 2012; Farris-Trimble and McMurray, 2018) and the eye-movement patterns from studies using artificial lexica are remarkably similar to those from studies using real words: they show frequency effects, phonological neighbourhood effects, cohort competition, and rhyme competition, all with a time course
similar to that observed for real words (Magnuson et al., 2003). We therefore believe that comparing eye-movement data across different conditions using artificial words will indeed allow us to study the cognitive process we are interested in, namely the formation of predictions during spoken word recognition. The design of our experiment, described in detail in the next section, closely follows the design of Farris-Trimble and McMurray’s (2018) second experiment, differing mainly in the phonological content of the artificial languages being used.

4 Methods

4.1 Participants

Sixteen native English-speakers between the ages of 18 and 39 participated in the experiment (eight in the control group and eight in the harmony group). Participants were recruited through the University of Ottawa’s Integrated System of Participation in Research (ISPR). In light of the results from Finn and Hudson Kam (2008), which suggest that prior linguistic knowledge can interfere with a learner’s ability to acquire novel phonotactics, we screened out native English-speakers with intermediate or higher proficiency in another language to ensure that all participants had a comparable language background. To be sure that all participants had zero or low proficiency in an additional language to English, all participants also completed a language background questionnaire before beginning the experiment. Participants received course credit for taking part in the experiment. All participants reported normal hearing and normal or corrected-to-normal vision. Three additional participants were tested but their data were excluded from the analyses for poor eye-tracking quality, not being a native speaker of English, and for having under 50% accuracy on test trials.

4.2 Artificial languages

Two artificial languages were created. The first language, which acted as the control, contained no morphophonological alternations, whereas the second language contained a prefix that was subject to regressive sibilant harmony. Both languages contained the same 36 noun bases, each of which had three morphological forms: a bare form with no prefix, one form prefixed with underlying /ʃo/ and one form prefixed with underlying /fə/ These 36 morphological paradigms of an un-affixed base and its two prefixed versions were themselves divided into 12 groups of three paradigms called “triplets”. Where C represents a stop consonant, every triplet consisted of one paradigm whose base had the shape /CVʃV/, one paradigm whose base had the shape /CVsV/, and one paradigm whose base had the shape /CVʃV/. There were thus 12 morphological paradigms for each of the three base types. All three bases in a given triplet began with the same CV sequence but shared no further material. By sharing an initial CV sequence, the three bases in a given triplet all exerted the same co-articulatory effect on the prefix vowel but were otherwise easily distinguishable from one another.

The two artificial languages used prefixes to mark noun number: the bare form of the noun always meant ‘singular’, and the two prefixes /ʃə/ and /ʃo/ marked either dual number or triple/plural number. The meanings of the two prefixes were counterbalanced such that for half of the participants, /ʃə/ marked dual number and /ʃo/ marked triple/plural number, and vice versa for the other half such that /ʃə/ marked plural number and /ʃo/ marked dual number. Each language thus contained 12 triplets x 3 bases x 3 forms = 108 words. Participants were randomly assigned to either the control group or the experimental group. The difference between the two groups was in how the /ʃo/ prefix behaved. The control group was taught that the /ʃo/ prefix never alternated and always surfaced as [ʃo]. The experimental group was taught that the /ʃo/ prefix alternated and became [so] when attaching to /CVsV/ bases. Table 1 shows a sample of one triplet for the control group, and Table 2 shows the same triplet for the experimental group, with the morphophonological alternation highlighted in grey.
We also tested whether participants learned an abstract (generalizable) rule of sibilant harmony rather than memorizing which base went with which alternant of /fo/. To do this, participants were exposed to the full morphological paradigm (i.e., the unaffixed form and the two prefixed forms) for only nine out of the 12 triplets, or 27 out of the 36 bases. For the remaining three triplets (nine out of 36 bases), participants were exposed only to the unaffixed version. Following Farris-Trimble and McMurray (2018) we call the former set of words the trained items and the latter set of words the generalization items. The trained items were further split into two groups based on how often they were seen in the training phase. Three of the trained triplets (nine of the 27 trained bases) were seen more frequently than the remaining six trained triplets (18 of the 27 trained bases). Only the high-frequency trained items and the generalization items were used in the test phase; the low-frequency trained items appeared during training so as to provide ample evidence of the morpho-phonological alternation and therefore encourage its acquisition.

