Chapter 39: Psycholinguistic Approaches to Vowel Harmony

Anne Pycha, University of Wisconsin, Milwaukee, pycha@uwm.edu
Sara Finley, Pacific Lutheran University, finleysr@plu.edu

39.1 Introduction

Despite the tremendous diversity exhibited by the languages of the world, certain grammatical patterns tend to recur frequently, while others are rare or altogether unattested. Indeed, as Greenberg's (1963) work highlighted, the set of actually-attested patterns – in phonology as well as other domains, such as syntax – is much smaller than the set of possible patterns. This broad observation also holds true for vowel harmony (Hyman 2002): to cite just one simplified example, front vowels are more likely to be targets of height harmony than back vowels (Linebaugh 2007). What lies at the source of such typological asymmetries?

Sometimes the recurrence of a particular pattern is due to common ancestry among languages, or to borrowing between them. But in cases where such factors are ruled out, an intriguing third possibility emerges, namely, that the human mind/brain may possess a *bias* for certain patterns over others. A given bias might, for example, ensure that harmony systems which target back vowels will be less likely to arise over historical time and, when they do arise, be less likely to survive over generations. The over-arching hypothesis is that characteristics of the human/mind brain – which may be either endowed with a universal grammar, or constrained by imperfect signal transmission, or both – place fundamental limitations on what a language can look like.

In this chapter, we analyze the evidence for psycholinguistic biases with regard to vowel harmony, and evaluate the relevant conclusions about how harmony patterns are represented in the mind and brain. Most of the relevant studies are behavioral, meaning that they allow us to make inferences about how harmony might be represented in the mind. Only a small, albeit growing, number of studies report measurements of brain activity; within this line of research, the focus is on questions of phonological processing and its time course. We do not know of any work that has attempted to anatomically "localize" vowel harmony, and, given the fact that language processing generally tends to be distributed across many different areas of the brain (e.g., Hickok & Poeppel 2007), we would not necessarily expect such attempts to be fruitful.

Biases that shape the phonological patterns of the world's languages could conceivably arise from a variety of potential sources, although, as we will see, research on vowel harmony systems remains mostly inconclusive on this point. With regard to height harmony, for example, it is possible that the human mind/brain is endowed with a universal grammar which simply dictates that such systems should target front vowels. This type of grammar-based proclivity has been referred to as an *analytic* bias (Moreton 2008). A different possible source of bias, sometimes referred to as *channel* bias, originates in errors of transmission between speakers and hearers (Moreton 2008). For example, the back vowel [o] is relatively more difficult to perceive than other vowels (Terbeek 1977), which might mitigate against its inclusion in height harmony systems. We touch on these issues for several of the studies that we analyze, although our discussion remains speculative.

In the sections that follow, we first introduce the common experimental paradigms for researching vowel harmony in the mind (39.2) and the brain (39.3). We summarize research showing that vowel harmony patterns can be learned with only brief exposure, in a variety of training paradigms (39.4). This set of results indeed suggests that human listeners may be endowed with a set of biases that favor rapid acquisition of certain linguistic patterns, a broad hypothesis that has been tested in many different ways (39.5). Learning after brief exposure also suggests that participants are able to generalize well beyond the particular stimuli that they train with, and the appropriate level of generalization has been explored in numerous studies (39.6). As mentioned above, many experimental explorations of vowel harmony have used American-English speaking participants, but some studies have explicitly investigated the role of L1 in learning new harmony patterns (39.7). Harmony languages sometimes contain exceptions that pose additional challenges for learners (39.8). Finally, we turn our attention to studies that take explicit measurements of brain activity while participants produce or listen to harmony patterns (39.9). Because many of the questions raised by consonant harmony are similar to those raised by vowel harmony, we include some relevant studies that focus on consonant harmony. Our focus is on studies using adult participants. Discussion of vowel harmony in during child language acquisition can be found in Chapter 38.

39.2 Techniques for studying the mind

The majority of research conducted on vowel harmony in the mind and brain has used the artificial grammar learning paradigm. With artificial grammar learning, the experimenter typically constructs a small and highly controlled set of nonsense words, often in pairs of stem and stem-plus-affix that exemplify a particular pattern. The experimenter then presents these nonce words to participants in a laboratory setting. The participants are trained on the given artificial language, and then tested for their learning and generalization to the given linguistic pattern. For example, an artificial back vowel harmony language might demonstrate this harmony with a suffix that alternates between [-ɛk] for front vowel stems, and [-ɔk] for back vowel stems. On a training trial, a participant may hear [gip], [gip-ɛk] or alternatively, [gup], [gup-ɔk]. This training phase often consists of passive listening (but can also involve participants repeating the word), and the participants are usually 'naïve' listeners with no previous exposure to the pattern in question. After training, participants complete a test phase in which they make grammaticality judgments for both old and new affixed stimuli, often in the form of a two-alternative forced-choice task (but can also involve a rating or accepting/rejecting a given item). For example, on a test trial, a participant may hear [gip-ɛk] and [gip-ɔk] and then be asked to choose which form is most likely to belong to the language that they just heard. Or, a participant may hear [gip-ɛk] and select 'correct' or 'incorrect'. If participants' judgments are overall more accurate than a baseline measurement (say, 50% for random guessing), this is interpreted as evidence that participants successfully learned the pattern, particularly if their judgments are accurate for new, unheard words (e.g., [bit-ɛk]). There are many variations on this basic theme, which we mention in the studies reviewed below. For all of them, however, the linking hypothesis is that if participants can learn the pattern exhibited by the artificial grammar, and especially if they can extend the pattern to novel items after only limited exposure, then they possess a presumably universal bias for such patterns in the mind.

In interpreting the results of artificial grammar learning studies, there are several caveats. The training procedure often relies on explicit learning of a pattern, using meaningless words presented in an isolated context, which does not mimic natural acquisition. The testing method uses forced-choice tasks, which are a rather blunt means of capturing what participants have learned. And, at least in the studies to date, the overwhelming majority of participants have been speakers of American English, whose results need not be representative of any universal truths. Despite these drawbacks, artificial grammar learning has been an unusually rich source of findings about the representation of phonological patterns in the mind, and we discuss them extensively in the sections that follow.

Although they are less well-represented in the literature, there are other useful techniques for probing the representation of harmony in the mind, which we describe on in the sections that follow. In speech segmentation tasks (Saffran, Newport & Aslin 1996), for example, participants listen to a stream of nonsense syllables in which 'words' are defined by transitional probabilities between vowels or consonants. For example, words may have higher transitional probabilities across syllables than partial words do (e.g., in *pretty baby*, the syllable [be] is more likely to be followed by the syllable [bi] than it is to be preceded by [ti]). If all of the words in a segmentation task observe vowel harmony, then disharmony between adjacent syllables could indicate a word boundary. The linking hypothesis is similar to that of artificial grammar learning: namely, that if participants can segment the speech into 'words' after brief exposure, they possess a presumably universal sensitivity to vowel co-occurrence restrictions.

Speech segmentation tasks have also been given to L1 speakers of vowel harmony languages. Such studies have investigated, for example, whether Finnish or Turkish listeners use harmony patterns to make decisions about the location of word boundaries (Suomi, McQueen & Cutler 1997; Kabak, Maniwa & Kazanina 2010). To our knowledge, speakers of other languages exhibiting vowel harmony, in e.g., Akan, Iberian, and Bantu languages, have not been investigated with these methods, and patterns from those languages may therefore be under-represented in our understanding.

Other studies recruit native speakers of harmony languages to see how they respond to various types of stimuli. In "wug" tests (Berko 1958), the stimuli are typically novel, and the goal is to investigate the extent to which harmony patterns that occur in actual words generalize to new nonsense words. As we shall see, such generalizations suggest that harmony has an abstract representation in the mind, although there can nevertheless be a high degree of variation in terms of when and how harmony is applied.

