A Computational Model of General Rule Learning with Unnatural Classes

Shira Calamaro (shira.calamaro@yale.edu)
Department of Linguistics, 370 Temple Street
New Haven, CT 06511 USA

Abstract

This paper presents the results of a computational model of generalized phonological rule learning (Calamaro and Jarosz, 2012), which is used to model experimental studies on the learning of phonotactic patterns governed by natural and unnatural classes. I focus on two papers with conflicting results on the learnability of natural and unnatural rules. Saffran and Thiessen (2003) find that a phonotactic pattern of positional voicing restrictions governed by a natural class of segments is learned by infants, but a similar pattern governed by an unnatural class is not learned. In contrast, Chambers, Onishi, and Fisher (2003) find that infants can learn a phonotactic pattern governed by an unnatural class of segments. The computational model presented in this paper is able to account for these seemingly conflicting results, explaining both the learnability and unlearnability of rules governed by unnatural classes.

Keywords: Linguistics; Phonology; Language Acquisition; Computational Model; Statistical Learning

Introduction

Many artificial-language experiments have explored the learnability of sound patterns in acquisition and how these may reflect biases in the phonology. The interpretation of experimental results can often be attributed to a number of different theoretical models. In this paper, I explore the results of experimental studies on the learnability of unnatural rules and provide an analysis in a computational model.

Experiments in language acquisition have found conflicting results in the learnability of unnatural sound patterns. In one study, Saffran and Thiessen (2003) found that infants were able to learn phonotactic voicing restrictions governed by a natural class of segments, but were unable to learn the same pattern when governed by an unnatural class. In contrast, Chambers, Onishi, and Fisher (2003) have shown that phonotactic patterns governed by unnatural classes may be learnable.

In this paper, I present a computational model of generalized rule learning (Calamaro and Jarosz, 2012), which offers an account of the results found by Saffran and Thiessen (2003) and Chambers et al. (2003). This model uses statistical regularities in the input, as well as linguistic filters, to learn phonological alternations. It encodes these patterns as generalized rules over natural classes of segments, which can explain the inability to generalize certain patterns that do not fall into a natural class.

The results of these computational experiments will help to further clarify the nature of the results of the acquisition studies. The preference for patterns governed by natural classes over unnatural classes may be explained by the inability to generalize over certain classes of segments. This preference is realized in the model through a generalization bias, or preference for general rules. The learnability of some types of unnatural rules is also explained by the model, which can identify the robust patterns present in the data and distinguish between them through the interaction of complexity and competition.

Background

In their artificial-language experiments on phonological acquisition, Saffran and Thiessen (2003) attempt to find the types of patterns that are learnable by infants and identify the types of pattern that are more difficult to learn. 9-month-old infants were trained on a set of language data exhibiting the specified pattern. They were then tested using the head-turn preference procedure, in which listening times for familiar and novel words were measured. A significant difference in listening times would indicate which patterns had been learned by the infants after a brief training period.

In one experiment, they looked at the learning of voicing restrictions in different positions of a syllable. Using two conditions, they restricted the types of consonants that could appear in the onset, the position preceding the vowel, and the coda, the position following the vowel. In one condition, the onset position was restricted to the set of voiceless stops [p,t,k], while the coda was restricted to voiced stops [b,d,g]. For example, words of the form *pibtad* were permitted, but not **bipdat*. In the second condition, the restrictions were reversed, with voiced stops in the onset and voiceless stops in the coda. The sets [p,t,k] and [b,d,g] each form a natural class of stop consonants because they can be distinguished using a single feature, [voice]. The results showed that infants were indeed capable of learning this distinction, with a significant difference in looking times between familiar and novel words. In this experiment, the infants were able to learn a phonotactic pattern governed by a natural class of segments.

The next experiment investigated the learning of voicing restrictions of unnatural classes in different prosodic positions. Unlike the previous experiment, in which the sets [p,t,k] and [b,d,g] could be distinguished by the [voice] feature, the sets used in the second experiment cannot be distinguished by any feature, making them unnatural. In one condition, [p,d,k] appeared in the onset while [b,t,g] appeared in the coda. The reverse was true in the second condition, with [b,t,g] appearing in the onset and [p,d,k] occurring in the coda. The experimental results differed, with no significant difference in looking times between the familiar and novel words. In the experiment, the infants failed to learn a phonotactic pattern governed by an unnatural class of segments.