### 4.3 Auditory stimuli

The 36 novel noun bases (see Appendix) were composed of English phonemes and consisted of only open syllables. The stimuli were recorded in a sound-attenuated booth at the Ottawa University Sound Patterns Laboratory using version 2.2.1 of Audacity® recording and editing software (Audacity Team, 2017). The words were produced by a female native English speaker who was naïve to the purpose of the experiment. Recordings were made at a sampling rate of 44.1 kHz using a Shure Digital microphone. Farris-Trimble and McMurray (2018) recorded full words in order for the stimuli to be as natural as possible, however, we opted to record the prefixes and bases separately and then splice them together so that we could make the time-locking of the eye-tracking data more precise by normalizing the duration of the prefixes and the base-initial syllables. Using Praat (Boersma and Weenik 2018), we spliced six tokens each of the prefix allomorphs [fe-], [fo-], and [so-] out of carrier words of the shape [fe-Cfe], [[fo-Co]o], and [so-Ćoso], where C represents the six possible base-initial consonants [p, t, k, b, d, g]. For each prefix, there was thus one dummy base per possible base-initial consonant in order to create stimuli in which the prefix vowel’s formant transitions are consistent with the following base-initial consonant. All 36 bases were recorded in a carrier word of the form [C_sV_f-C_sV_fXX]; for example, the base [pas] was recorded in the carrier word [pa-pasi]. Again using Praat, we spliced the 36 bases out of their carrier words we separated them into their component syllables, calculated the average duration of the initial syllables, and modified the initial syllables so that they all had this average duration of 204ms (range = 183–224, sd = 12). We then calculated the average duration of the 18 prefix tokens and modified them so that they all had this average duration of 176ms (range = 158–196, sd = 12). These manipulations made the time-locking of the eye-tracking data more precise by having the length of the prefix constant in all prefixed words. Accordingly, the onset of the base-medial fricative always occurred at 380ms after word onset. Finally, we added 50ms of silence to the beginning of each audio file.
4.4 Visual stimuli

Words were paired with pictures of everyday items, and the word-image pairings were the same across all participants (see Appendix). The images were obtained from the University of California San Diego Center for Research in Language’s International Picture Naming Project database. There were three images for each object: one singular, one containing two instances of the object, and one with three instances of the object. All images were 300x300 pixels. The dual and plural images were constructed by making copies of the singular image, shrinking them, then arranging them in a 300 x 300 pixel square.

4.5 Procedure

The experiment took place in two sessions spaced one or two days apart, each lasting 45 minutes to an hour. The first day was split into two training phases. In the first training phase, participants learned the novel words in their singular, unaffixed forms. This was done to ensure that participants were sufficiently accurate at mapping the noun base to its object before exposing them to the two prefixes. On each trial, four pictures of different singular objects appeared on the computer screen and an audio recording of a word from the artificial language was played 750ms later. Upon hearing the word, the participants clicked on the image that they thought matched the word. If they were correct, they would hear a bell ding upon making their selection, and if they were incorrect, they instead heard a short buzz upon making their selection. In either case, all images except the correct image disappeared and a blue box appeared around the correct image simultaneously with the audio feedback. This first training phase consisted of 162 trials. The three high-frequency trained triplets and the three generalization triplets were presented six times each (6 triplets x 3 bases x 1 form x 6 repetitions = 108 trials), and the remaining six low-frequency trained triplets were presented three times each (6 triplets x 3 bases x 1 form x 3 repetitions = 54 trials).

In the second training phase (still on day one), participants learned the prefixed forms of the trained bases, but continued to see the generalization items in their unaffixed forms only. The procedure was the same as in the first learning phase. This second learning phase consisted of 378 trials. Fifty-four of these trials contained only singular images, like in the previous learning phase, and were broken down as follows: one trial per low-frequency triplet (6 triplets x 3 bases x 1 form x 1 presentation = 18 trials), two trials per high-frequency trained triplet (3 triplets x 3 bases x 1 form x 2 presentations = 18 trials), and two trials per generalization triplet (3 triplets x 3 bases x 1 form x 2 repetitions = 18 trials). The remaining 324 trials in this second learning phase mixed the various morphological forms of the bases. In these mixed trials, the four pictures on screen were composed of two pairs such that two different noun bases appeared in the same two out of three morphological numbers. For example, one such display contained two elephants, two banjos, three elephants, and three banjos. Each of the three bases in the six low-frequency trained triplets was a target three times in each of the three morphological forms (6 triplets x 3 bases x 3 forms x 2 presentations = 108 trials). Each of the three bases in the three high-frequency trained triplets was a target eight times in each of the three morphological forms (3 triplets x 3 bases x 3 forms x 8 presentations = 216 trials). The generalization items did not occur as a target in these mixed trials, nor were they the foils in any of these mixed trials.

The second day consisted of a short review session and a testing session. The review session was included in order to refresh participants’ memory of the noun-image pairings. This review session contained 171 trials and was broken down as follows: each of the three bases in all 12 triplets appeared as the target once in a trial containing only singular items like in the first training phase (12 triplets x 3 bases x 1 form x 1 repetition = 36 trials), each of the three bases in the six low-frequency trained triplets appeared as the target once in each of the three morphological forms in a mixed trial like in the second learning phase (6 triplets x 3 bases x 3 forms x 1 presentation = 54 trials), and each of the three bases in the three high-frequency trained triplets appeared thrice as the target in each of the three morphological forms in a mixed trial like in the second learning phase (3 triplets x 3 bases x 3 forms x 3 presentations = 81 trials). The generalization items did not occur as a target in these mixed trials, nor were they the foils in any of these mixed trials.
The role of long-distance phonological processes in spoken word recognition

