Generalization Beyond Similarity: Support for Abstract Phonology[*]

Sara Finley
Pacific Lutheran University

1 Background

A growing debate within the cognitive science of language concerns how to create a psychologically plausible theory of phonological representations. A divide exists as to whether abstract approaches or more gradient, exemplar based approaches provide a better account of speakers' knowledge, intuitions and language use (Port and Leary 2005). The present paper presents evidence from an artificial grammar learning experiment that supports the abstract approach to phonology and phonological representations.

Traditional generative approaches to phonology have emphasized abstract symbol manipulation (Chomsky and Halle 1968; Prince and Smolensky 2004) in which an underlying form becomes a surface form as a result of structural changes. In a rule-based approach, abstract symbol manipulation creates a change between a potentially arbitrary underlying form and a surface form (e.g., 'A' becomes 'B' in the context of 'C', /A/ → [B] _C). In Optimality Theory (OT), a language-specific balance (determined by constraint rankings) between minimizing structural change and minimizing markedness violations determines the mapping between the input and the output. The language-specific constraints that determine the optimal mapping tend to be relatively abstract, referring to abstract concepts such as stress, articulatory features and syllable structure.

These traditional generative formalizations of input-output mappings in phonology are typically based in a formalism that distinguishes between competence and performance (Chomsky 1965; Hale and Reiss 1998). Competence-based grammars tend to keep performance-based factors, such as rate of speech, outside of grammatical influence. Recently, generative analyses have begun to embrace more gradient notions of markedness and abstraction, potentially blurring the line between competence and performance. These innovations include the introduction of constraints based in phonetic grounding (Zuraw 2007; Wilson 2001) that can account for fine-grained phonetic details, including variability at a sub-phonemic level (Flemming 1995) and articulatory gestures (Gafos 2002). Additions to traditional generative models have also involved formalized accounts of morphological variation, including analyses of morpheme-specific effects (Wolf 2005; Finley 2009), phonological exceptions (Finley 2010; Mahanta 2012), and lexical variation (Coetzee and Pater 2011; Coetzee and Kawahara 2013; Zuraw 2010). These adaptations of Optimality Theory are required in order to account for the full range of phonological patterns that exist in the world's languages, including variations within phonological patterns. Such variations cannot be accounted for without a flexible notion of markedness and grammaticality. While these relatively recent changes in formal phonological theory show promise for moving the field towards a better understanding of the mental representation of phonological grammars, they raise the possibility that generative models, even with modifications, may be unable to account for phonological knowledge as it exists in the minds of language users. If this is true, a model of phonology, such as the models proposed in exemplar theory, grounded in performance factors rather than abstract symbol manipulation, may be better able to capture true phonological knowledge (Port and Leary 2005).

Exemplar models of phonology focus primarily on these so-called extra-grammatical properties, specifically talker-specific and lexical frequency effects. It has been shown that language users remember

highly detailed phonetic properties of the speakers they encounter, and use those details when making decisions about speech (Goldinger 1996; Nygaard 1998). In addition, the application of phonological rules is strongly influenced by lexical frequency (Bybee 2002; Bybee 2012) and social context (Drager, Hay, and Walker 2010). These factors are easily accounted for in an exemplar model of phonology. Exemplar models propose that speakers store all instances of every utterance they speak and hear, including the talker, the social context, and other idiosyncratic phonetic details. In order to determine grammaticality of a novel utterance, the speaker will compare a novel word (or novel production of a familiar word) to known exemplars. The acceptability of this novel form is based on similarity; the more similar a novel word is to known exemplars, the more acceptable it is (Johnson, 1997). The exemplar approach to phonology does not rely on abstract, categorical symbol manipulation, but gradient speech processing, allowing for a variety of lexical and talker-specific effects.

Proponents of exemplar models imply that extra-grammatical properties of speech (such as the gradient variations found in articulatory and perceptual phonetics) that are often ignored by generative linguists, are vital for understanding how phonological patterns are represented in the minds of speakers (Port and Leary 2005). In addition, there is an implication that exemplar models are able to account for the same phenomena that are addressed in formal, generative models of phonology. If this were the case, an exemplar model would make a generative model unnecessary, because the relevant factors in an exemplar model can explain the same facts as a generative model, in addition to several findings that cannot be explained by traditional generative models.