Overall, the Saffran and Thiessen (2003) results show that infants are capable of learning patterns over a set of segments that form a natural class and can be described by a minimal number of features, and it is more difficult to learn a pattern over an unnatural class of segments which cannot be described by any set of features.

In contrast, Chambers et al. (2003) have shown that infants are capable of learning phonotactic patterns governed by an unnatural class of segments. In this experiment, 16.5-month-old infants were tested using a head-turn preference test. The training data consisted of artificial CVC words in which the set of segments [b, k, m, t, f] and [p, g, n, ʧ, s] were restricted by position, appearing in either the onset or coda. There is no combination of features that can be used to define these segments, so these sets of segments constitute an unnatural class. In the testing phase of the experiment, infants were able to distinguish between legal and illegal words, meaning they had learned the phonotactic pattern they had been trained on.

The results found by Chambers et al. (2003) seem to be in conflict with the results found by Saffran and Thiessen (2003) on the learnability on rules governed by unnatural classes. In addition to the distinction between natural and unnatural classes, a learning model should also be able to account for these different results on the learnability of unnatural classes. In the next section, I present such a model to account for these results.

Generalized Rule Learning Model

The Generalized Rule Learning model (GRL: Calamaro and Jarosz 2012) presented here is used to test the learning of the acquisition data in a computational setting. The GRL is a statistical model with linguistic constraints and generalized rule learning. The generalization component of the model is motivated by experimental evidence showing that infants are able to generalize rules using features (Maye, Weiss, and Aslin 2008; Cristiá and Seidl 2008). Given a set of segmented data, the model learns general rules for alternations in the data at the contexts in which they occur, as well as a score reflecting the strength of the rule. The original goal of the GRL model was the learning of alternations, such as word-final devoicing in Dutch, but in this paper it is applied to static phonotactic patterns. The GRL model is based on an earlier model (PLND: Peperkamp, Le Calvez, Nadal, and Dupoux, 2006) for learning pairs of alternating segments by calculating their statistical distribution in the data with an application of KL-divergence (Kullback and Leibler, 1951) and linguistic filters.

The GRL model maintains the use of the two linguistic filters[1] from PLND, which remove spurious pairs that should not be considered as alternating segments for linguistic reasons. The first filter removes pairs which have an intervening segment in the phonetic space based on their features, which are represented by a vector with values for place, sonority, voicing, nasality, rounding, and vocalic.[2] The second filter removes pairs in which the allophone is not more similar to its context than the default segment. Overall, these filters are able to introduce phonological knowledge not available to a purely statistical model.

The GRL model also maintains use of KL-divergence, though the formulation is somewhat changed, with the new calculation shown in (1) :

(1) $$\boldsymbol{Score}_{(\mathbf{c,s1,s2})} = \boldsymbol{D}\,(\boldsymbol{c}, \boldsymbol{s_1}, \boldsymbol{s_2}) * \mathbf{Z^2}$$

Where:

$$\boldsymbol{D}\,(\boldsymbol{c}, \boldsymbol{s_1}, \boldsymbol{s_2}) = \boldsymbol{P}(\boldsymbol{c}|\boldsymbol{s_1}) \log \frac{P(c|s_1)}{P(c|s_2)} + \boldsymbol{P}(\boldsymbol{c}|\boldsymbol{s_2}) \log \frac{P(c|s_2)}{P(c|s_1)}$$

and $\boldsymbol{Z} = \frac{KL\,(c, s_1, s_2) - \mu}{\sigma}$

The equation in (1) is used to calculate scores for pairs of alternating pairs of segments at a context *c*, defined as the following segment. The use of KL-divergence to find alternating pairs captures the intuition that segments which have highly distinct distributions in the data are likely to governed by some phonological or phonotactic rule.

The model creates general rules by merging alternating pairs which undergo an identical structural change, as represented by a feature vector. For example, the alternating pair (d,t) has a structural change of [0,-1,-1,0,0,0], calculated as the difference between the feature vector of segments *t*: [4,1,0,0,0,0] and *d*: [4,2,1,0,0,0]. This difference vector represents the devoicing pattern of the (d,t) pair.