The eye-tracker was calibrated immediately before the test session (see the next subsection for details regarding the eye-tracking). The test session had 360 trials, broken down into 288 critical trials (80% of trials) and 72 filler trials (20% of trials). During testing, items were presented only in their prefixed forms (i.e., there were no singular pictures / unaffixed forms). The procedure and visual displays differed slightly in the test phase. A red circle appeared at the centre of the screen along with the four images, and this circle turned blue 750ms later. Participants were instructed to click on this circle once it turned blue in order to hear the trial audio. This alternative procedure ensured that the mouse (and likely the participants’ gaze) was at the centre of the screen when the audio began to play. In all visual displays during the test phase, three of the four images matched in number (dual or triple/plural) and corresponded to the three bases from a single triplet all preceded by the same prefix. The fourth picture was the opposite-prefix form of one random member of the same triplet. An example image display for a trial in this testing phase is presented below in Figure 1.

![Example screen display of a trial from the test phase (participants saw no text)](image)

Figure 1. Example screen display of a trial from the test phase (participants saw no text)

On all critical trials, the target image was among the three images that matched in number. Each of the three bases in the three high-frequency trained triplets and the three generalization triplets appeared as the target in each of the two prefixed forms eight times each (3 triplets x 3 bases x 2 forms x 8 repetitions = 288 trials). The remaining 72 trials were fillers, in which the target image was the one that did not match in number with any other image on screen (i.e., when the prefix of the auditory stimulus matched only one item on the screen). Each of the three bases in the three high-frequency trained triplets and the three generalization triplets appeared in each of the two prefixed forms three times each (6 triplets x 3 bases x 2 forms x 2 repetitions = 72 trials).

4.6 Eye-tracking

Eye movements were recorded with an Eyelink 1000 eye tracker (SR Research). Participants were seated in front of the screen, and their head was supported by a fixed chin and forehead rest. The standard nine-point calibration occurred immediately before the testing phase began. Eye position was sampled every two milliseconds and a drift correction was run every trial to correct for small drifts in the calibration; the participants were recalibrated at this point if necessary. The Eyelink 1000 automatically divides the
recording into fixations, saccades, and blinks. Any fixation that fell within a 300x300 pixel image or a surrounding 50-pixel buffer (to allow for drift or slight error in the eye-track) was considered a fixation on that item. This additional 50 pixels did not result in any overlap among the regions of interest.

5 Results

5.1 Accuracy

Mean accuracy including the filler trials was 78.0% (range = 60.3–95.3, sd = 11.9), which was appreciably lower than the 88% mean accuracy reported in Farris-Trimble and McMurray (2018). Accuracy was particularly variable and overall low for critical trials on which a generalization item was the target (mean = 70.7, range = 29.9–96.5, sd = 16.7). Figure 2 shows a boxplot of mean accuracy broken down by experimental group, the training status of the target, and filler status.

![Boxplot of mean accuracy](image)

**Figure 2.** Mean accuracy in test trials by participant group, learning status, and filler status.

A possible reason for the lower-than-expected accuracy is the fact that present experiment had more items with cohort competition relative to Farris-Trimble and McMurray’s experiment. In their experiment, the three bases of a triplet overlapped in all word-medial phonemes but had distinct initial and final phonemes (e.g., their triplet [diʃ], [tip] and [zif]), whereas in the present experiment, the three bases of a triplet shared an entire initial syllable (e.g., our triplet [beso], [beʃa] and [befu]). This higher level of initial overlap could have negatively affected accuracy for one or both of the following reasons. First, the overlap could have interfered with learning. Second, the three “same-prefix” items in critical test displays might have competed
more strongly amongst themselves for recognition in the current experiment than in Farris-Trimble and McMurray’s.

5.2 Eye-tracking

The main hypothesis that the current experiment wished to assess was whether hearing the [so] alternant of the /ʃo-/ prefix would lead the harmony group to expect an upcoming CVsV item before the base fricative was realized. If this were the case, we could expect that on trials with a /ʃo-CVsV/ word as the target, the harmony group but not the control group would look significantly more towards the target image than the competitor images. In all other circumstances, we would expect neither group to look more to the target image than the competitor images. The onset of the base’s fricative occurred 380ms after word onset in all test items. Since it takes approximately 200ms to plan and execute an eye movement (Viviani, 1990), the 200–580ms time window is therefore the period in which we would expect to observe early activation of targets and/or early inhibition of competitors due solely to the [ʃo]/[so] alternation. Figure 3 shows how the mean fixation proportions to the four images change over time in critical CVsV trials. The vertical black bars mark the 200–580ms time window used in the statistical analyses below.

![Figure 3](image)

**Figure 3.** Time course of mean fixation proportions in CVsV critical trials. Vertical black bars demarcate the widow chosen for statistical analysis.