While exemplar models may be able to account for a variety of effects that traditional generative models cannot account for, the recent adaptations applied to OT imply that the factors that were once believed to be extra-grammatical may be incorporated into a generative model as part of the grammar. The question that remains to be understood is whether a theory based on abstract symbol manipulation or a theory based in exemplars better captures the cognitive representation of phonological knowledge. In exemplar models, abstraction is epiphenomenal, while in a generative model such as OT, phonological patterns are inherently abstract, and can potentially interact with other, less abstract levels of representation (Smolensky and Legendre 2005). Because the goal of phonological models is primarily to understand how language users represent the patterns of sounds in the languages they speak, being able to differentiate between these two very different approaches to phonological representations is extremely important.

There are several issues that arise when differentiating between exemplar and generative models of phonology. First, while both generative and exemplar models of phonology may be able to account for the same set of phenomena, the focus of these models is different. Generative models tend to focus on categorical patterns, while exemplar models tend to focus on gradient patterns, making the distinction between the two appear to be bigger than it may actually be. Second, the predictions between abstract, generative models and exemplar models of phonology tend to be very similar, particularly when exploring adult speakers' language usage for known words, making it relatively difficult to compare the theories using words from the lexicon. While it is possible to compare variation across speakers (Cohen-Goldberg 2015) and responses to nonce forms (Gouskova and Becker 2013), another method of theory comparison is to use an artificial grammar learning paradigm to test each theory's predictions for how speakers respond to novel forms in a newly learned phonological system (Cristia et al. 2013; Finley and Badecker 2009a; Finley 2013). In an artificial grammar learning paradigm, the language can be specifically manipulated so that talker effects, frequency and similarity can all be controlled for. In addition, test questions can be manipulated to directly ask learners how they have represented the novel language.

In the present study, English speaking participants were exposed to a novel vowel harmony pattern. Vowel harmony is a phonological process whereby all vowels in a word must share the same phonological feature. Typically, vowel harmony results in morphophonological alternations in which a particular morpheme will have multiple allomorphs that vary depending on the vowel features of the stem (Rose and Walker 2011). Previous artificial grammar learning experiments have shown that adult English speaking language users can easily learn a novel vowel harmony pattern (Pycha et al. 2003) and generalize that novel pattern to novel vowels (Finley and Badecker 2009a), novel affixes (Finley and Badecker 2009b) and novel talkers (Finley 2013). Finley and Badecker (2009a) showed that learners who were trained on a novel back/round vowel harmony pattern that contained only mid and low vowels generalized that harmony pattern to high vowels, even though high vowels did not appear in the training set. Finley (2013) showed that language users will generalize a novel harmony pattern to novel talkers, even when the talkers have a

different native language, and are of a different gender. In addition, Finley and Badecker 2009b showed that learners of a back/round vowel harmony pattern could generalize a suffixation pattern to a prefixation pattern, even when prefixes never appeared in the training set. The ability to generalize far beyond the training stimuli in these experiments may provide evidence against an exemplar model because learners extend patterns beyond what was heard during training. However, it is possible that an exemplar model could account for the data if learners are using some abstract notion of similarity. If the novel affixes are similar enough to the familiar items, then participants may use an exemplar-based strategy to accept a novel item. For example, if a novel grammatical item is more similar to a familiar grammatical item than a corresponding novel ungrammatical item, then the exemplar model will choose the grammatical item over the ungrammatical item, even though both are unfamiliar.

In order to better understand whether learners base grammatically judgments on similarity (as predicted by exemplar models), participants in the present study perform both a two-alternative forced choice grammaticality judgment task and a similarity judgment task. If grammaticality judgments are guided by the same principles as similarity judgments, as predicted by an exemplar model, then learners will respond similarly to two-alternative forced choice grammaticality judgments and similarity judgments. If grammaticality judgments are guided by abstract phonological principles, as predicted by abstract phonological models, then learners will respond differently to grammaticality judgments and similarity judgments. Specifically, the exemplar theory predicts that grammaticality judgments should be mediated by similarity; items judged as similar to the training items will be more likely to be judged as grammatical, while abstract phonology theory predicts that grammaticality and similarity will be independent; learners will select items as grammatical even if they see these items as highly dissimilar.