The scores of alternating pairs as calculated in (1) are summed for all pairs whose change in features is the same, giving a contextualized rule score. Each contextualized rule is represented by a structural change vector, the context in which it occurs, and a rule measuring its strength. The calculation of contextualized rule scores is shown in (2):

(2)

$$\boldsymbol{Score}_{(c,\vec{f})} = \sum_{(\mathbf{s1,s2})\ \text{where}\ \mathbf{s1-s2}=\vec{f}} \boldsymbol{Score}_{(\mathbf{c,s1,s2})}$$

The output of the formula in (2) is a set of rules which each apply at a single context. Many phonological rules apply at multiple, related contexts, such as a vowel nasalization rule that applies in the context of all nasal segments. The contextualized rule scores can be further generalized, by merging rules whose contexts are phonologically related to each other and the change undergone by the rule. The formal calculation of the rule merging is defined in (3):

(3)

$$\boldsymbol{Score}_{(\{C\},\vec{f})} = \sum_{\mathbf{c} \in \mathbf{C}\ \text{where SCC and SVC hold}} \boldsymbol{Score}_{(\mathbf{c},\vec{f})}$$

[1] See Appendix for formal definitions of the two filters.

[2] See Appendix for the set of phonetic features.

Shared Change Condition (SCC): To merge, contexts must share feature values for any non-zero values in $\vec{f}$.

Shared Values Condition (SVC): To merge, contexts must not differ along more than one feature.

The formula in (3) is used to calculate the score of a generalized rule as the sum of all rules whose contexts meet two conditions: the Shared Change Condition (SCC) and the Shared Values Condition (SVC). Like the linguistic filters from Peperkamp et al (2006), the merging conditions provide linguistic information in assigning classes of sounds that pattern together. The SCC requires that contexts must be related to the rule change in the same way by restricting merging to contexts which share non-zero values of the rule vector. The SVC requires that contexts be related to each other by restricting merging to contexts which only differ along a single feature dimension, thus approximating a natural class. The merging of contextualized rules into generalized rules can capture generalizations about the data as well as assign increased scores to more robust rules occurring in a set of related contexts.

This model learns generalized rules as a difference vector of features, a set of contexts of application, and a score indicating the goodness of the rule. The rules learned by the model will need to be interpreted somewhat differently from the results of the Saffran and Thiessen (2003) and Chambers et al. (2003) experiments, which measured successful learning by significant differences in looking times in a head-turn test. Instead, this model will need to look for rules which reflect the regularities found in the training data. Additionally, the model looks at alternations conditioned by contexts defined as following segments and does not have access to syllabic structure. Due to these limitations of the model, this discussion will focus on the results as they relate to the learning of the pattern in coda position, which is defined by the following segment. With these restrictions in mind, successful replication of results in the model will mean the learning of a word-medial and word-final voicing/devoicing rule in Experiment 1, no successful learning of any such a rule in Experiment 2, and the learning of meaningful rules in Experiment 3.

Experiment 1: Learning rules governed by natural classes

In Experiment 1, I replicate the results of an experiment by Saffran and Thiessen (2003), in which infants were able to learn voicing restrictions by position.

Method

The Generalized Rule Learning Model, as described in the previous section, was used.

Data

The same training data from Saffran and Thiessen (2003) was used. Each condition in the training data consisted of 30 unique CVCCVC words for each condition, made from an alphabet with four vowels [a, i, o, u], three voiceless stops [p, t, k], and three voiced stops [b, d, g]. In condition a, voiced stops were restricted to coda position and voiceless stops were restricted to onset position. The opposite was true for condition b, with voiceless stops occurring in coda position and voiceless stops occurring in onset position.

While the model does not specifically reference syllable structure, successful learning of this data would find a rule of voicing/devoicing in the word-final context and before voiceless/voiced consonants.

Results

The results from experiment 1 are shown in Figure 1, reflecting the highest scoring rules found by the model.