A participant’s eye movement data from the 200–580ms window were aggregated across each prefix and base combination (i.e., each word), yielding the total number of samples that fell within any of the on-screen images on trials where the participant heard that word, as well as separate counts of how many of those samples were allocated to a particular image. After doing so, we followed Farris-Trimble and
McMurray (2018) by adding the equivalent of half of the experiment’s average fixation (137.5ms or 69 samples) to each data point in order to avoid problems caused by the zero values from trials in which a participant did not look to any of the images during the time window. We then calculated the empirical logit of each proportion, rather than use the raw proportions in the statistical analyses below.

We fit linear mixed effects models with the lme4 package (Bates et al. 2015: version 1.1-18-1) in R Studio (version 1.1.453), using the empirical logit of the fixation proportions as the dependent variable. Separate models were fit for the trained and generalization items. Both models included the fixed effects of PREFIX (/fe-/ = -1, /fo-/ = -1), and GROUP (Control = -1, Harmony = 1). Both models additionally included a treatment-coded fixed effect of IMAGE that used the target CVsV image as the baseline (the other two levels being CVfV and CV\textsuperscript{V}). We followed Farris-Trimble and McMurray (2018) by aggregating the data across a triplet, calculating the empirical logit of the average proportions of looks to the unrelated image (i.e., the opposite-prefix item), centering these values, and including them as a covariate in both models, rather than include triplet as a random factor. To determine the optimal random effects structure for each model, we fit baseline models that had only per-participant intercepts, and then fit separate models that each added one of the possible per-participant random slopes. These latter models were individually compared to their appropriate baseline model using a chi-square test, and a random slope was added to a new baseline model if it significantly improved model fit ($p < 0.05$). The random slopes not added on this round were tested against the new baseline, and this process was repeated until no further random slopes significantly improved model fit.

The final model for the trained items is summarized in Table 3. It only included a per-participant slope of IMAGE. Note that the $p$-values on fixed effects were computed by using the Satterthwaite approximation for degrees of freedom, as implemented in the lmerTest package (Kuznetsova et al. 2017: version 3.0-1) in R Studio.

**Table 3.** Linear mixed-effects model of fixation proportions in trained CVsV trials

| Term                             | estimate | std. error | df    | $t$-value | $p$-value |
|----------------------------------|----------|------------|-------|-----------|-----------|
| (Intercept)                      | -1.08    | 0.12       | 16.54 | -9.12     | < 0.001   |
| F Image                          | -0.25    | 0.20       | 14.00 | -1.22     | 0.24      |
| F Image                          | -0.25    | 0.17       | 15.05 | -1.47     | 0.16      |
| Prefix                           | 0.064    | 0.084      | 258.11| 0.76      | 0.45      |
| Group                            | 0.025    | 0.12       | 16.55 | 0.21      | 0.83      |
| Unrelated (covariate)            | -0.36    | 0.37       | 248.22| -0.96     | 0.33      |
| F Image * Prefix                 | -0.15    | 0.12       | 258.26| -1.26     | 0.21      |
| F Image * Prefix                 | -0.048   | 0.12       | 258.12| -0.40     | 0.69      |
| F Image * Group                  | 0.0096   | 0.20       | 14.00 | 0.047     | 0.96      |
| F Image * Group                  | -0.017   | 0.17       | 15.05 | -0.10     | 0.92      |
| Prefix * Group                   | 0.083    | 0.084      | 258.11| 0.99      | 0.32      |
| F Image * Prefix * Group         | -0.21    | 0.12       | 258.26| -1.76     | 0.079     |
| F Image * Prefix * Group         | -0.075   | 0.12       | 258.12| -0.63     | 0.53      |

If the harmony group did indeed use the alternation to anticipatorily activate the target and/or inhibit the competitors during this early time window, we expected to see a significant three-way interaction between IMAGE, PREFIX, and GROUP such that the harmony group would make earlier fixations to the target CVsV image than to the competing CVfV or CV\textsuperscript{V} images when they encountered the /fo-/ prefix, but would show no preferential early looking otherwise; the control group would show no preferential looking amongst the three image types during this time window no matter the prefix. As can be seen in
Table 3, the expected interaction approached significance when comparing looks to the CVsV target and the CVfV competitor ($t(258.12) = -0.63, p = 0.53$). This marginal interaction is illustrated in Figure 4, which shows the harmony group’s mean fixation proportions in trained CVsV trials.

![Figure 4](image.png)

**Figure 4.** Mean fixation proportions in trained CVsV trials for the harmony group. Error bars represent the standard error of the mean.