3   Method

The present study is a replication with extension of Finley and Badecker (2009b) in which English-speaking participants were exposed to a novel artificial back/round vowel harmony pattern. Participants were exposed to a language in which a suffix alternated between [-mi] and [-mu] in accordance to vowel harmony, and were then tested on novel items, some of which contained prefixes ([gi-]/[gu-]) instead of suffixes. If participants generalize to the novel prefixes, it suggests that learners formed a general, bi-directional vowel harmony pattern that applies to both prefixes and suffixes.

3.1   *Participants*   Thirteen speakers, fluent in American English (with no knowledge of vowel harmony), participated for course credit at a small liberal arts college in the Pacific Northwest of the United States of America. No participant had any exposure to a natural or an artificial vowel harmony language.

3.2   *Exposure Phase*   Participants were exposed to pseudo-morphological alternations with [CVCV] 'stems' that were followed by [CVCV-mi/-mu] affixed forms. The suffix [-mu] followed back vowel stems that contained [u, o], while the [-mi] suffix followed front vowel stems that contained [i, e].

(1) Examples of Training Stimuli

Front Vowel Stems	Back Vowel Stems
kene kenemi	kogo kogomu
midi midimi	mobo mobomu
nege negemi	goto gotomu
tepe tepemi	bono bonomu
biki bikimi	punu punumu
dimi dimimi	tudu tudumu
gebe gebemi	nupu nupumu

The stem vowels always contained two identical vowels (e.g., [ee], [uu]), and were drawn from the set [i, e, u, o]. The stem vowels were always identical in order to control for the generalization to novel affixes at

test, so that the vowel that occurred before the suffix would be the same vowel that occurred after the novel prefix for all items. This made it possible to rule out the possibility that the learners would reject the novel prefix simply because the initial vowel was different from the final vowel. While the vowels in the stems were identical, the suffix vowels were only identical to the stem vowels on half of the exposure trials (e.g., [ketemi] and [midimi]). The consonants were of the drawn from the set [p, t, k, b, d, g, m, n]. There were 24 sets of stem+suffix pairs repeated five times each in a randomized order. Examples of training stimuli can be found in the table in (1) above.

3.3 *Test Phase*   Participants responded to two different sets of test items, each presented to all participants with the order of test items counterbalanced. However, due to a randomization error, only two of the 13 participants received the two-alternative forced choice items first, while the other 11 participants received the distance judgment task first. The two-alternative forced choice test used the identical items from Finley and Badecker (2009b). Each item compared a grammatical (harmonic) item to an ungrammatical (disharmonic) item. The stems for each alternative were identical, with the difference being the suffix or prefix vowel (e.g., [mi] vs. [mu]). There were three types of test items, with 12 items for each of the three types of items: Old items, New Stem items and Prefix Affix items. Old items contained the same forms that appeared in the training set. Selecting the harmonic alternative for Old items means that the participant is able to recognize the item heard in the training set. New Stem items did not appear in the training set, but contained the same suffix that participants heard in training, with novel stems. Selecting the harmonic alternative for New Stem items means that the participant is able to generalize the harmony pattern to novel items not heard at training. Prefix items contained the same stems heard in training, but with a prefix attached to the stem. These items differed based on the prefix which was either [gi-] or [gu-]. Selecting the harmonic alternative for Prefix items means that the participant is able to generalize the harmony pattern to a novel affix, from suffix to prefix.

(2)   Examples of Test Stimuli: *Ungrammatical Items Counterbalanced Across Items

Old Items	New Stem Items	Prefix Items
bememi vs. *bememu	bigimi vs. *bigimu	gibiki vs. *gubiki
digimi vs. *digimu	gedemi vs. *gedemu	gidimi vs. *gudimi
gomomu vs. *gomomi	gobomu vs. *gobom	gugoto vs. *gigoto
nupumu vs. *nupumi	mutumu vs. *mutumi	gupunu vs. *gipunu
tudumu vs. *tudumi	nukumu vs. *nukumi	gutunu vs. *gitunu
gebemi vs. *gebemu	pekemi vs. *pekemu	gitepe vs. *gutepe

The Distance Judgment test items required participants to rate the similarity of each alternative from the two-alternative forced choice test to the similarity of the words that were heard during the exposure phase. The scale ranged from 1 to 5, where 1 indicated that the word was identical to an item in the exposure phase, and 5 indicated that the item was extremely different from the words in the exposure phase.