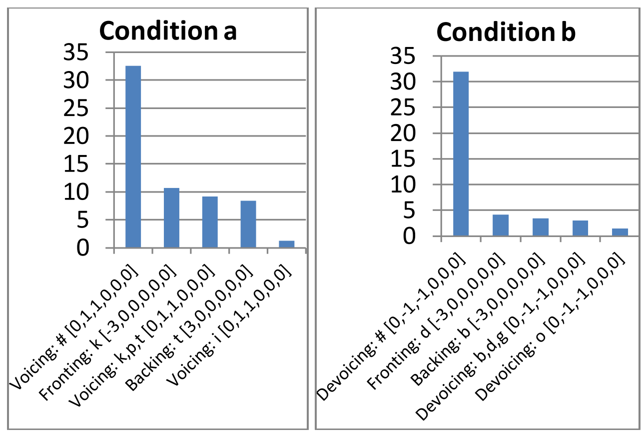

Figure 1: Exp. 1 results

Each bar in Figure 1 shows the score of a generalized rule. In Condition a, the highest scoring rule is the word-final voicing rule, (# [0,1,1,0,0,0]), where # represents the word-final context and [0,1,1,0,0,0] represents the structural change vector. The two non-zero values in the vector indicate a change in the sonority and voicing features in pairs such as (t,d).[3] The reverse rule is found in Condition b, with a structural change vector [0,-1,-1,0,0,0] indicating devoicing in pairs such as (d,t).

In each of the two conditions, the highest scoring rule is the desired voicing or devoicing rule. This rule reflects the change in voicing of the stops in coda position in the training data. The voicing/devoicing rule is quite robust in each of the two conditions, scoring much higher than the next highest scoring rule. A similar rule for word-medial codas is also found, which is the voicing/devoicing rule occurring in {p,t,k} or {b,d,g} contexts.

A number of spurious rules were also found by the model. These rules reflect a change in place of articulation, shown as fronting and backing rules. While these rules are not desired, they do reflect a generalization in the data, namely, a possible alternation between pairs like [p,t] or [t,k], in which the segments differ only in place of articulation. These spurious rules are likely an artifact of the small

[3] See Appendix for the full set of feature values.

segment inventory, making a minor statistical regularity appear to reflect a possible alternation. In artificial language learning, these types of spurious statistical regularities have the potential to affect the results to a greater extent than in natural language learning, as will be seen in Experiment 3.

Overall, the results in Experiment 1 show learning of the phonotactic pattern, aligning with the results found by Saffran and Thiessen (2003). The model successfully learned the voicing restrictions when they were governed by a natural class of segments.

Experiment 2: Failure to learn rules governed by unnatural classes

In experiment 2, I replicate the results of a second experiment from Saffran and Thiessen (2003), in which infants were not able to learn phonotactic restrictions of unnatural classes of segments which are specified by voice and place of articulation.

Method

The Generalized Rule Learning Model, as used in the previous experiment.

Data

The same training data from the Saffran and Thiessen (2003) experiment was used. The training data consisted of 30 CVCCVC words in each condition with the same alphabet as experiment 1. In condition a, the set of coda consonants was [b, t, g] and the set of onset consonants were [p, d, k]. In condition b the voicing specifications were reversed, with codas [p, d, k] and onsets [b, t, k].

Results

In Exp. 2, the model failed to learn voicing restrictions governed by unnatural classes. These results are shown in Figure 2, with the highest scoring rules represented.

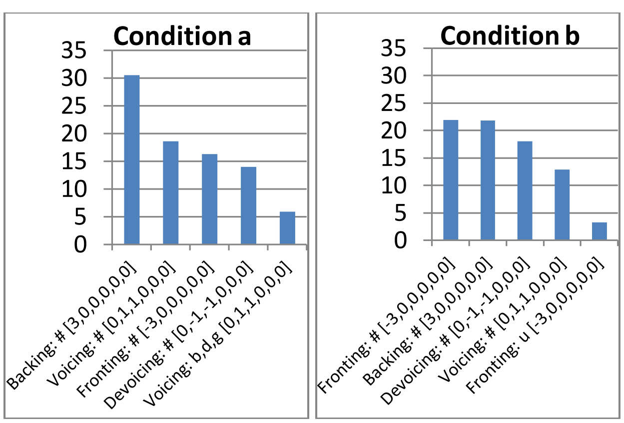

Figure 2: Exp. 2 results

The voicing and devoicing rules are no longer learned as the highest scoring rules, as seen in Figure 2. The highest scoring rule is now a spurious rule reflecting a change in the place of articulation. This rule would account for a possible alternation between pairs such as (p,t) or (d,g), which can be generalized from statistical regularities in the data.