39.3 Techniques for studying the brain

While there is less research that directly studies vowel harmony and the brain, the two main approaches are case studies of patients with aphasia (i.e., impairment to language abilities as a result of brain damage), and brain activity studies in healthy populations. In aphasia studies, the general goal is to understand what aspects of a patient's language abilities are spared, and which are impaired. The specific brain damage can be correlated with the impairment in order to draw inferences about the location of linguistic processes and representations.

Meanwhile, studies with healthy populations have used electrical or magnetic measurements to correlate brain activity with behavior. Electroencephalography (EEG) uses electrodes on the participant's scalp to measure electrical activity in the brain. This electrical activity can be correlated with linguistic behavior. While EEG has great temporal resolution, its spatial resolution is very poor. Magnetoencephalography (MEG), which records changes in magnetic fields in the brain, has both high spatial and temporal resolution, and so can be used to find an approximation of where in the cortex various linguistic behaviors might occur. Both EEG and MEG studies tend to make use of 'oddball' paradigms, where participants' brain activity is recorded while listening to stimulus sequences in which there is a 'standard' (majority) item type and an 'oddball' minority type. This oddball paradigm is advantageous because the brain's Mismatched-Negativity (MMN) response, which occurs when a deviant or 'surprise' stimulus occurs in an otherwise regular sequence, can be reliably produced in a variety of modalities, and has a long history in the neuroscience literature (for a review, see Luck 2014). These responses are tightly time-locked to the presentation of stimuli, allowing researchers to infer that a participant's brain responded to a specific event in the unfolding speech stream, such as the onset of a particular vowel or consonant. This can allow the researcher to address questions about the timing of processing, but also the representation of the patterns at play (e.g., different patterns of responses for harmonic vs. disharmonic items). EEG and MEG are non-invasive, and participants in these studies can typically listen passively to linguistic stimuli without performing any task. As we will see, linguistics researchers have leveraged these advantages to make some interesting conclusions about the representation of harmony in the brain.

Readers may be also familiar with functional magnetic resonance imaging (fMRI), a technique for measuring changes in blood flow to particular areas of the brain. Although fMRI has been used in studies investigating a variety of different language functions, none of these studies, to our knowledge, investigate harmony.

39.4 Rapid learning across several different training paradigms

A major finding, replicated in many different studies with various methodologies, is that naïve listeners can learn harmony patterns relatively well, even after limited exposure. Many of these studies used explicit training with artificial grammars. For example, Pycha, Nowak, Shin and Shosted (2003) trained adult American English listeners on nonsense words that conformed to front-back vowel harmony. During the listening phase, participants passively listened to eighteen pairs of stems and suffixes, such as [gip]…[gip-ɛk] and [sun]…[sun-ʌk]. In a subsequent training phase, participants heard suffixed forms that were either harmonic or disharmonic, made a forced-choice judgment about whether the word was grammatical or ungrammatical, and received feedback. Finally, during the testing phase, participants again heard suffixed forms and made grammaticality judgments, without feedback. Results showed that participants achieved a mean score of 86% correct in the testing phase, suggesting that they successfully learned the harmony pattern. Each participant spent only 15 to 20 minutes on the entire experiment, indicating that learning was very rapid.

Other studies have used implicit training, with similar results. For example, Skoruppa and Peperkamp (2011) exposed adult French listeners to a set of four short stories spoken in a novel 'accent'. The stories included actual French words that conformed to a front-vowel rounding harmony pattern, e.g., standard French *pudeur* [pydœʁ] 'prudity'. They also included words that were altered so as to conform to harmony, e.g., standard French *naturel* [natyʁel] 'natural' was pronounced [natyʁœl] and [likœʁ] *liquère* 'liquor' was pronounced [likɛʁ]. Participants were asked to simply attend to the stories, and took a short quiz on the content. At test, participants heard pairs of harmonic and disharmonic words, such as [likɛʁ]~ [pydɛʁ], and were asked to select the word that was pronounced in the 'accent' they had heard. Results showed that, for both old and new pairs, performance differed significantly from chance, suggesting that participants successfully learned the harmony pattern.

Vowel harmony can also be learned implicitly through a speech-segmentation task (Saffran, Newport, and Aslin 1996), where participants use implicit learning strategies to infer word boundaries. Bonatti, Pena, Nespor, and Mehler (2005) exposed adult French listeners to two seven-minute streams of artificial speech created by concatenation of nonsense words. In their Experiment 2, words were constructed such that the transitional probabilities between vowels was 1.0, while probabilities between consonants was 0.5. For example, one family of words consisted of [pɔ̃kima], [pɔ̃ʀila], [tɔ̃kila], [tɔ̃ʀima] and another consisted of [mopɛ̃ky], [motɛ̃ʀy], [lopɛ̃ʀy], [lotɛ̃ky]. At test, participants heard two strings, one word (e.g., [pɔ̃kima]) and one pseudo-word (a string that occurred at exposure, but spanned word boundaries, e.g. [mamopɛ̃], created from the final syllable of [pɔ̃kima] and the first two syllables of [mopɛ̃ky]). Their task was to choose the string that looked as if it came from the exposure language. Results indicated that listeners were more likely to choose words than pseudo-words, suggesting that they successfully learned the pattern of dependency between vowels (see comparable results from Newport & Aslin 2004). Taking these findings together with those of Pycha et al. (2003) and Skoruppa and Peperkamp (2011), we see that listeners show evidence for learning vowel harmony after relatively short exposure periods, with both implicit and explicit exposure.

39.5 Learning biases

39.5.1 Biases for attested patterns

How can naïve participants can learn harmony patterns so quickly? As discussed in Section 39.1, one possibility is that the human mind possesses a bias which facilitates learning of vowel harmony patterns, presumably to the exclusion of other patterns, such as vowel dis-harmony. Given the fact that harmony is well-attested in languages of the world while disharmony is only sparsely attested, such an explanation would seem plausible, although the experimental evidence is decidedly mixed. In their artificial grammar experiment, Pycha and colleagues (2003) compared learning of front-back vowel harmony, front-back disharmony, and an arbitrary pattern in which certain stem vowels triggered harmony while others triggered disharmony. Results showed that there was no significant difference between the harmony and disharmony conditions, but that both patterns were learned better than the arbitrary pattern. In their implicit learning experiment, Skoruppa and Peperkamp (2011) compared rounding harmony and disharmony, and also reported equivalent learning across conditions. More recently, however, Martin and Peperkamp (2020) critiqued the relatively low statistical power of these previous studies. In an explicit artificial grammar learning study that included 173 American English listeners, participants who learned a front-back vowel harmony pattern demonstrated significantly better learning than those who learned front-back disharmony. Although the authors note that performance was low in both conditions, this finding nevertheless suggests that a bias toward widely-attested patterns may be operative. Furthermore, Martin and White (2021) compared learning of harmony versus disharmony patterns that were iterative (i.e., spanning across multiple affixes, whereas most other studies employ just a single affix), and their results showed a bias towards harmony. This is consistent with the cross-linguistic distribution of harmony and disharmony; harmony can apply iteratively, while disharmony generally only applies once.

Other studies have found additional evidence supporting a bias for one type of long-distance pattern over another. For example, Lai (2015) compared sibilant consonant harmony with 'first-last' agreement (see Chapter 34 for details about computational approaches to vowel harmony and learnability). The harmony pattern requires that all sibilants in a word agree in anteriority (e.g., [s…s…C…s], [ʃ … ʃ…C… ʃ]), and is attested in languages such as Navajo. The first-last pattern requires that initial and final sibilants agree in anteriority (e.g., [s…C… ʃ…s], [ʃ …C… s… ʃ], and is not attested in any natural language. The results of this explicit artificial grammar learning experiment showed that while the sibilant harmony rule appeared to be internalized, the first-last rule was not.