For the harmony group, when the prefix was /ʃo-/ (left panel) there were more looks to the target (red bar) than looks to the CVfV competitor (green bar), and a post-hoc one-tailed $t$-test confirmed that this difference was significant ($t(23) = 2.15, p = 0.021$). Also for the harmony group, again when the prefix was /ʃo-/ (left panel), there were more looks to the target CVsV target (red bar) than to the CVfV competitor (blue bar), but a post-hoc one-tailed $t$-test revealed that this difference was not significant ($t(23) = 1.09, p = 0.14$). In all other circumstances, the proportion of looks to the target did not diverge from the proportion of looks to the other images. This seems to suggest that the harmony group was indeed using the prefix alternation to anticipate the presence of an upcoming [s]. The anticipation was not strong enough to prevent looks to the CVfV competitor, however, possibly due to the long-distance nature of the sibilant harmony driving the alternation. This would be in line with the results from Tobin et al. (2010), who found that anticipation due to regressive vowel-to-vowel coarticulation in English, a similarly long-distance phenomenon, was particularly small. It is also consistent with work by Kimper (2017) and Zellou and Pycha (2018) who found that distance can interfere with certain aspects of perception, and is consistent with the cross-linguistic dis-preference for increased distance in harmonic processes (Krämer 2003; Martin 2005; Hayes and Londe 2006; Hayes et al. 2009; Finley 2011, 2015, 2017; Kimper 2011; Jurgec 2011; McMullin and Hansson 2014; Zymet 2015; McMullin 2016).

All that being said, we have still only looked at the trained items, that is items for which participants in the harmony group directly saw the prefix alternation. We still do not know whether the anticipation is...
due to an abstract rule of sibilant harmony, or whether participants in the harmony group were simply memorizing which prefix alternant went with which base. In the latter case, we would expect them to not show any evidence of anticipation in the generalization items, since they did not directly witness which prefix alternant goes with which of those bases. The final model for the generalization items is summarized in Table 4. It did not include any random per-participant slopes. As above, the $p$-values on fixed effects were computed by using the Satterthwaite approximation for degrees of freedom, as implemented in the lmerTest package (Kuznetsova et al. 2017: version 3.0-1) in R Studio.

Table 4. Linear mixed-effects model of fixation proportions in generalization CVsV trials

|                      | estimate | std. error | df   | t-value | p-value |
|----------------------|----------|------------|------|---------|---------|
| (Intercept)          | -1.16    | 0.086      | 254.00 | -13.60  | < 0.001 |
| f Image              | -0.14    | 0.12       | 254.00 | -1.13   | 0.26    |
| F Image              | -0.010   | 0.12       | 254.00 | -0.083  | 0.93    |
| Prefix               | -0.12    | 0.086      | 254.00 | -1.43   | 0.15    |
| Group                | 0.12     | 0.086      | 254.00 | 1.38    | 0.17    |
| Unrelated (covariate)| -0.070   | 0.38       | 254.00 | -0.19   | 0.85    |
| f Image * Prefix     | 0.26     | 0.12       | 254.00 | 2.17    | 0.031   |
| F Image * Prefix     | 0.53     | 0.12       | 254.00 | 0.44    | 0.66    |
| f Image * Group      | -0.086   | 0.12       | 254.00 | -0.71   | 0.48    |
| F Image * Group      | -0.25    | 0.12       | 254.00 | -2.09   | 0.038   |
| Prefix * Group       | 0.033    | 0.086      | 254.00 | 0.39    | 0.70    |
| f Image * Prefix * Group | 0.016 | 0.12 | 254.00 | 0.13 | 0.90 |
| F Image * Prefix * Group | -0.048 | 0.12 | 254.00 | -0.40 | 0.69 |

As can be seen in Table 4, neither of the expected three-way interactions between IMAGE, PREFIX, and GROUP were significant. Instead, there were two significant two-way interactions. The first was between IMAGE = $f$ and PREFIX ($t[254] = 2.17, p = 0.031$), such that all participants looked to the CVJV image more in trials with /fjo-/ than they did in trials with /fe-. The second interaction was between IMAGE = $f$ and GROUP ($t[254] = -2.09, p = 0.039$), such that harmony group looked more to the CVfV image than did the control group. These two-way interactions were unexpected and difficult to interpret, though no matter how they are to be explained, the lack of the expected three-way interaction in the generalization items, despite its marginal presence in the trained items, suggests that participants in the harmony group were treating the prefix alternation as some sort of noun class marker. They did not show any preferential early looking to generalization CVsV target images when the prefix was [so-] because they never witnessed which prefix alternant went with which generalization base.

6 Discussion and conclusion

The fact that, in trials with a trained CVsV item as the target, the experimental group tended to fixate more on the target image than either competitor image during the 200–580ms time window seems to suggest that the harmony group did indeed use the prefix alternation to anticipate the identity of the upcoming word, since there is no other information available during this early time window to allow for disambiguation. The anticipation was not entirely significant, however, and did not extend to trials with a generalization CVsV item as the target. This mixed result is possibly due to the long-distance nature of the sibilant harmony driving the alternation, which would be in line with the results from Tobin et al. (2010), who found that anticipation due to regressive vowel-to-vowel coarticulation in English, a similarly long-distance
phenomenon, was particularly small. It would also be consistent with work by Kimper (2017) and Zellou and Pycha (2018) who found that distance can interfere with certain aspects of perception, and would be consistent with the cross-linguistic dis-preference for increased distance in harmonic processes (Krämer 2003; Martin 2005; Hayes and Londe 2006; Hayes et al. 2009; Finley 2011, 2015, 2017; Jurgec 2011; Kimper, 2011; McMullin and Hansson 2014; Zymet 2015; McMullin 2016).