The desired rules of voice alternations receive lower scores than some of the spurious rules. The overall strength of these desired rules has decreased, with the weight of each voicing/devoicing rule being split into two lower weighted rules. The reason for this decrease in the score is that the patterns cannot be fully generalized because they belong to an unnatural class. In experiment 1, the desired voicing rules were supported by three pairs of segments, one for each place of articulation. In this experiment, the scores were split between two separate rules, each supported by one or two pairs of segments, (t,d) or (b,p) and (g,k).

Both factors of decrease in rule rank and loss of rule strength contribute to the increased difficulty of learning the phonotactic pattern in experiment 2. This difficulty in learning is a desired result because infants failed to learn this same pattern in an experimental setting (Saffran and Thiessen 2003).

In this case, the unnatural voicing pattern was not the most robust pattern in the data. The model found other patterns which were generalizable from the given data, obscuring the desired patterns. From this result, a prediction of the model is that it would be able to learn a rule governed by unnatural classes, if the data did not contain any other patterns which could be inferred. Such a case is used by Chambers et al. (2003), which will be shown in the following experiment.

Experiment 3: Learning rules governed by unnatural classes

In a final experiment, I run the GRL model on the data from Chambers et al. (2003), in which infants were able to learn phonotactic patterns governed by an unnatural class of segments.

Method

The Generalized Rule Learning Model, as used in the previous experiments.

Data

The data used in this experiment were replicated from Chambers et al. (2003). A set of CVC words were creating using two groups of consonants belonging to an unnatural class: [b, k, m, t, f] and [p, g, n, ʧ, s]. The onsets were drawn from one group and codas from another, creating a phonotactic pattern governed by an unnatural class of segments.

Results

While the data could not be generalized, the patterns were learnable as separate rules, as shown in Figure 3:

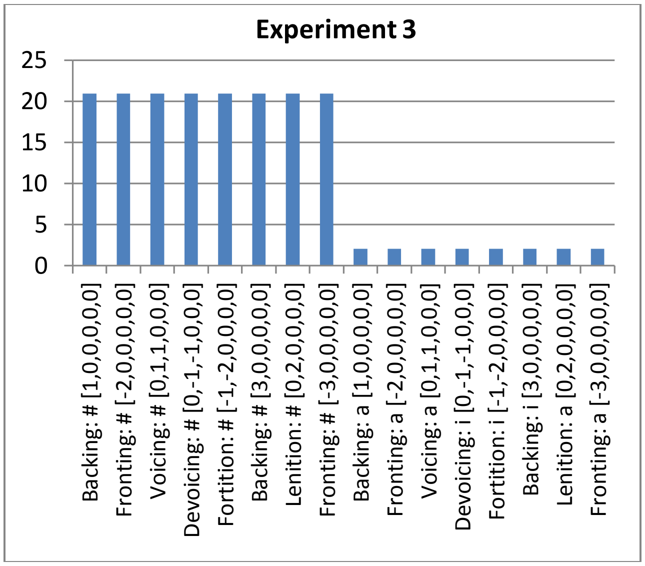

Figure 3: Exp. 3 results

The rules shown in Figure 3 are striking due to the uniformity of the data. While the segments could not be generalized by position, the model was able to find relationships among between-group segments. For example, the (b, p) pair is reflected by the word-final devoicing rule (# [0,-1,-1,0,0,0]), while the (k, g) pair is reflected by the word-final voicing rule (# [0,1,1,0,0,0]).

With a one-to-one mapping of segments to learned rules, we would expect five rules, but instead find eight. While each segment belongs to at least one rule, some segments are learned as multiple rules. For example, 'p' is found in both the devoicing rule (# [0,-1,-1,0,0,0]) with the pair (b,p), but also in the fortition rule (# [-1,-2,0,0,0,0]) with (f,p).