Another way to test for learning biases is to provide learners with exposure items that are ambiguous between two possible rules, and test how they generalize to novel items that disambiguate the rule. For example, Finley and Badecker (2008) trained participants on items that were ambiguous between a "majority rules" pattern versus a directional harmony pattern. Crucially, "majority rules" is an unattested harmony pattern in which the trigger is determined by the number of vowels in the input. For example, if the input contains two round vowels and one unround vowel, the output target will be round, regardless of direction of spreading or featural dominance. On each trial, participants were exposed to an alternation consisting of a single disharmonic word followed by its harmonic counterpart, e.g., [beguto] → [boguto], in which the pattern could be interpreted either as majority-rules (because there are two round vowels in the first item) or as right-to-left directional spreading (a left-to-right condition was also included). At test, participants were presented with two different alternations, such as [kukope] → [kikepe] vs.

[kukope] → [kukopo], and were asked to choose which one was correct. Note that both alternations satisfy majority-rules, but [kukope] → [kikepe] is consistent with right-to-left directional spreading, while [kukope → kukopo] is not. Results showed that participants were more likely to choose the directional alternation, supporting a bias towards the typologically more common directional pattern. Interestingly, a subsequent experiment showed that the bias did not hold for a non-linguistic version of this task (Finley & Badecker 2010).

White and colleagues (2018) also presented participants with exposure items that were ambiguous regarding the source (trigger) for harmony, and tested to see whether learners made generalizations conforming to a local pattern (between two adjacent syllables), or a non-local pattern (between two non-adjacent syllables). At training, participants heard stem-affixed pairs that conformed to front-back vowel harmony in either a suffixing condition, e.g., [gini], [gini-be] or a prefixing condition [gini], [be-gini]. At test, participants heard disharmonic stems, and made a forced choice between two affixed forms (e.g., [guni-be] versus [ginu-be] in the suffix condition, or [be-guni] versus [be-ginu] in the prefix condition). They key question was whether participants would assume a local harmony rule, in which the affix vowel is affected by the adjacent root vowel, or a non-local rule, in which the affix vowel is affected by the non-adjacent root vowel. Results showed that there was a greater preference for local harmony compared to a non-local process (furthermore, this preference was greater in the suffix condition compared to prefix condition).

39.5.2 Biases as tendencies, rather than absolutes

Other evidence, however, tempers the conclusion for strong learning biases. Avcu (2018) used the same stimuli and training procedure as Lai (2015), but implemented an oddball testing paradigm, in which participants listened to sequences of words and were asked to detect the ungrammatical ones. Notably, participants in both the harmony and first-last conditions could successfully discriminate grammatical and ungrammatical words, although performance was better in the harmony condition. Finley (2012a) also investigated the learnability of a first-last pattern, but varied the morphological relationship among the training words. In the morphological condition, participants were exposed to pairs of nonsense words that observed first-last agreement in consonant voicing, such as /kidat gidad/, and were told that the first word was “singular” and the second was “plural”. In the phonotactic condition, by contrast, participants were exposed to the same words, but they were presented individually rather than in pairs, and with no morphological information. At test, listeners heard a single word and decided whether or not it was a word in the exposure language. Results showed an effect of condition, such that participants in morphological condition performed significantly better than those in the phonotactic condition, who were at chance. Taken together, the result of Avcu (2018) and Finley (2012a) suggest that, although a bias for attested patterns may exist, it does not completely rule out the successful learning of unattested patterns (for a related set of results, see Baer-Henney, Kügler, and van de Vijver (2015)).

39.5.3 Biases for sub-patterns within harmony

Studies have also addressed the question of learning biases by comparing asymmetries within a particular type of harmony. For example, Finley (2012b) and Kimper (2016) both investigated the learnability of vowel rounding harmony with high versus non-high vowel triggers (e.g., [gubu-mu], [gibi-mi] versus [gobo-mu], [gebe-mi]). Finley (2012b) reported that participants exposed to non-high triggers performed significantly better than those exposed to high triggers in generalizing the rounding harmony to new stems (for old stems, there was no difference between the groups). In addition, Kimper (2016) showed that participants exposed to non-high triggers were more successful at generalizing the pattern to other vowel heights. Specifically, the non-high group generalized rounding harmony to high vowels (e.g., after learning [gobo-mu], they extended this to [gubu-mu]), while the high group did not generalize to non-high vowels. A learning bias for non-high triggers in rounding harmony is consistent with typological investigations, which have indicated that, across languages, vowel rounding harmony is more likely to occur when the trigger vowel is non-high (Kaun 2004) (see Chapter 5 for more discussion on round harmony). This is exemplified in Yakut, where non-high vowels trigger rounding harmony regardless of target vowel height (oɣo-nu ‘child-Acc’, ohoχ-tor ‘stoves-Pl’), but high vowels do so only when the target agrees in height (tuːnnuːg-u ‘window-Acc’, tuːnnuːk-ter ‘window-Pl’) (Kaun 1995; Kimper 2016). Citing perceptual evidence from Terbeek (1977), Kaun (2004) suggested that such patterns could have their origins in a type of channel bias, with the idea being that roundness is more difficult to perceive in non-high vowels than in high vowels, and that harmony can enhance the perceptibility of such vowels by prolonging the duration of its features across a domain (Suomi 1983; Kaun 2004). Terbeek's original work (1977), however, actually shows that non-high vowels are well-separated perceptually from other vowels. The precise nature of this particular bias thus remains unresolved.

Finley and Badecker (2012) also examined an asymmetry, with a focus on vowel height harmony (see Chapter 6 for further discussion on vowel height harmony). They asked some participants to learn a pattern in which suffix front vowels alternated according to stem vowel height (e.g., [bemeg-e], [dunig-i]), and other participants to learn a pattern in which suffix back vowels alternated ([bemeg-o], [dunig-u]). Results from the testing phase indicated that participants in the front vowel condition successfully learned the pattern, while those in the back vowel condition did not. In addition, participants extended the pattern to novel front vowels, but not to novel back vowels. A bias for front vowel targets in height harmony is consistent with typological investigations. In a study of over 100 languages with height harmony, Linebaugh (2007) cited at least fifty-nine languages with symmetrical patterns (i.e., both front and back vowels serve as targets), fifty-two languages with asymmetric patterns favoring front vowels (i.e., front vowels are always targets, but back vowels are targets only when another back vowel is the trigger), and only two languages with potentially asymmetric patterns favoring back vowels. In addition, as discussed in Finley and Badecker (2012), a number of other factors may also be relevant, such as the relatively poor perceptibility of the back vowel [o] (Terbeek 1977), stronger F1 co-articulation across front vowels than across back vowels (Recasens & Pallarès 2000), or greater homogeneity in constriction location for front vowels compared to back vowels. One or more of these factors could potentially form the basis of a channel bias, although further research would be necessary to link them directly to the patterns in question.