3.4 *Stimuli*   The stimuli used in the present experiment were identical to the stimuli used in Finley and Badecker (2009b). An adult, male speaker of American English who was trained in phonetics, but was unaware of the hypothesis of the present experiment served as the talker for the stimuli in the present experiement. The speaker produced all items in list format, repeating each item four times. One item was taken from the second or third item in the list in order to keep prosody of the item the same (avoiding differences that occur for first and last items in the list). Stem items were not modified, except to control for intensity (normalized average of 70dB for each item). Suffixed items were recorded as CVCVma (e.g., bemema) and cross-spliced spliced with a recording of babami/babamu. This ensured that the stem portion of each suffixed item was identical for the harmonic and disharmonic items in the test. Prefixed items were created in the same fashion (e.g., gaCVCV).

4   Results

4.1 *Two Alternative Forced Choice*   Means and standard errors of proportion of harmonic responses are

presented in the figure in (3). The data were analyzed using a generalized linear mixed effects logistic regression fit by the Laplace approximation using the lme4 package in R (R Development Core Team 2011) with test condition as the fixed factor. The alpha for significance was p <0.05, based on z scores (z > 1.96). The model was created with the maximal set of random slopes and intercepts for subjects and items that would successfully converge (Barr et al. 2013), which included random intercepts for subjects and items, but models with random slopes failed to converge. Each test condition was compared to 50% chance by setting the reference point of the model to the test condition in question. Harmonic responses Old items (Mean= 0.63, SD = 0.16) were marginally above chance (β= 0.66, *SE* = 0.38, z = 1.71, p = 0.087). Harmonic responses to New Stem items (Mean = 0.66, SD = 0.17) were significantly above chance (β= 0.84, *SE* = 0.38, z = 2.10, p = 0.029). Harmonic responses to Prefix items (Mean = 0.68, SD = 0.13) were significantly above chance (β= 0.88, *SE* = 0.39, z = 2.24, p = 0.025). There were no significant differences between Old and New Stem items (β= 0.18, *SE* = 0.47, z = 0.38, p = 0.70), Old and Prefix items (β= 0.22, *SE* = 0.49, z = 0.44, p = 0.65), or between New Stem and Prefix items (β= 0.037, *SE* = 0.48, z = 0.076, p = 0.94).

(3) Two Alternative Forced Choice Results (Means and Standard Errors)
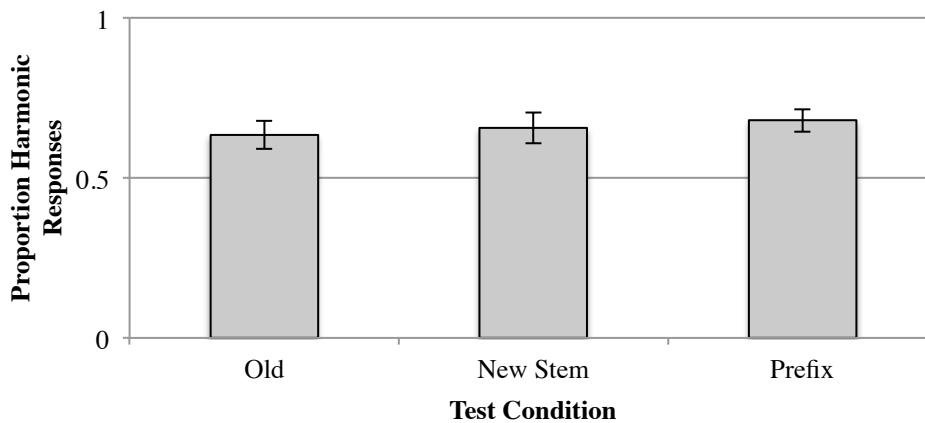
While the Old items did not reach statistical significance, it is likely that this is due to statistical power, as there were only 13 participants in the entire experiment, and there were no differences between New Stem and Old or Prefix and Old items. One factor that may have lead to lower rates of selection of the harmonic items in Old items (and across the experiment, as compared to Finley and Badecker 2009b) is that the majority of participants responded to the distance judgment items first. Exposure to ungrammatical, disharmonic items at test may have primed participants to select the disharmonic item during the two-alternative forced choice test. The two participants who completed the two-alternative forced choice test both selected the harmonic items at a rate greater than chance for all three types of test items, while, of the 11 participants who completed the distance judgment task first, two selected the harmonic item at a rate at or below chance across all items, and five selected the harmonic item at a rate at or below chance for Old items. Interestingly, even if the distance items primed participants accept familiar, but disharmonic items, participants still selected the harmonic alternative in the Prefix items at a rate greater than chance, suggesting that participants learned an abstract vowel harmony pattern that can extend to novel affixes.

