

Structural Biases in Phonology: Infant and Adult Evidence from Artificial Language Learning

Elika Bergelson and William J. Idsardi
University of Pennsylvania and University of Maryland

Introduction

Do biases in infant and adult language learning follow (or better yet, follow from) typological biases observed cross-linguistically? The evidence obtained thus far has been equivocal; the present study examines a previously uninvestigated type of bias: are certain phonological processes preferentially associated with certain positions in the word? In particular, do different phonological processes invoke different notions of *finality*? We investigated this question by switching the environments between two phonological processes (final devoicing and final stress) and testing adults and infants on their learning preferences between the typologically attested and unattested generalizations.

The infant language learner faces seemingly insurmountable challenges in her first few years of life. Not only must she eventually figure out that words exist, what they mean, and how they go together, but she must also (and perhaps first) figure out the extremely complicated phonological system of her native language. Fortunately, phonological systems are patterned and systematic. However, certain aspects of this systematicity seem quite difficult to induce from just the raw input: while there are a plethora of phonological rules, many of which simply vary on a given parameter (e.g. assimilation can be for various kinds of place features as well as for laryngeal postures) if you look crosslinguistically, there are certain systematic gaps in what kinds of patterns apply to what kinds of phonetic elements.

Previous work on rule naturalness and its relationship to learnability has provided mixed results. Some evidence that learners are biased towards attested patterns comes from Berent *et al* (2008), who found that adult Korean speakers prefer ‘bl’ onsets to ‘lb’ onsets, even though neither onset type is attested in Korean. This preference reflects general linguistic universals. Other evidence, however, such as Seidl and Buckley (2005) has shown that infants find some ‘unnatural’ rules just as learnable as natural ones.

Recently, Gerken and Bollt (2008), following up on work showing that multiple exemplars make it easier for infants to generalize phonetic rules, found that 9 month olds were unable to learn an unnatural rule (stress syllables starting with ‘t’), but 7.5 month olds were able to learn this rule, arguing that perhaps ability to learn unnatural kinds of rule alters over development.

Given the mixed results of previous work, we sought to test the relationship between learnability and the attestation of phonological rules in the world’s languages. We tested 8-month-old infants and adults on rules that occur

in natural languages, but not in English, through two artificial language learning studies. Our work differs from that of Buckley & Seidl (2005) and Gerken & Boltt (2008) in that we did not directly test whether infants could learn a single unattested rule if that were the only single generalization consistent with the data; rather, we presented our subjects with data that was equally compatible with either an attested or an unattested rule, leaving the subjects' preferences in the test phase as the determinants of which rule was posited by them. Thus, our experiments demonstrate learning preferences rather than absolute abilities or inabilities. Further studies using other methodologies will need to be done to address this issue. In our studies we crossed the focus and the environment of a pair of rules, one for word-final devoicing and the other for stress.

In our studies we examine notions of *finality*. Unfortunately, this word is used ambiguously in the linguistic literature. For example, in a rule like the French stress rule, 'stress the final vowel' the vowel only needs to be *relatively* final, that is, the vowel can be followed by one or more consonants (or by none). Thus, both *petit* [pəti] 'small masc.' and *petite* [pətit] 'small fem.' are stressed on their final vowels, even though only *petit* has a vowel at the absolute end of the word in its pronunciation. In SPE notation (Chomsky and Halle 1968) this is as in (1).

$$(1) \quad [-\text{consonantal}] \rightarrow [+ \text{stress}] / _ [+ \text{consonantal}]^* \# \quad (\text{relatively final})$$

Conversely, in a phonological process like the Russian devoicing rule 'devoice the final obstruent,' final is meant in an *absolute* sense: the rule only applies to the string-final element, applying in хлеб [xlɐp] 'bread nom. sg.' but not in хлеба [xlɐba] 'bread gen. sg.' even though there are no other obstruents following the /b/ in either word. In SPE notation this is as in (2). Note that (2) does not contain any use of the Kleene star operator (Kleene 1956) whereas (1) does.

$$(2) \quad [-\text{sonorant}] \rightarrow [-\text{voice}] / _ \# \quad (\text{absolutely final})$$

Repeating the point for additional clarity: the stress rule scans the form from right to left looking for a vowel to stress, it continues on, skipping consonants until it finds a vowel. Conversely, the devoicing rule examines the final segment, if it is an obstruent it devoices it, if the final segment is not an obstruent, the rule stops and looks no further.

However, one can easily imagine that each of these processes *could* apply in the switched senses of finality. That is, we can easily formulate a rule like 'stress the final segment if it is a vowel', (3).

$$(3) \quad [-\text{consonantal}] \rightarrow [+ \text{stress}] / _ \# \quad (\text{absolutely final})$$

Or 'devoice the final obstruent in the word, wherever in the word it falls', (4).

$$(4) \quad [-\text{sonorant}] \rightarrow [-\text{voice}] / _ [+ \text{sonorant}]^* \# \quad (\text{relatively final})$$

However, for whatever reasons, rules such as (3) and (4) are not attested in the languages of the world.