It is, however, also possible that the lack of the expected result could have been due to certain aspects of the experimental design. Analyses of the accuracy data revealed that the experimental task may have been more difficult than intended, and we identified two potential reasons for the overall lower-than-expected accuracy. First, the test trials may have been more difficult than intended due to the large amount of cohort competition amongst three of the four on-screen images (all images shared the same base-initial syllable), and this added difficulty may have pushed participants to adopt a strategy whereby they mostly focused on the fricatives in the base, and only paid cursory attention to the fricatives in the prefixes. Second, the high amount of cohort competition within the artificial lexicon as a whole may have interfered with learning such that participants might not have been able to form representations of the individual bases that were strong enough to prevent confusion between words. In the case of the experimental group, these insufficiently strong lexical representations would in turn likely prevent any learning of the sibilant harmony rule.

We accordingly propose certain modifications for a future version of the current experiment. First, the artificial lexicon will be organized into pairs of bases rather than triplets of bases. This change will reduce the amount of cohort competition. Second, in light of the poorer accuracy on the generalization items during the test phase, we propose to add training trials with generalization items as the target in their /fθ/-prefixed forms (but still crucially never in their /fθ/-prefixed forms). This would increase these items’ frequency and therefore improve their acquisition, while still allowing us to test for generalization. Finally, we propose to have some of the low-frequency trained items begin with a sibilant, rather than have a sibilant as their second consonant. With such items included, participants in the experimental group would see sibilant harmony applying at shorter distances as well as longer distances, which may facilitate the acquisition of the pattern.

Further work is needed to clarify the role that long-distance phonological processes play in spoken word recognition, and we see two major directions that could be pursued. First, if the proposed amendments to the present experiment yield a positive result, such that participants indeed learn an abstract rule of sibilant harmony and utilize it to anticipatorily recognize spoken words, it would be revealing to directly compare the strength of this anticipation to the strength of a parallel anticipatory effect due to a local phenomenon. Second, sibilant harmony is but one of a wide variety of long-distance processes, and whether or not the amended experiment produces a positive result, other long-distance patterns should be investigated, since the properties of sibilant harmony aside from its long-distance nature may modulate its potential effect on spoken word recognition. In particular, it would be revealing to see whether and how a long-distance dissimilatory pattern is utilized during spoken word recognition, since some have theorized that long-distance assimilation and dissimilation arise from separate mechanisms (e.g., Agreement-by-Correspondence models: Rose and Walker 2004; Hansson 2010; Bennet 2015). The answers to these questions, whatever they may be, will have consequences for theories of spoken-word recognition and may shed new light on the typology of long-distance phonological processes.

Acknowledgments. This research was supported by the Social Sciences and Humanities Research Council of Canada. Thank you also to Bob McMurray and Ashley Farris-Trimble for their feedback on an early version of this experiment. Finally, thank you to the attendees at the 2019 Montreal-Ontario-Toronto Phonology/Phonetics Workshop for their insightful questions and comments.
References

Allopenna, Paul, James S. Magnuson, and Michael K. Tanenhaus. 1998. Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language*, 38(4), 419–439.

Audacity Team. 2017. Audacity(R): Free Audio Editor and Recorder [Computer application]. Version 2.2.1 retrieved July 12th, 2018 from https://audacityteam.org/

Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.

Bedder, Patrice S., Kevin B. McGowan, Julie E. Boland, Andries Coetzee, and Anthony Brasher. 2013. The time course of perception of coarticulation. *Journal of the Acoustical Society of America*, 133, 2350–2366.

Bennett, William G. 2015. *The phonology of consonants: harmony, dissimilation and correspondence*. Cambridge: Cambridge University Press.

Berkson, Kelly H. 2013. Optionality and locality: Evidence from Navajo sibilant harmony. *Laboratory Phonology*, 4(2), 287–337.

Blazej, Laura J., and Ariel M. Cohen-Goldberg. 2015. Can we hear morphological complexity before words are complex? *Journal of Experimental Psychology: Human Perception and Performance*, 41(1), 50–68.

Boersma, Paul, and David Weenik. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.42, retrieved 15 August 2018 from http://www.praat.org/

Carbary, Kathleen, Meredith Brown, Christine Gunlogson, Joyce M. McDonough, Aleksandra Fazlipour, and Michael K. Tanenhaus. 2015. Anticipatory deaccenting in language comprehension. *Language, Cognition and Neuroscience*, 30(1–2), 197–211.

Creel, Sarah C., Michael K. Tanenhaus, and Richard N. Aslin. 2006. Consequences of lexical stress on learning an artificial lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(1), 15–32.

Dahan, Delphine, James S. Magnuson, Michael K. Tanenhaus, and Ellen Hogan. 2001. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16, 507–534.