Moreton (2008) also examined vowel height harmony, albeit with a somewhat different goal. In an effort to disentangle analytic bias (e.g., arising from a universal grammar) from channel bias (i.e., arising from errors of transmission between speakers and hearers), he compared height-height patterns, in which there is a dependency between the height of neighboring vowels, with height-voice patterns, in which there is a dependency between the height of a vowel and the voicing (or aspiration, or fortis-lenis status) of the following consonant. In a survey of languages, Moreton (2008) showed that the height-height pattern is more commonly attested than the height-voice pattern, which suggests that some factor encourages height-height patterns more than height-voice patterns. In a review of previously-published experiments, he concluded that the phonological height of a vowel exerts similar effects on a) F1 of a target vowel and b) voicing of a following consonant, which suggests that height-height and height-voice patterns are both phonetically natural, meaning that they both "resemble exaggerated or stylised expressions of some phonetic fact" (Moreton 2008: 87). Finally, Moreton conducted an artificial grammar learning experiment. At training, participants were exposed to CVCV nonce words that conformed to either height-height (e.g., [tiku], [tækɔ]) or height-voicing patterns (e.g., [tigɔ], [tæku]); at test, they made a forced-choice judgment between one item that conformed to the pattern they had learned, and one item that did not. Results showed that participants chose the correct item significantly more often in height-height condition than in height-voice condition. The overall conclusion of Moreton's (2008) study is that analytic bias – that is, some sort of universal grammatical tendency or constraint – is sufficient to shape to typological asymmetries, without the contribution of channel bias.

39.5.4 Biases for directionality

Finley and Badecker (2009b) examined another potential learning bias, namely directionality. As they noted, Hyman's (2002) cross-linguistic investigation suggests that harmony patterns tend to be either stem-controlled, with features spreading from stems to affixes, or directional from right-to-left (see Chapter 24 on directionality in vowel harmony). When working with an attested natural language, it can be difficult to distinguish between these factors. In Hungarian, Finnish, and Turkish, for example, vowel features always spread rightward, suggesting a role for directionality. However, we could alternatively analyze their harmony patterns as stem-controlled, and directionality would be merely a by-product of the fact that these languages are suffixing (Bakovic, 2000). Since these languages have no attested prefixes, it is not possible to distinguish between these two accounts.

Working with artificial grammar learning allows us to take a different angle on this issue. In Finley and Badecker (2009b)'s Experiment 1, American English-speaking adults were exposed to nonsense words exhibiting stem-controlled rounding harmony in one of two conditions: exclusively prefixing, or exclusively suffixing. At test, they made judgments about new words containing the affix type they had not previously been exposed to (suffixes for the prefixing condition, and prefixes for the suffixing condition). Results showed that both groups successfully generalized the harmony pattern to the new types of words. In Experiment 2, participants were exposed to a dominant-recessive phonological pattern, in which the presence of any round vowel triggered spreading to the remaining vowels in the word. In their design, the trigger vowel always occurred on an affix. As in Experiment 1, there were two training conditions in which listeners were exposed to nonsense words that were either exclusively prefixing or exclusively suffixing; at test, participants heard new words containing the affix type they had not previously been exposed to. Results showed that participants in the prefixing condition did not learn the pattern. The bias against prefix-triggered harmony matches Hyman's (2002) assessment; in his analysis, this is exactly the type of harmony that is rarely or never attested. These results also support the distinction between stem-controlled and dominant-recessive harmony (Baković 2000), and are consistent with other studies which show that biases can be crucially modulated with morphological information (Finley 2012a; White et al. 2018).

39.6 Level of Generalization

Although phonologists typically discuss harmony in terms of rules or constraints, listeners do not typically experience it this way. Rather, their exposure to harmony – whether as native speakers, or in the lab – is always instantiated through words. How do we know that listeners generalize beyond the individual words that they hear? In other words, how do we know that they form a pattern at all? The literature, including some studies reviewed in Sections 39.4 and 39.5, holds a wealth of evidence on this point. One set of evidence comes from "wug tests" (Berko 1958), in which native speakers are asked to make judgments about nonsense words. Zimmer (1969) showed that Turkish participants preferred nonsense words that conform to Turkish front/back and rounding harmony. Yavaş (1980) showed that Turkish participants applied harmony to epenthetic vowels in nonsense words, and also to vowels in suffixes attached to foreign words. Using a word game with novel inputs, Campbell (1986) reports similar results for Finnish. In all of these studies, speakers readily extended their native harmony pattern to words that they had never heard before. If speakers simply memorized individual words without generalizing over them, such a scenario would be impossible. Note, however, that many experiments with nonce words find both inter-speaker and intra-speaker variability in responses (Gósy 1989; Hayes &

Londe 2006; Duncan 2015). This variability suggests that knowledge of vowel harmony may be stochastic (i.e., randomly determined by the speaker's lexicon; Hayes & Londe 2006) and raises questions about the categorical nature of phonological alternations like vowel harmony.

The second set of evidence comes from artificial grammar learning studies, many of which were discussed in Sections 39.4 and 39.5. In the majority of these studies, the test phase crucially includes both 'old' words (which participants heard during training) as well as 'new' words (which they did not hear during training). This key feature of the experimental design allows researchers to evaluate whether participants merely memorized a set of words (in which case, test performance would be good on old words, but not new words), or whether they actually learned a pattern that could be applied more generally (in which case test performance would be good on both 'old' and 'new' words). In general, experimenters find support for learning when participants extend the harmony pattern to these new items, which demonstrates that participants rapidly learn a phonological pattern, rather than a list of words (Pycha et al. 2003). In some cases, participants fail to learn the harmony pattern, but still succeed on familiar items. For example, in her artificial grammar learning study of rounding harmony, Finley (2012b) found that participants in two conditions performed similarly on old test items, but that only participants in the non-high-trigger condition generalized their learning to new test items.

A related issue concerns how learners encode phonological patterns, specifically whether learners make use of abstract representations like phonological features. Most vowel harmony patterns can be captured with feature-based statements, such as *within a word, all vowels agree for the feature [back]*. In theory, however, the same patterns can also be captured by listing co-occurrence restrictions between vowels, such as *if the first vowel is [u], the next vowel must be [u], [o], or [a]*. How do we know that listeners are actually making feature-based generalizations, rather than learning a simple list of co-occurrence restrictions? One approach is to compare co-occurrence restrictions that can be captured by features to those that cannot. For example, Pycha and colleagues (2003) compared the learning of vowel front/back harmony and front/back disharmony with an 'arbitrary' condition, in which stems containing [i, æ, ʊ] triggered front vowel suffix [-ɛk], while stems containing [ɪ, u, a] triggered the back vowel suffix [-ʌk]. Results showed that test performance in the arbitrary condition was at chance, and significantly below performance in the other conditions. Another approach is to expose participants to a harmony pattern with one set of vowels, and then test to see if they extend the pattern to new vowels. For example, Finley and Badecker (2009a) trained English-speaking participants on a back/round vowel harmony pattern, exposing them to four vowels from a six-vowel inventory. At test, they introduced the two new vowels. Results of Experiment 1 showed that participants readily generalized to novel mid vowels, although not to novel low vowels. This result is consistent with the fact that, cross-linguistically, low vowels do not typically participate in rounding harmony. However, American English does not have a low, round counterpart to the stimulus vowel [æ], a factor that may have influenced participant responses. In Experiment 2, stimuli were altered so as to be compatible with back harmony only (and not rounding harmony); results showed that participants successfully generalized to both novel mid vowels and novel high vowels.

Interestingly, feature-based generalizations do not seem to extend to cases of consonant harmony. In a study designed to reproduce the learning conditions of Finley and Badecker (2009a), Finley (2011) exposed listeners to a pattern of continuant harmony in which a stop in the root triggered a stop in the suffix (e.g., [didu-bi]), while a fricative in the root triggered a fricative in the suffix ([vozo-vi]). The authors note that this type of consonant harmony is not common (Hansson 2001) but was nevertheless well-suited for constructing nonsense words using English consonants. At training, participants heard six consonants, out of an eight-consonant inventory. At test, they made grammaticality judgments for words that included the two new consonants. Results showed that, while participants learned the harmony pattern for the consonants they had been exposed to, they failed to extend it to the new consonants. Interestingly, a second experiment showed that participants had no difficulty extending a deletion rule to new consonants. Since the deletion rule was not feature-based, however, it is difficult to know whether these divergent results are due to rule type, or to feature generalizations.