4.2   *Distance Judgments*   Means and standard errors of the proportion of harmonic responses are presented in the figure in (4). The data were analyzed using a generalized linear mixed effects logistic regression fit by the Laplace approximation using the lme4 package in R (R Development Core Team 2011) with test condition (Old, New Stem and Prefix) crossed with Harmony as the fixed factors, with Old items set as the baseline for comparison. The alpha for significance was p <0.05, based on z scores (z > 1.96). The model was created with the maximal set of random slopes and intercepts for subjects and items that would successfully converge (Barr et al. 2013), which included random intercepts for subjects and items, and random slopes for subjects (as models with random slopes for items failed to converge).

Overall, participants rated Harmonic items as more similar than Disharmonic items ($\chi^2(3) = 21.43$, p<0.001), based on model comparison; a full model with crossed Harmony and Test items compared to a model without harmony as a factor. There was also a significant interaction between Harmony and Test item, as determined by a comparison of the model with and without the interaction term items ($\chi^2(2) = 10.79$, p=0.0045), as the difference between Harmonic and Disharmonic items was significantly greater for Old items compared to Prefix items ($\beta = 0.73$, $SE = 0.43$, $z = 3.41$, $p = 0.0028$). Contrasts were also performed in order to compare differences between Old and Prefix items and New Stem and Prefix items. There was a significant difference between similarity ratings between Old (Mean = 1.96, SD = 0.45) and Prefix (Mean = 3.60, SD = 0.80) items, and New Stem (Mean = 2.26, SD = 0.37) and Prefix items, suggesting that prefix items were rated as more dissimilar than both Old and New Stem items.

Pairwise comparisons were made using the least squared means package in R (Lenth 2016) with an mvt correction for multiple comparisons, which is equivalent to a step-wise correction in the multiple comparison (multcomp) package (note that the same general results emerge when with model comparisons when a subset, such as only Old items is selected in the data). There were significant differences between Harmonic and Disharmonic items for Old ($\beta = 1.04$, $SE = 0.22$, $z = 4.72$, $p < 0.001$) and New Stem ($\beta = 0.31$, $SE = 0.22$, $z = 1.39$, $p = 0.71$) test items, but not for Prefix items ($\beta = 0.71$, $SE = 0.22$, $z = 3.23$, $p = 0.024$). In order to test the hypothesis distance judgments do not reflect grammaticality, pairwise comparisons between distance judgments for harmonic prefix items and disharmonic New Stem ($\beta = -0.81$, $SE = 0.22$, $z = -3.67$, $p = 0.0070$) and Old items ($\beta = -0.97$, $SE = 0.22$, $z = -4.40$, $p < 0.001$) were made, both of which were significantly different, suggesting that learners found disharmonic Old and New items to be more similar to the training items than harmonic Prefix items.