Desmeules-Trudel, Felix, and Tania S. Zamuner. 2019. Gradient and categorical patterns of spoken-word recognition and processing of phonetic details. *Attention, Perception, & Psychophysics*, 1–19.

Farris-Trimble, Ashley, and Bob McMurray. 2018. Morpho-phonological regularities influence the dynamics of real-time word recognition: Evidence from artificial language learning. *Laboratory Phonology*, 9(1), 1–34.

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

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

Finley, Sara. 2017. Locality and harmony: Perspectives from artificial grammar learning. *Language and Linguistics Compass*, 11(1).

Finn, Amy S., and Carla L. Hudson Kam. 2008. The curse of knowledge: first language knowledge impairs adult learners’ use of novel statistics for word segmentation. *Cognition*, 108, 477–499.

Friederici, Angela D., Karsten Steinhauser, and Erdmut Pfeifer. 2002. Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. In *Proceedings of the National Academy of Sciences* (Vol. 99, pp. 529–534).

Gow, David, and Bob McMurray. 2007. Word recognition and phonology: The case of English coronal place assimilation. In Jennifer Cole and José Ignacio Hualde (Eds.), *Papers in Laboratory Phonology* (Vol. 9, pp. 173–200). New York: De Gruyter Mouton.

Hansson, Gunnar Ölaufur. 2010. *Consonant harmony: long-distance interaction in phonology*. Berkeley, CA: University of California Press.
Hayes, Bruce, and Zsuzsa Londe. 2006. Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology*, 23, 59–104.

Hayes, Bruce, Kie Zuraw, Péter Siptar, and Zsuzsa Londe. 2009. Natural and Unnatural Constraints in Hungarian Vowel Harmony. *Language*, 85, 822–863.

Hayward, Richard J. 1990. Notes on the Aari language. In Richard J. Hayward (Ed.), Omotic language studies (pp. 425–493). London: School of Oriental and African Studies.

Hudson Kam, Carla L., and Elissa L. Newport. 2005. Regularizing unpredictable variation: the roles of adult and child learners in language variation and change. *Language Learning and Development*, 1, 151–195.

Hudson Kam, Carla L., and Elissa L. Newport. 2009. Getting it right by getting it wrong: when learners change languages. *Cognitive Psychology*, 59, 30–66.

Huetttig, Falk, Joost Rommers, and Antje S. Meyer. 2011. Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151–171.

Jurgec, Peter. 2011. *Feature spreading 2.0: A unified theory of assimilation* (Doctoral Dissertation). University of Tromso.

Kaplan, Elizabeth, Tatyana Levari, and Jesse Snedeker. 2016. Eye-tracking and Spoken Language Comprehension. In Christopher Was, Frank Sansosti, & Bradley Morris (Eds.), *Eye-tracking technology applications in educational research* (pp. 88–105). IGI Global.

Kaun, Abigail R. 1995. The typology of rounding harmony: An Optimality Theoretic approach (Doctoral Dissertation). University of California, Los Angeles.

Kimper, Wendell. 2011. *Competing triggers: Transparency and opacity in vowel harmony* (Doctoral Dissertation). University of Massachusetts Amherst.

Kimper, Wendell. 2017. Not crazy after all these years? Perceptual grounding for long-distance vowel harmony. *Laboratory Phonology*, 8(1), 19. https://doi.org/10.5334/labphon.47

Krämer, Martin. 2003. *Vowel Harmony and Correspondence Theory*. Berlin and New York: Mouton de Gruyter.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26.

Liversedge, Simon, and John M. Findlay. 2001. Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4(1), 6–14.

Magnuson, James S., Michael K. Tanenhaus, Richard N. Aslin, and Delphine Dahan. 2003. The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, 132, 202–227.

Mahr, Tristan, Brianna McMillan, Jenny R. Saffran, Susan E. Weismer, and Jan Edwards. 2015. Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition*, 142, 345–350.

Martin, Andrew. 2005. *The effects of distance on lexical bias: sibilant harmony in Navajo compounds* (Master’s Thesis). University of California, Los Angeles.

McMullin, Kevin. 2016. *Tier-Based Locality in Long-Distance Phonotactics: Learnability and Typology* (Doctoral Dissertation). University of British Columbia, Vancouver, BC.

McMullin, Kevin, and Gunnar Ö. Hansson. 2014. Locality in long-distance phonotactics: evidence for modular learning. In Jyoti Iyer & Leland Kusmer (Eds.), *Proceedings of the 44th meeting of the North Eastern Linguistic Society* (Vol. 2, pp. 1–14). Amherst, Massachusetts: GSLA Publications.

Moore-Cantwell, Claire, Joe Pater, Lisa Sanders, Robert Staubs, Benjamin Zobel. 2018. Event-related potential evidence of abstract phonological learning in the laboratory. Unpublished ms., University of Massachusetts.

Moreton, Elliot. 2008. Analytic bias and phonological typology. *Phonology*, 25, 83–127.