A very different possibility is that listeners make stochastic (i.e., randomly determined) generalizations based on patterns in the lexicon. To date, research in this area has focused primarily on variable patterns, in which the relationship between vowels is not stable across words. In Hungarian (Hayes & Londe 2006), for example, there is a set of back-neutral stems, such as [ɔrze:n] 'arsenic', which can take either a back or front vowel suffix (e.g., dative [-nɔk] ~ [-nɛk]). Corpus work indicates that certain generalizations help predict the suffix vowel. These include the *height effect* (front-vowel suffixes occur more often when the neutral vowel is [ɛ], somewhat less when it is [e:], and least when it is [i] or [i:]) and the *count effect* (stems with back + two neutral vowels take front-vowel suffixes more often than stems with back + one neutral vowel). Hayes and Londe (2006) used frequency data to calculate the probability of each stem + suffix variant, and then administered a wug-test in which they presented Hungarian speakers with nonsense stems modelled after [ɔrze:n], and asked them to choose a front or back suffix. Results showed that, for each stem type, the probability of participants' choosing a back vowel closely matched its calculated probability (the correlation was anywhere from 0.82 to 0.90, depending on how it was calculated) – in other words, the wug data largely replicate the stochastic pattern exhibited by the lexicon as a whole. Hayes and Londe 2006) model the wug test data within Stochastic Optimality Theory (Boersma & Hayes 2001), showing how speakers can use lexical information to learn the grammar of their language, even when that grammar is variable (for further discussion on vowel harmony in Optimality Theory, see Chapter 30).

39. L1 influences in learning and processing

In many of the artificial grammar learning studies of harmony discussed in this chapter, the participants are native speakers of English. While the data from these experiments is obviously valuable, their generalizability is nevertheless limited. Even though the participants were exposed to words and phonological rules with which they presumably had no previous experience, we cannot rule out the strong possibility that their L1 phonology impacted their performance in some way. To our knowledge, a relatively small number of studies have explicitly investigated this issue.

LaCross (2015) conducted an extension of Bonatti et al. (2005) (discussed in Section 39.2) with five experiments examining statistical learning of vowel-vowel dependencies. Participants included either American English-speaking or Khalka Mongolian-speaking participants. Importantly, Khalka Mongolian has a rule of harmony whereby vowels must agree in the feature [ATR], as well as labial harmony among non-high vowels. Participants were exposed to a stream of concatenated syllables, where 'words' were characterized by vowel-to-vowel transitional probabilities of 1.0. For example, one set of words was [pemait^{h}ʊ], [gesaixʊ], [pesait^{h}ʊ], [gemaixʊ], where the vowel [e] always predicts that the next vowel will be [ai], which in turn always predicts that the next vowel will be [ʊ]. At test, when asked to choose between a ‘word’ and a ‘partword’, the American English participants performed no better than chance while Khalka Mongolian participants did perform better than chance, although their scores were still somewhat low (e.g., 58% correct in Experiment 1, 55% correct in Experiment 4). Interestingly, the Khalka Mongolian participants performed similarly across experiments in which the vowel dependencies either did or did not obey the rules of Khalka Mongolian ATR harmony, suggesting that their L1 phonological bias was rather general, and consists of more than just a simple “transfer” of L1 properties. It is not clear why the American English-speaking participants in LaCross (2015) failed to learn harmony, while the French-speaking participants in Bonatti et al. (2005) did learn harmony. It is possible that differences in vowel inventories (e.g., French has front rounded vowels) and harmony type may have played a role.

The study by White and colleagues (2018), described in Section 39.3, included participants from five different native languages: Dutch, English, French, German, and Greek. Recall that their overall results indicated a greater preference for local harmony compared to a non-local process, and that this preference was greater in the suffix condition compared to prefix condition. Stress location, which was also manipulated, played no role. When examined individually, the English, French, and German participant groups showed patterns consistent with the overall pattern. For the Dutch group, however, only the interaction effect reached significance – that is, these participants exhibited a greater preference for local harmony only when there was a suffix, as well as local stress. Meanwhile, the Greek group showed no clear preference for locality, regardless of suffixation or stress location. It is not entirely clear why the Greek and Dutch participants behaved differently than the English, French, and German participants. While this study is impressive for including multiple native languages, it should be noted that all of the languages sampled are Indo-European, and all are primarily suffixing languages. Thus, it is possible that even greater differences across language samples might be found with samples from more diverse languages.

In a word segmentation study, Vroomen, Tuomainen, and Gelder (1998) exposed Finnish, Dutch, and French participants to streams of synthesized nonsense syllables which contained ‘words’. In four conditions, the words were constructed such that they either exhibited palatal harmony (e.g., [mymɛty], [vomuvu]) or disharmony (e.g., [mumɛty], [vœmyvu]), and such that they either had equal stress on all syllables or a pitch accent on the first syllable. At test, participants heard both a word or a foil (a string that contained syllables from the exposure stream, but in an order which was not identical to any of the words), and were asked to choose which item was a word. In the no-stress condition, Finnish listeners identified harmonic words better than disharmonic words. In the stress condition, however, they did not, suggesting that stress serves as a more reliable cue to segmentation than harmony in Finnish. Dutch and French participants, on the other hand, showed no difference in responding to harmonic versus disharmonic words in either stress condition.

In another word segmentation study, Kabak, Maniwa, and Kazanina (2010) exposed Turkish and French participants to orthographic target nonsense words, such as *pavo.* Then, in a word-spotting paradigm, participants heard a five-syllable CVCVCVCVCV string such as [golushopavo]. The task was to press ‘yes’ if they heard the target within that string, and ‘no’ if they did not. Results showed that Turkish participants were more accurate when the preceding context was disharmonic (e.g., [gølyshøpavo]) compared to when it was harmonic. No such effect occurred for the French speakers.

In an ERP study, which we review here due to its emphasis on L1 influences, Aaltonen and colleagues (2008) used an oddball paradigm to compare the brain responses of Finnish participants, whose language exhibits front/back harmony, and Estonian participants, whose language has a similar vowel inventory but no harmony. In the isolated condition (where the vowel was presented in isolation, rather than as part of a word), the standard stimulus was [æ] and the deviant was [æ̲], a vowel created so as to be ambiguous between [æ] and [ɑ] in both languages. In the word condition, the standard stimulus was the nonsense word [tækæ] and the deviant was [tækæ̲]. The Mismatched Negativity (MMN) response did not differ significantly between the language groups in the isolated condition, but was significantly stronger for the Finnish participants in the word condition. This suggests that the Finnish speakers reacted to the violation of vowel harmony, while the Estonian speakers, who do not have vowel harmony, did not react. These results show that speakers of a vowel harmony language rapidly process violations of vowel harmony.

Taken together, these studies demonstrate the strong and perhaps unsurprising impact that L1 has on the processing of harmony patterns. Such findings should be borne in mind when interpreting studies with homogenous participant groups.