While some of these are the same rules which were unlearnable in Experiment 2, namely voicing and devoicing, a potential difference here is the lack of interference from spurious rules. While in the case demonstrating unlearnability, the desired rules were dominated by spurious rules. In this experiment, the desired rules were the highest scoring rules.

Discussion

The computational experiments presented in this paper seek to address two fundamental questions about the learnability of phonotactic patterns: Why are patterns governed by natural classes easier to learn than those governed by unnatural ones? How can we explain results in which unnatural patterns are learnable? The first question is addressed by comparing the results of Experiments 1 and 2, and the second by comparing the results of Experiments 2 and 3.

Natural vs. Unnatural Classes

In Experiments 1 and 2, the GRL replicated the results found by Saffran and Thiessen (2003), that a phonotactic pattern governed by a natural class is learned, while one governed by an unnatural class is not. Specifically, infants can learn patterns which occur over a set of segments that all agree in voicing and differ in place, they cannot learn patterns which occur over a set of segments that differ in both place and voicing.

The GRL finds an asymmetry in the learning of natural and unnatural classes due to an inherent bias in the generalization mechanism. Generalized rules receive higher scores from the model because they have support from more pairs of segments. The strength of general rules is computed by summing the scores of rules governing alternations between a single pair. Therefore, the more pairs of segments contributing to a general rule, the higher its score will be. In the case of the Saffran and Thiessen (2003) data, the rules governed by natural classes are supported by more segments than the unnatural ones. This inherent generalization bias assigns higher scores to the natural rules in Exp.1 than the unnatural rules in Exp. 2.

The asymmetry in the learning of natural and unnatural rules has previously been explained by a Complexity Bias (Moreton and Pater, 2011). Under this account, the more complex set of features needed to describe unnatural classes makes the learning of unnatural patterns more difficult. Natural classes, which can be described with fewer features, can be learned more easily.

The generalization mechanism in the GRL accounts for the same patterns as the Complexity Bias, but for a different reason. While the Complexity Bias asserts that unnatural rules are more difficult to learn because they require the encoding of additional feature values, the GRL attributes this asymmetry to weaker statistical regularities due to the more complex data. This prediction of the GRL can be seen by the difference in rules scores in Exp. 1 versus Exp. 2.

The GRL model has an additional property that interacts with complexity: competition. In the results from Exp.1, the desired pattern was learned because of the high score relative to other rules. In Exp. 2, the lower scoring unnatural rules were dominated by competing spurious rules, interfering with their learnability. This interaction between complexity and competition allows the GRL to make additional predictions beyond complexity alone, which will play a role in the learning of different types of unnatural classes.

Unnatural vs. Unnatural Classes

In Experiments 2 and 3, the learning data contained phonotactic patterns governed by unnatural classes. In the original experimental setting, infants did not learn the unnatural pattern in Experiment 2 (Saffran and Thiessen 2003), but did learn the pattern in Experiment 3 (Chambers, et al. 2003). Likewise, the GRL found a similar difference in the learnability of the two unnatural patterns, as shown in this paper. The distinction to be made between these two unnatural patterns lies in the nature of the data.

Both experiments presented artificial data in which syllable positions were restricted to a specific set of consonants. In Saffran and Thiessen (2003) the sets were [p, d, k] and [b, t, g]; in Chambers et al. (2003) they were [b, k, m, t, f] and [p, g, n, ʧ, s]. While both sets of data are unnatural to some extent, there is a striking difference in the segment inventories of the two experiments.

While the pattern presented in Saffran and Thiessen (2003) is unnatural, the segment inventory is well-balanced among the feature set it uses, with a voicing distinction present for each place of articulation. In contrast, the segment inventory of Chambers et al. (2003) is not as balanced, with a mix of voicing, place and sonority distinctions that do not apply across all pairs of segments. For example, there is a voicing distinction for the pairs (p,b) and (k,g), but there exists no pair (t,d).

The effects of the overall naturalness of the data are seen directly in the computational results of Experiments 2 and 3. In Experiment 2, the more balanced data allowed the GRL to make a number of spurious generalizations, obscuring the robustness of the desired unnatural rules. In Experiment 3, the less balanced data could not be generalized by the model, leaving the set of desired unnatural rules as the most robust in the data.