Morgan-Short, Kara, Karsten Steinhauer, Cristina Sanz, and Michael T. Ullman. 2010. Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, 60, 154–193.
Morgan-Short, Kara., Karsten Steinhauer, Cristina Sanz, and Michael T. Ullman. 2012. Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience, 24*, 933–947.

Ohala, John J. 1993. The phonetics of sound change. In Charles Jones (Ed.), *Historical linguistics: Problems and perspectives* (pp. 237–278). Routledge.

Paquette-Smith, Melissa., Natalie Fecher, and Elizabeth K. Johnson. 2016. Two-year-olds’ sensitivity to subphonemic mismatch during online spoken word recognition. *Attention, Perception and Psychophysiology, 78*, 2329–2340.

Rose, Sharon, and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language, 80*, 475–531.

Salverda, Anne Pier, Delphine Dahan, and James M. McQueen. 2003. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90*, 51–89.

Salverda, Anne Pier, Dave Kleinschmidt, and Michael K. Tanenhaus. 2014. Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language, 71*(1), 145–163.

Shatzman, Keren B., James M. McQueen. 2006. Prosodic knowledge affects the recognition of newly acquired words. *Psychological Science, 17*, 372–377.

Silva, Susana, Vasiliki Folia, Peter Hagoort, and Karl M. Petersson. 2016. The P600 in Implicit Artificial Grammar Learning. *Cognitive Science, 41*(1), 137–157.

Sulpizio, Simone, and James M. McQueen. 2012. Italians use abstract knowledge about lexical stress during spoken-word recognition. *Journal of Memory and Language, 66*, 177–193.

Suomi, Kari. 1983. Palatal harmony: a perceptually motivated phenomenon? *Nordic Journal of Linguistics, 6*, 1–35.

Tanenhaus, Michael K., James S. Magnuson, Delphine Dahan, and Craig Chambers. 2000. Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*(6), 557–580.

Tanenhaus, Michael K., Michael Spivey, Kathleen Eberhard, & Julie Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Tobin, Stephen, Pyeong Whan Cho, Patrick Jennet, and James S. Magnuson. 2010. Effects of Anticipatory Coarticulation on Lexical Access. In Stellan Ohlsson & Richard Catrambone (Eds.), *Proceedings of the 32nd annual meeting of the cognitive science society* (pp. 2200–2205). Austin, TX: Cognitive Science Society.

Trueswell, John C. 2008. Using eye movements as a developmental measure within psycholinguistics. In Irena A. Sekerina, Eva M. Fernández, & Harald Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children’s language processing* (pp. 71–96). Philadelphia, PA: John Benjamins Publishing Company.

Viviani, Paolo. 1990. Eye movements in visual search: Cognitive, perceptual, and motor control aspects. In Eileen Kowler (Ed.), *Eye Movements and Their Role in Visual Cognitive Processes. Reviews of Oculomotor Research* (Vol. 4, pp. 353–383). Amsterdam: Elsevier.

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

Zamuner, Tania S., Charlotte Moore, and Félix Desmeules-Trudel. 2016. Toddler’s sensitivity to within-word coarticulation during spoken word recognition: Developmental differences in lexical competition. *Journal of Experimental Child Psychology, 152*, 136–148.

Zellou, Georgia, and Anne Pycha. 2018. The gradient influence of temporal extent of coarticulation on vowel and speaker perception. *Laboratory Phonology, 9*(1), 12.

Zymet, Jesse. 2015. Distance-based decay in long-distance phonological processes. In Ulrike Steindl, Thomas Borer, Huilin Fang, Alfredo G. Pardo, Peter Guekguezian, Brian Hsu, Charlie O’Hara, and Iris C. Ouyang (Eds.), *Proceedings of the 32nd West Coast Conference on Formal Linguistics* (pp. 72–81). Sommerville, MA: Cascadilla Press.
Appendix: Nouns used in the experiment

| Triplet | Bases and their meanings |
|---------|--------------------------|
| 1       | pasi ‘wolf’  pafe ‘pear’  pafo ‘penguin’ |
| 2       | dosu ‘tent’  dofe ‘bird’  dofa ‘slingshot’ |
| 3       | kisu ‘trophy’ kife ‘safety pin’ kifo ‘wheelbarrow’ |
| 4       | busa ‘mouse’  bufo ‘skateboard’ bufi ‘broom’ |
| 5       | tesa ‘camel’  tefi ‘fan’  tefu ‘teapot’ |
| 6       | gasi ‘baby’  gafu ‘bandage’ gafe ‘drum’ |
| 7       | puse ‘bench’  pufi ‘moose’  pufa ‘slide’ |
| 8       | tosi ‘purse’  tofa ‘squirrel’  tofe ‘panda’ |
| 9       | guso ‘monkey’ gufi ‘glass/cup’ gufa ‘shovel’ |
| 10      | beso ‘toilet’ befu ‘banjo’  befu ‘snail’ |
| 11      | dise ‘burger’  difu ‘lamp’  difo ‘shirt’ |
| 12      | kase ‘clock’  kafe ‘leaf’  kafo ‘elephant’ |