(4) Similarity Results (Means and Standard Errors)
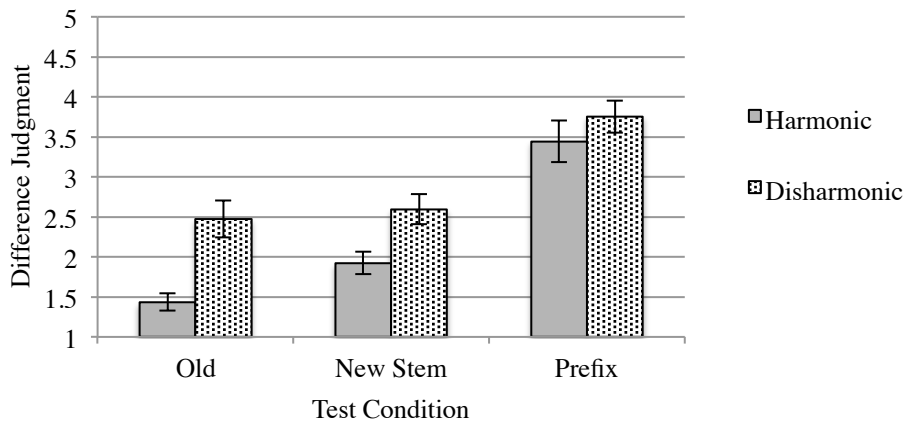
The results of the distance judgment task demonstrated that similarity reflects a difference between harmony and disharmony only for Old and New Stem items, suggesting that the ability to generalize beyond the input to novel items cannot be explained in terms of comparison to known items, as learners are able to generalize the harmony pattern to Prefix items in the two alternative forced choice task, despite the fact that the same learners judged these prefixed items to be less similar than Old or New Stem items, even comparing grammatical prefixed items to ungrammatical Old and New Stem items. These results support a view in which learners make use of abstract representations when they are learning novel phonological patterns, and that their decisions about whether to include a novel item into the representation of a novel pattern is not based solely on similarity to existing items.

5   Discussion

The results of the present study provide a replication of Finley and Badecker (2009b) in which participants were exposed to a novel back/round vowel harmony pattern with suffix alternations and then

generalized that pattern to items containing prefixes. This provides evidence that the participants in artificial grammar learning paradigms are able to extend patterns to novel structures, and are able to generalize beyond what is learned, supporting an abstract, general representation of the novel vowel harmony pattern.

The present study was designed to explore the possibility that the generalization to novel affixes found in Finley and Badecker (2009b) was based on similarity judgments, rather than an abstract rule. In an exemplar model, abstract processes are epiphenomenal, and grammaticality judgments to novel items are based on comparisons to known items. A similarity judgment cannot explain the present results, as the learners in the present study accepted the Prefix items as grammatical, even though learners rated grammatical prefix items as less similar than ungrammatical items from Old and New Stem items. In addition, it was possible to distinguish between grammatical and ungrammatical items based on similarity judgments for Old and New items, but not for Prefix items. Participants rated prefixed items as highly dissimilar, and showed no significant distinction between grammatical and ungrammatical items.

While the results of the present study show evidence against the predictions made by similarity-based models of generalizations, there are some reasons to be cautious about the interpretation of the results. First, the two alternative forced choice task directly pit harmonic and disharmonic items against each other, while the distance judgment task did not directly ask which of two items was more similar to the training items. Thus, it is difficult to directly compare the results of the two alternative forced choice task, and the similarity task. Second, the participants were asked to make conscious judgments in the distance judgment task. These conscious judgments about similarity may not be an accurate reflection of how participants make grammaticality judgments. It is possible that consciously deciding the similarity of items circumvents an unconscious model of grammaticality that makes use of similarity. Future research will work to resolve some of these issues. One possibility would be to create grammaticality and similarity judgment tasks that are more comparable in nature, either by using a similar scale or a similar set of instructions. Compatibility of tests will make it possible to directly compare the results across scale types, rather than making inferences across tasks. This will make it easier to understand the underlying mechanisms used in of each type of task, and to address whether similarity and grammaticality judgments tap into the same type of knowledge. One of the issues in comparing exemplar models and abstract models of language processing is that the tasks used to test each theory tend to be different, making it difficult to compare across tasks. In addition, the focus of research and the focus of the research tend to be different depending on the theory addressed. In order to better understand predictions that both theories make, it is extremely important to test both theories using a variety of comparable tasks, over a variety of effects.

The results of the present study suggest that metrics of similarity, as proposed by exemplar models of phonology, cannot account for the ability to distinguish between grammatical and ungrammatical items when the structure of the novel item differs significantly from known items. This is in line with experimental results that show gradient acceptability judgments for words with a lexical frequency of zero, and relatively equal frequency to known items (Berent and Lennertz 2010; Moreton 2002). The present study, along with previous research, suggests that speakers form abstract knowledge about phonological representations that goes beyond the lexicon.