39.8 Exceptions

In natural languages, many vowel harmony patterns contain exceptions. In Hungarian, for example, the temporal suffix /-kor/ exceptionally does not alternate based on stem vowel: e.g. [øt-kor] 'at five (o'clock)' and not *[øt-kør] (Kenesei, Vago & Fenyvesi 2002). These facts raise questions about how learners cope with exceptionality and variance, an issue that has been explored in a couple of studies. Finley (2015) trained American English-speaking participants on a back/round vowel harmony pattern in which one suffix alternated between [-me] ~ [-mo] and another suffix, [-go], did not alternate. Results showed that participants learned the behavior of both alternating and the non-alternating affix, suggesting that learning exceptions is relatively straightforward, at least when they are categorical (i.e., one morpheme alternates, and another does not). Interestingly, a separate analysis of the test items showed that although both front vowel and back vowel stems differed significantly from chance, participants were more likely to choose the [-go] when the stem vowels were back. This suggests a bias for harmony even in non-alternating affixes (other authors working in different paradigms have also found a non-alternation bias; Coetzee 2009; Stave, Smolek & Kapatsinski 2013; White 2014). These results could potentially shed light on the evolution of vowel harmony; if learners prefer harmony even in a non-alternation context, this could, over time, drive harmony to be more frequent across items.

In another study of exceptional patterns, Baer-Henney, Kügler, and van de Vijver (2015) trained German-speaking participants on a front-back vowel harmony pattern. In one condition, 85% percent of the exposure items were harmonic and 15% were disharmonic; and in another condition, 65% were harmonic and 35% were disharmonic. Results showed that participants in the 85% condition were more likely to generalize the harmony pattern to novel items. This difference persisted even in a second experiment, where participants were provided with increased exposure to training items. These results suggest that the ease of learnability of vowel harmony patterns can extend to patterns that apply non-categorically, but that a strong majority of items must follow harmony in order for this ease of learning to persist.

39.9 Brain responses to vowel harmony

As reviewed in the sections above, many studies have used behavioral experiments to investigate the representation of vowel harmony in listener's minds. Meanwhile, comparatively few studies have used neurological measurements to investigate physical brain responses. The research that has been done thus far, which includes studies of people with aphasia as well as studies of healthy people, suggests that the brain encodes vowel harmony quite robustly.

39.9.1 Studies of people with aphasia

Studies of people with aphasia suggest that vowel harmony patterns remain intact, even when there is clear disruption in other language domains. MacWhinney and Osmán-Sági (1991) conducted a speech production study with fourteen Hungarian speakers with aphasia. Nine of their participants had damage to Broca's area (in the frontal lobe of the cortex), which is associated with speech production, and five participants had damage to Wernicke's area (in the temporal lobe of the cortex), which is associated with speech comprehension. The researchers were primarily interested in case-marking, and therefore elicited sentences in which target nouns occupied various positions, such as direct object, indirect object, and so on. Their results revealed several patterns of incorrect or missing case-marking but, interestingly, there was not a single violation of the basic rules of vowel harmony, for either type of aphasia. Slobin (1991) reported similar results for Turkish speakers. He conducted a production study with ten people with Broca's aphasia and seven people with Wernicke's aphasia. His findings showed that, despite interesting patterns of errors and omissions in other domains, errors of vowel harmony did not occur. These results suggest that vowel harmony may be robust to some of the neurological damage that affects language.

39.9.2 EEG and MEG studies of healthy people

Studies of healthy people have used electrophysiological measurements, typically electroencephalography (EEG) or magnetoencephalography (MEG), to record brain activity. With EEG, electrodes attached to multiple locations on a participant's scalp record electrical activity in the brain. With MEG, sensitive magnetometers measure the magnetic fields produced by electrical currents in the brain. Researchers working with speech stimuli often analyze a specific type of EEG/MEG output called event-related potentials (ERPs), which are averaged responses time-lock to a specific sensory event. A common paradigm in ERP research is the "oddball" paradigm, in which a standard stimulus is repeated many times in sequence, interspersed occasionally with a deviant stimulus. For example, in the oddball sequence [dɑ, dɑ, dɑ, dɑ, dɑ, dɑ, tɑ, dɑ, dɑ], [dɑ] is the standard and [tɑ] is the deviant. The oddball paradigm can be used to elicit several different ERP 'components', which are reliable changes in the neural activity at a specific point in time following stimulus presentation. Mis-matched negativity, or MMN, is a negative deflection in electrical activity that occurs approximately 150 to 250 ms. after the onset of the oddball (e.g., [ta]). Another important ERP component that can be elicited with the oddball paradigm is known as P300 (or P3). This is a positive component that is usually occurs around 250-300 ms after the onset of the oddball (and is typically the third positive waveform). The P3 is often elicited in attention shifts and decision making (Luck 2014).

Ylinen and colleagues (2016) used EEG to investigate the brain responses of vowel harmony in Finnish speakers. They used an oddball paradigm with two conditions. In the harmony condition, the standard stimulus was the pseudoword [ɑkɑ], and the deviant stimuli were either [ɑko] (which is consistent with the rules of Finnish front-back harmony) and *[ɑkø] (which violates harmony). In the neutral condition, the standard stimulus was [ikɑ], and the deviants were [iko] and [ikø]. Since [i] behaves as a neutral vowel, both of these deviant stimuli are consistent with the rules of Finnish front-back harmony. The key comparison concerned responses to deviant [ikø] versus deviant *[ɑkø]. Crucially, the vowel that elicits the response for both of these deviants is the same (i.e., [ø]); only its preceding phonological context differs. Results showed a significantly larger MMN response for *[ɑkø] compared to [ikø]. Unlike other stimuli, the deviant *[ɑkø] also elicited a P3a component (referred to as the 'novelty P3' reflecting a shift in attention for a novel response), which may indicate a shift of attention to the unpredicted word form. The authors argue that this pattern of brain responses reflects their Finnish participants' implicit knowledge of vowel harmony patterns.

Scharinger, Idsardi, and Poe (2011) used MEG to investigate the brain responses of Turkish participants. They used the oddball paradigm with two different pairs of isolated vowels, each of which were presented in contrasting orders of standard-deviant: [ɯ, œ], [œ, ɯ], [ɛ, œ], [œ, ɛ]. The authors were particularly interested in MMN responses to acoustic versus phonological differences. They state that the acoustic difference between [ɯ] and [œ] is small, whereas the acoustic difference between [ɛ] and [œ] is large. However, since MMN is presumably sensitive to phonological contrast and not acoustic contrast, they predicted that disharmony rather than acoustic difference should mediate the size of the MMN. Furthermore, since the Turkish harmony rule does not spread rounding features onto low vowels, they predicted that sequences of [ɛ, ɛ, ɛ, œ] should produce a larger MMN than sequences of [œ, œ, œ, ɯ]. Results show that both predictions are confirmed. Magnitude of MMN responses for [ɯ]-[œ] and [œ]-[ɯ] sequences were similar (despite differences in acoustics), and the deviant in the sequence [ɛ]-[œ] elicited a larger MMN amplitude than the other sequences. The authors argue that these results support a view of harmony as an abstract phonological process, based on phonological features, rather than a comparison of acoustic stimuli. They also argue that there are no ‘graded’ forms of disharmony (e.g., that acoustically distinct disharmonic forms are ‘more’ disharmonic than acoustically similar disharmonic pairs). In interpreting these results, one caveat is that the authors characterized “acoustic difference” as the Euclidean distance between vowels on the basis of the first three formants. However, listeners generally tend to be more sensitive to differences in lower frequencies than higher frequencies (e.g., Sek & Moore 1995), a factor which could level out the differences between [ɯ]-[œ] and [œ]-[ɯ] for reasons unrelated to phonology.

Miglietta, Grimaldi, and Calabrese (2013) used EEG to investigate brain responses from speakers of the Tricase dialect of Italian. In this language, a metaphony process occurs, such that stressed low-mid front vowels raise to high-mid when they occur before a front vowel: [ˈdɛnte] ‘tooth’, [ˈdenti] ‘teeth’. The authors compared MMN responses to the allophones [ɛ, e] versus the phonemes [e, i]. Results showed that the amplitude of the MMN response was the same in both conditions. The latency of the MMN response, however, was significantly earlier in the phonemic condition. The authors suggest that listeners perform a single neural computation which is sensitive to the contrastive versus allophonic status of isolated vowels, as determined by the listeners’ L1 grammar.