In the distinction between these two sets of unnatural patterns, the GRL is better able to predict these results than a model using complexity alone. The Complexity Bias (Moreton and Pater 2011) predicts difficulty in the learning of both types of unnatural patterns, but would predict even greater difficulty in Exp. 3 due to the greater number of features needed to describe the unrelated set of segments. However, the experimental evidence shows the opposite is true, with the data in Exp. 3 being learned more easily. The predictions of the GRL align with the experimental evidence due to the interaction of competition and complexity in the model. The unnatural pattern in Exp. 3 is learned more easily than that in Exp.2 because the desired rules are not in competition with any high scoring spurious rules as is the case in Exp. 2.

Conclusion

The GRL is able to model the results of experimental data showing the learning of phonotactic patterns by infants. It can account for the preference for learning natural rules over unnatural ones, as well as the distinction between the learnability of different patterns of unnatural classes. This preference for natural classes is an inherent property of the model, due to the rule generalization component. While the generalization component of the model does facilitate the learning of natural rules, it does not exclude the learning of rules governed by unnatural classes. Indeed, rules governed by unnatural classes were learned by the model, when there were no other more robust rules in the data.

These experiments provide some promising results for the GRL model, with its ability to account for attested cases of phonological learning. While there remains a possibility that differences in infant learning can be attributed to differences in experimental methodologies, these results show compelling evidence for further exploration of this topic. Future work will explore other predictions made by the model and extensions needed to account for additional data.

Appendix

Linguistic filters (Peperkamp et al. 2006)

Allophonic distributions of s_a and s_d are spurious if:

$$\exists s \begin{bmatrix} \forall i \in \{1, \ldots, 5\}, v_i(s_a) \leq v_i(s) \leq v_i(s_d) \\ or\ v_i(s_d) \leq v_i(s) \leq v_i(s_a) \end{bmatrix}$$

With $v_i(s)$ the ith component of the vector representation of s.

Allophonic distributions of s_a and s_d are spurious if:

$$\exists i \in \{1, \ldots, 5\}, \left| \sum_{s \in C[s_a]} (v_i(s_a) - v_i(s)) \right| > \left| \sum_{s \in C[s_a]} (v_i(s_d) - v_i(s)) \right|$$

Feature values

Segments are represented as feature vectors with the following values:

Place: bilabial 1, labio-dental 2, dental 3, alveolar 4, post-alveolar 5, palatal 6, velar 7, uvular 8, glottal 9

Sonority: voiceless stop 1, voiced stop 2, voiceless fricative 3, voiced fricative 4, nasal 5, lateral 6, rhotic 7, glide 8, high vowel 9, mid vowel 10, low vowel 11

Voicing: voiceless 0, voiced 1

Nasality: oral 0, nasal 1

Rounding: unrounded 0, rounded 1

Vocalic: non-vowel 0, vowel 1

Acknowledgments

Many thanks for feedback from Gaja Jarosz, CLAY Lab, audiences at LSA 2013, and an anonymous reviewer.

References

Calamaro, S. and Jarosz, G. (2012). A computational model of general rule learning for phonological alternations. Ms.

Chambers, K.E., Onishi, K.H., and Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition* 87 B69-B77.

Christiá, A. and Seidl, A. (2008). Is infants' learning of sound patterns constrained by phonological features? *Language Learning and Development* 4(3), 203-227.

Kullback, S. and Leibler, R. (1951). On information and sufficiency. *Annals of Mathematical Statistics* 22, 76-86.

Maye, J., Weiss, D.J., and Aslin, R.N. (2008) Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental* Science(11)1, 122-134.

Moreton, E. and Pater, J. (2011). Learning artificial phonology: A review. Ms.

Peperkamp, S., Le Calvez, R., Nadal, J.P., and Dupoux, E. (2006). The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition* 101(3), B31-B41.

Saffran, J. and Thiessen, E. (2003). Pattern Induction by Infant Language Learners. *Developmental Psychology* Vol. 39, No. 3, pp. 484-494.