While the present results do not discount the context specific effects that exemplar models were designed to account for (such as talker-specificity, lexical frequency and social context), it does suggest that exemplar models cannot replace abstract models of phonology. This leads to a question of how to account for the wide range of phenomena in phonology: from the very abstract to the very fine-grained. This relates to the dichotomy between categorical and gradient phonological processes, discrete and distributed representations, and competence and performance. One possible way to understand these dichotomies is to think about each end of the spectrum as a separate level of abstraction that can be integrated together through a general cognitive architecture. Under this approach, different systems generate phonological knowledge, some highly abstract and categorical, others highly gradient and context dependent. A mechanism of integration between levels of abstraction creates phonological knowledge that allows for high levels of abstraction, but also highly detailed representations based on lexical and social context. The idea of cognitive system that integrates multiple levels of abstraction has been proposed in Smolensky and Legendre's (2005) 'Integrated Connectionist/Symbolic Cognitive Architecture'. In this model, abstract symbols are integrated with distributed networks that are designed to handle large amounts of variability. An integrated cognitive system of phonology may allow for an adaptation of Optimality

Theory that takes into account the insights from exemplar models. Recent work using Harmonic Grammar, a variant of OT that allows for gradient constraint violations (Pater 2009; Potts et al. 2010), and other work that integrates gradient, phonetic and contextual effects into the abstract models (Hayes et al. 2009; Zuraw 2007; Flemming 1995; Hayes and Steriade 2004), demonstrates that OT may be an excellent starting point for working towards an integrated model of phonological representation, in which both fine-grained, distributed effects and abstract, symbolic effects can be accounted for in a way that provides an explanation for the mental processes that underlie phonological knowledge.

The present paper provides evidence from an artificial grammar learning task that supports abstract representations for phonology. Learners exposed to a novel back/round vowel harmony pattern were able to discriminate between grammatical and ungrammatical items that contained novel affixes, even though participants rated these novel items as highly dissimilar to exposure items. These results suggest that learners do not use exemplar-based notions of similarity as a metric for acceptability, but use abstract notions of grammaticality and markedness.

References

Barr, Dale J.; Roger Levy.; Christoph Scheepers.; and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68.255–278. doi:10.1016/j.jml.2012.11.001.

Berent, Iris.; and Tracy Lennertz. 2010. Universal constraints on the sound structure of language: phonological or acoustic? *Journal of Experimental Psychology: Human Perception and Performance* 36.212–223. doi:10.1037/a0017638.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14.261–290. doi:10.1017/S0954394502143018.

Bybee, Joan. 2012. Patterns of lexical diffusion and articulatory motivation for sound change. *The initiation of sound change. Perception, production, and social factors*, ed. by Maria-Josep Solé and Daniel Recasens, 211–234. Philadelphia, PA: John Benjamins Publishing.

Chomsky, Noam. 1965. *Aspects of a theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam.; and Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

Coetzee, Andries W.; and Joe Pater. 2011. The place of variation in phonological theory. *The handbook of phonological theory (2nd ed.)*, ed. by John Goldsmith, Jason Riggle, and Alan Yu, 401–431. Blackwell Publishers.

Coetzee, Andries.; and Shigeto Kawahara. 2013. Frequency biases in phonological variation. *Natural Language and Linguisic Theory* 31.47–89.

Cohen-Goldberg, Ariel. 2015. Abstract and lexically specific information in sound patterns: Evidence from /r/-sandhi in rhotic and non-rhotic varieties of English. *Language and Speech* 58.522–548.

Cristia, Alejandrina.; Jeff Mielke.; Robert Daland.; and Sharon Peperkamp. 2013. Similarity in the generalization of implicitly learned sound patterns. *Laboratory Phonology* 4. doi:10.1515/lp-2013-0010.

Drager, Katie.; Jennifer Hay.; and Abby Walker. 2010. Pronounced rivalries: Attitudes and speech production. *Te Reo* 53.27–53.

Finley, Sara. 2009. Morphemic harmony as featural correspondence. *Lingua* 119.478–501.