The studies just discussed examined native speakers of harmony languages, but studies have also looked at vowel harmony in second language learning. McLaughlin and colleagues

(2010) compared native Finnish speakers to English-speaking adults who were in a Finnish language instruction program at various stages in their first year of instruction (recorded at 3, 6, and 8 months in). Participants performed a lexical decision task while EEGs were recorded. Items were real words, legal pseudowords, and harmony-violating nonwords. Native Finnish speakers showed a P600 (also referred to as a 'late positive component') response to the harmony-violating nonwords, and an N400 response to pseudowords. The P600 response is typically associated with grammatical violations, while the N400 is typically associated with semantic anomalies. The native English speakers showed an N400 response to all pseudowords in the first session, but participants shifted to showing a P600 for harmony-violating words by the final session. These results suggest that as L2 learners gain more knowledge of a harmony language, their brain responses to harmony violations look similar to that of native speakers.

Avcu, Rhodes, and Hestvik (2019) trained naïve American-English speakers on a sibilant harmony pattern. During exposure, they heard and repeated 20 CVCV nonsense words, such as [sasi] and [ʃeʃu], that conformed to sibilant harmony. At test, participants' EEGs were recorded while performing a grammaticality judgment task. This task made use of the oddball paradigm, where harmony-conforming words served as the standards, and non-conforming words such as [ʃasi] and [ʃesu] served as the deviants. Results for the behavioral task showed that ungrammatical words were detected with a mean d' sensitivity of 0.557, which is significantly different from zero (d' is a measure of how sensitive people are to the difference between two signals). Results from the electrophysiological measurements, which focused on two ERP components called the auditory evoked potential and Bereitschafts potential, suggested that participants' brain detected deviant stimuli exactly at the point in time when harmony was violated (e.g., during the [s] in [ʃa<u>s</u>i]). These measurements therefore provide information about the time course of responses to disharmonic words, which cannot be inferred solely on the basis of behavioral responses.

All of the ERP studies discussed here address questions about phonological processing and the time course of processing. While these papers report findings from several different electrode sites, they generally report from the central midline (Cz) and frontal midline (Fz) locations of the brain, suggesting the possibility that these are general areas that elicit the responses in question. As noted above, while techniques with better spatial resolution (e.g., MEG as discussed in the current section or, eventually, fMRI) could provide more insight into the localization of phonological patterns such a vowel harmony, it is likely that several neurological pathways are implicated in the production and processing of vowel harmony.

39.10 Summary and Conclusions

This chapter discussed several experimental studies addressing vowel harmony in the mind and brain. While there are many studies looking at how vowel harmony might be represented and processed in the mind, very few studies addressed the brain. The two studies that examined patients with aphasia showed no deficit to vowel harmony, and the neurological studies with healthy participants addressed the timing and representation of vowel harmony.

This chapter discussed a large and growing body of literature that makes use of artificial grammar learning paradigms in vowel harmony. Artificial grammar learning paradigms can be used to address a wide range of questions related to phonological processes, and the fact that participants can learn vowel harmony patterns after a brief exposure period makes the paradigm attractive. However, the use of convenience samples in experimental work on vowel harmony raises important questions about the generalizability of the findings. Aside from a few exceptions, work on vowel harmony in learnability makes use of speakers of European languages, and work addressing native vowel harmony processing focuses on a few languages (e.g., Finnish, Hungarian, and Turkish). Thus, more research is needed from a wider language sample to better understand how vowel harmony is processed in the mind and brain.

Despite these limitations, the research on vowel harmony in naïve participants as well as native speakers strongly suggests that vowel harmony is an abstract process, governed by grammatical rules. It also suggests that speakers possess biases for particular patterns, which may ultimately underpin the cross-linguistic typology of vowel harmony.

Acknowledgments

We thank Amanda Rysling and John Kingston, an anonymous reviewer, and editors Nancy Ritter and Harry van der Hulst for helpful comments that improved this chapter.

References

Aaltonen, Olli, Åke Hellström, Maija S. Peltola, Janne Savela, Henna Tamminen & Heidi Lehtola. 2008. Brain responses reveal hardwired detection of native-language rule violations. *Neuroscience Letters* 444(1). 56–59. DOI: https://doi.org/10.1016/j.neulet.2008.07.095

Avcu, Enes. 2018. Experimental investigation of the Subregular Hypothesis. In Wm. G. Bennet, Lindsay Hracs, & Dennis Ryan Storoshenko (eds.), *Proceedings of the 35th West Coast Conference on Formal Linguistics*, 77–86. Somerville, MA: Cascadilla Proceedings Project.

Avcu, Enes, Ryan Rhodes & Arild Hestvik. 2019. Neural Underpinnings of Phonotactic Rule Learning. In *Proceedings of the Annual Meetings on Phonology*, Vol. 7.

Baer-Henney, Dinah, Frank Kügler & Ruben van de Vijver. 2015. The Interaction of Language-Specific and Universal Factors During the Acquisition of Morphophonemic Alternations With Exceptions. *Cognitive Science* 39(7). 1537–1569. DOI: https://doi.org/10.1111/cogs.12209

Baković, Eric. 2000. *Harmony, dominance and control*. PhD Dissertation Rutgers University dissertation.

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14(2–3). 150–177.

Boersma, Paul & Bruce Hayes. 2001. Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32(1). 45–86. DOI: https://doi.org/10.1162/002438901554586

Bonatti, Luca L., Marcela Peña, Marina Nespor & Jacques Mehler. 2005. Linguistic Constraints on Statistical Computations: The Role of Consonants and Vowels in Continuous Speech Processing. *Psychological Science* 16(6). 451–459. DOI: https://doi.org/10.1111/j.0956-7976.2005.01556.x

Campbell, Lyle. 1986. Testing phonology in the field. *Experimental Phonology* 163–186.

Coetzee, Andries W. 2009. Learning lexical indexation. *Phonology* 26(1). 109–145. DOI: https://doi.org/10.1017/S0952675709001730

Duncan, Liisa. 2015. *Productivity of Finnish Vowel Harmony: Experimental Evidence*. ProQuest Dissertations Publishing dissertation. Retrieved from http://search.proquest.com/docview/1762524990/?pq-origsite=primo

Finley, Sara. 2011. Generalization to novel consonants in artificial grammar learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 33.

Finley, Sara. 2012a. Learning unattested languages. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 34.

Finley, Sara. 2012b. Typological Asymmetries in Round Vowel Harmony: Support from Artificial Grammar Learning. *Language and Cognitive Processes* 27(10). 1550–1562. DOI: https://doi.org/10.1080/01690965.2012.660168

Finley, Sara. 2015. Learning Exceptions in Phonological Alternations. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, Vol. 37, 698–703.

Finley, Sara & William Badecker. 2008. Analytic biases for vowel harmony languages. In *WCCFL*, Vol. 27, 168–176.

Finley, Sara & William Badecker. 2009a. Artificial language learning and feature-based generalization. *Journal of Memory and Language* 61(3). 423–437. DOI: https://doi.org/10.1016/j.jml.2009.05.002

Finley, Sara & William Badecker. 2009b. Right-to-left biases for vowel harmony: Evidence from artificial grammar. In *Proceedings of the 38th North East linguistic society annual meeting*, Vol. 1, 269–282. University of Massachusetts.

Finley, Sara & William Badecker. 2010. Linguistic and non-linguistic influences on learning biases for vowel harmony. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 32.

Finley, Sara & William Badecker. 2012. Learning biases for vowel height harmony. *Journal of Cognitive Science* 13(3). 287–327.

Gósy, Mária. 1989. Vowel harmony: Interrelations of speech production, speech perception, and the phonological rules. In *Acta Linguistica Hungarica*, Vol. 39, 93–118. Budapest: Akadémiai Kiadó.

Hansson, Gunnar Ólafur. 2001. *Theoretical and typological issues in consonant harmony*. University of California, Berkeley dissertation. Retrieved from http://faculty.arts.ubc.ca/gohansson/pdf/GH_diss.pdf

Hayes, Bruce & Zsuzsa Cziráky Londe. 2006. Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology* 23(1). 59–104. DOI: https://doi.org/10.1017/S0952675706000765

Hickok, Gregory & David Poeppel. 2007. The cortical organization of speech processing. *Nature Reviews Neuroscience* 8(5). 393–402. DOI: https://doi.org/10.1038/nrn2113

Hyman, Larry M. 2002. Is there a right-to-left bias in vowel harmony. In *9th International Phonology Meeting, Vienna*, Vol. 1. Retrieved from http://www.linguistics.berkeley.edu/~hyman/Hyman_Vienna_VH_paper_forma.pdf

Kabak, Barış, Kazumi Maniwa & Nina Kazanina. 2010. Listeners use vowel harmony and word-final stress to spot nonsense words: A study of Turkish and French. *Laboratory Phonology* 1(1). 207–224. DOI: https://doi.org/10.1515/labphon.2010.010

Kaun, Abigail. 1995. The Typology of Rounding Harmony: An Optimality Theoretic Approach. DOI: https://doi.org/10.7282/T3R49PM2

Kaun, Abigail. 2004. The typology of rounding harmony. *Phonetically Based Phonology* 87416.

Kenesei, István, Robert M. Vago & Anna Fenyvesi. 2002. *Hungarian*. Routledge.

Kimper, Wendell. 2016. Asymmetric Generalisation of Harmony Triggers. In *Proceedings of the Annual Meetings on Phonology*, Vol. 3.

LaCross, Amy. 2015. Khalkha Mongolian speakers' vowel bias: L1 influences on the acquisition of non-adjacent vocalic dependencies. *Language, Cognition and Neuroscience* 30(9). 1033–1047. DOI: https://doi.org/10.1080/23273798.2014.915976

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

Linebaugh, Gary Dean. 2007. *Phonetic grounding and phonology: Vowel backness harmony and vowel height harmony*. University of Illinois at Urbana-Champaign dissertation.

Luck, Steven J. 2014. *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT press.

MacWhinney, Brian & Judit Osmán-Sági. 1991. Inflectional marking in Hungarian aphasics. *Brain and Language* 41(2). 165–183. DOI: https://doi.org/10.1016/0093-934X(91)90151-P

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

Martin, Alexander & James White. 2021. Vowel Harmony and Disharmony Are Not Equivalent in Learning. *Linguistic Inquiry* 52(1). 227–239. DOI: https://doi.org/10.1162/ling_a_00375

Miglietta, Sandra, Mirko Grimaldi & Andrea Calabrese. 2013. Conditioned allophony in speech perception: An ERP study. *Brain and Language* 126(3). 285–290. DOI: https://doi.org/10.1016/j.bandl.2013.06.001

Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25(1). 83–127. DOI: https://doi.org/10.1017/S0952675708001413

Newport, Elissa L. & Richard N. Aslin. 2004. Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology* 48(2). 127–162. DOI: https://doi.org/10.1016/S0010-0285(03)00128-2

Pycha, Anne, Pawel Nowak, Eurie Shin & Ryan Shosted. 2003. Phonological rule-learning and its implications for a theory of vowel harmony. In *Proceedings of the 22nd west coast conference on formal linguistics*, Vol. 22, 101–114. Cascadilla Press Somerville, MA.

Recasens, Daniel & Maria Dolors Pallarès. 2000. A Study of F1 Coarticulation in VCV Sequences. *Journal of Speech, Language, and Hearing Research* 43(2). 501–512. DOI: https://doi.org/10.1044/jslhr.4302.501

Saffran, Jenny R., Elissa L. Newport & Richard N. Aslin. 1996. Word segmentation: The role of distributional cues. *Journal of Memory and Language* 35(4). 606–621. DOI: https://doi.org/10.1006/jmla.1996.0032

Scharinger, Mathias, William J. Idsardi & Samantha Poe. 2011. Neuromagnetic reflections of harmony and constraint violations in Turkish. *Laboratory Phonology* 2(1). 99–123. DOI: https://doi.org/10.1515/labphon.2011.003

Sek, Aleksander & Brian C.J. Moore. 1995. Frequency discrimination as a function of frequency, measured in several ways. *The Journal of the Acoustical Society of America* 97(4). 2479–2486. DOI: https://doi.org/10.1121/1.411968

Skoruppa, Katrin & Sharon Peperkamp. 2011. Adaptation to Novel Accents: Feature-Based Learning of Context-Sensitive Phonological Regularities. *Cognitive Science* 35(2). 348–366. DOI: https://doi.org/10.1111/j.1551-6709.2010.01152.x

Slobin, Dan I. 1991. Aphasia in Turkish: Speech production in Broca's and Wernicke's patients. *Brain and Language* 41(2). 149–164. DOI: https://doi.org/10.1016/0093-934X(91)90150-Y

Stave, Matthew, Amy Smolek & Vsevolod Kapatsinski. 2013. Inductive bias against stem changes as perseveration: Experimental evidence for an articulatory approach to output-output faithfulness. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 35.

Suomi, Kari. 1983. Palatal Vowel Harmony: A Perceptually Motivated Phenomenon? *Nordic Journal of Linguistics* 6(1). 1–35. DOI: https://doi.org/10.1017/S0332586500000949

Suomi, Kari, James M. McQueen & Anne Cutler. 1997. Vowel Harmony and Speech Segmentation in Finnish. *Journal of Memory and Language* 36(3). 422–444. DOI: https://doi.org/10.1006/jmla.1996.2495

Terbeek, Dale. 1977. A cross-language multidimensional scaling study of vowel perception. *UCLA Working Papers in Linguistics* 37. 1–271.

Universals of language. 1963. Oxford, England: M.I.T. Press.

Vroomen, Jean, Jyrki Tuomainen & Beatrice de Gelder. 1998. The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language* 38(2). 133–149.

White, James. 2014. Evidence for a learning bias against saltatory phonological alternations. *Cognition* 130(1). 96–115. DOI: https://doi.org/10.1016/j.cognition.2013.09.008

White James, Kager René, Linzen Tal, Markopoulos Giorgos, Martin Alexander, Nevins Andrew, … van de Vijver Ruben. 2018. Preference for locality is affected by the prefix/suffix asymmetry: Evidence from artificial language learning. In Hucklebridge Sherry & Nelson Max (eds.), *NELS 48: Proceedings of the Forty-Eighth Annual Meeting of the North East Linguistic Society : Volume 3*, 207–220. Amherst, MA, USA: Graduate Linguistics Student Association. Retrieved from http://real.mtak.hu/88495/

Yavaş, Mehmet. 1980. Some pilot experiments on Turkish vowel harmony. *Paper in Linguistics* 13(3). 543–562. DOI: https://doi.org/10.1080/08351818009370510

Ylinen, Sari, Milla Huuskonen, Katri Mikkola, Emma Saure, Tara Sinkkonen & Petri Paavilainen. 2016. Predictive coding of phonological rules in auditory cortex: A mismatch negativity study. *Brain and Language* 162. 72–80. DOI: https://doi.org/10.1016/j.bandl.2016.08.007

Zimmer, Karl E. 1969. Psychological Correlates of Some Turkish Morpheme Structure Conditions. *Language* 45(2). 309–321. DOI: https://doi.org/10.2307/411662