Finley, Sara. 2010. Exceptions in vowel harmony are local. *Lingua* 120.1549–1566.

Finley, Sara. 2013. Generalization to unfamiliar talkers in artificial language learning. *Psychonomic Bulletin & Review* 20.780–9. doi:10.3758/s13423-013-0402-7.

Finley, Sara.; and William Badecker. 2009a. Artificial language learning and feature-based generalization. *Journal of Memory and Language* 61.423–437.

Finley, Sara.; and William Badecker. 2009b. Right-to-left biases for vowel harmony: Evidence from artificial grammar. *Proceedings of the 38th North East Linguistic Society Annual Meeting*, ed. by Anisa Shardl, Martin Walkow, and Muhammad Abdurrahman, 1:269–282. Amherst, MA: GLSA.

Flemming, Edward. 1995. Auditory Representations in Phonology. PhD Dissertation, UCLA.

Gafos, Adamantios. 2002. A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20.269–337.

Goldinger, Stephen D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22.1166–1183.

Gouskova, Maria.; and Michael Becker. 2013. Nonce words show that Russian yer alternations are governed by the grammar. *Natural Language & Linguistic Theory* 31.735–765. doi:10.1007/s11049-013-9197-5.

Hale, Mark.; and Charles Reiss. 1998. Formal and empirical arguments concerning phonological acquisition. *Linguistic Analysis* 29.656–683.

Hayes, Bruce.; and Donca Steriade. 2004. The phonetic bases of phonological markedness. *Phonetically based phonology*, ed. by Bruce Hayes, Robert Kirchner, and Donca Steriade, 1–33. Cambridge: Cambridge University Press.

Hayes, Bruce.; Kie Zuraw.; Péter Siptár.; and Zsuzsa Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85.822–863.

Lenth, Russell. 2016. Using lsmeans. https://cran.r-project.org/web/packages/lsmeans/vignettes/using-lsmeans.pdf.

Mahanta, Shakuntala. 2012. Locality in exceptions and derived environments in vowel harmony. *Natural Language and Linguisic Theory Linguistic Theory* 30.1109–1146. doi:DOI:10.1017/S0025100312000096.

Moreton, Elliott. 2002. Structural constraints in the perception of English stop-sonorant clusters. *Cognition* 84.55–71.

Nygaard, Lynne C. 1998. Talker-specific learning in speech perception. *Perception & Psychophysics* 60.355–376.

Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33.999–1035.

Port, Robert.; and Adam Leary. 2005. Against formal phonology. *Language* 81.927–964.

Potts, Christopher.; Joe Pater.; Karen Jesney.; Rajesh Bhatt.; and Michael Becker. 2010. Harmonic Grammar with linear programming: from linear systems to linguistic typology. *Phonology* 27.77–117.

Prince, Alan.; and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar.* Cambridge: Blackwell.

Pycha, Anne.; Pawel Nowak.; Eurie Shin.; and Ryan Shosted. 2003. Phonological rule-learning and its implications for a theory of vowel harmony. *West Coast Conference of Formal Linguistics* 22.101–113.

R Development Core Team, R. 2011. R: A Language and Environment for Statistical Computing. (Ed.) R Development Core Team. *R Foundation for Statistical Computing*. R Foundation for Statistical Computing. R Foundation for Statistical Computing. doi:10.1007/978-3-540-74686-7.

Rose, Sharon.; and Rachel Walker. 2011. Harmony systems. *Handbook of Phonological Theory (2nd ed.)*, ed. by John Goldsmith, Jason Riggle, and Alan Yu, 240–290. Oxford, UK: Blackwell.

Smolensky, Paul.; and Géraldine Legendre. 2005. Foundational implications of the ICS architecture. *The harmonic mind: From neural computation to Optimality-Theoretic grammar*, ed. by Paul Smolensky and Géraldine Legendre, 99–122. Cambridge, MA: MIT Press.

Wilson, Colin. 2001. Consonant cluster neutralisation and targeted constraints. *Phonology* 18.147–197.

Wolf, Matthew. 2005. An autosegmental theory of quirky mutations. *West Coast Conference of Formal Linguistics* 24.370–378.

Zuraw, Kie. 2007. The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog. *Language* 83.277–316.

Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguisic Theory* 28.417–472.