We sought to investigate whether this disparity in what kind of finality applies to which kinds of processes or elements is simply an accidental idiosyncrasy, or whether the human language learner finds it easier to posit and learn attested rules. In the case of the infant, it is likely that much of the data in the input is ambiguous, increasing the difficulty of the learner's task. Specifically, since the definition of 'finality' as relative or absolute differs depending on the context it's being applied to, the language learner is faced with a particularly steep challenge of not only discovering the existence of both kinds of finality, but the differentiated application of each in varying phonological settings.

1. Experiment 1

Experiment 1 tested the prediction that given ambiguous data that would allow subjects to generalize a stress rule that invokes absolute finality or relative finality, subjects would posit the (typologically attested) relative-finality rule, and in the test phase, would prefer novel words that follow this rule over novel words that do not. The rules in question are given in (A) and (B) (C=consonant, V=vowel, a colon indicates length):

(A) Stress a relatively-final long vowel, otherwise, stress is initial

$$V: \rightarrow 'V: / _ C^* \#$$

(B) Stress an absolutely-final long vowel, otherwise, stress is initial

$$V: \rightarrow 'V: / _ \#$$

Rule A appears in the languages of the world, for example, in Eastern Chereemis (Sebeok and Ingeman, 1961) while Rule B is unattested in the world's language.

1.1. Experiment 1a

1.2. Materials

Materials for the familiarization were open trisyllabic words in a CVCVCV pattern. They were constructed using the consonants /p, t, k, m/ and the vowels /a, i, o, u/ exhaustively in each C or V slot, resulting in 16 possible CV syllables. Of these, 4 CV pairings (/tu, ko, pa, mi/) were held back for use in the test phase, so that the test phase of the experiment would feature entirely novel syllables. Every token featured three different consonants and vowels.

The familiarization words were created in each of two types. In the first, the final syllable was stressed and the vowel was long (cvcvCV:). In the second familiarization pattern, the first syllable was stressed, but all syllables had short vowels (CVcvcv). Both types of familiarization words ended with open syllables.

The test tokens were trisyllabic words with final long, closed syllables of the form CVCVCV:C. They were created using the syllables /tu, ko, pa, mi/ in all permutations, adding /p, t, k, m/ to the ends of the syllables. These tokens were created in two versions: one set had final length and stress (cvcvCV:C); the other had final length but initial stress (CVcvcv:c). The former set are consistent with stress rule A, the latter with stress rule B.

All tokens were digitally recorded by a native Japanese speaker using a Marantz solid state recorder PMD660 with a Sennheiser head-mounted, noise-reducing microphone. These were then analyzed, edited and normalized with Praat (Boersma and Weenick 2008) to ensure consistent length, stress, and volume across tokens. The choice to use a Japanese speaker stemmed from our concern that the length and stress come across clearly in the tokens (both of these distinctions occur naturally in Japanese; English lacks a vowel-length distinction). All tokens were presented in random order for every participant

1.3. Participants/Procedure

Participants were 28 native English-speaking undergraduates at the University of Maryland. All gave signed, informed, consent.

Adults were seated in a sound proof room where they listened passively to ten minutes of the artificial language over headphones. They were permitted to draw during this time. Then, the experimenter instructed them to respond ‘yes’ or ‘no’ by pressing different keys on a standard computer keyboard as to whether new items ‘belonged to the language they had been listening to’ or not.

1.4. Results and Discussion

The adult participants were significantly more likely to accept items with final stress (those consistent with A, the attested rule) over items with initial stress (those consistent with B; mean difference = 0.063, $t(27) = 2.17$, $p < 0.02$), as shown in Figure 1. Again, this is in line with language typologies: only rule A exists in the languages of the world.

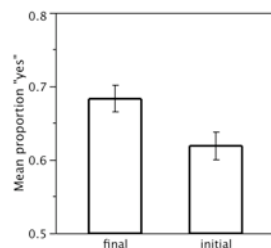


Figure 1: Mean proportion of “yes” responses by adults to cvcvCV:C and CVcvcv:c items (see text).

2. Experiment 1b

2.1. Materials/Participants

Materials were identical to those for Experiment 1a. Participants were 14 8-month old infants (6 females; range 7;16-9;0). They were recruited from English-speaking homes near College Park, Maryland.

2.3. Procedure

Infants were seated in their mother's laps in a dimly lit 14'x7' sound-proof room. They were seated approximately 6' from a Samsung 50" widescreen television in the center, which displayed images, and flashing lights and speakers on the left and right, which were used for infant attention-getting and auditory stimuli delivery, respectively. Mother and experimenter were unable to hear the stimuli playing for the infant. Mothers wore Sennheiser PXC250 Noise Canceling Stereo Headphones, through which they heard music, and were instructed not to communicate with their child through pointing or talking. In an adjacent room, the experimenter controlled and coded the experiment online, based on a close-up video image of the infant, recorded from a Sony EVI-D100 color video camera that was above the television. Infants were tested using the Head-Turn Preference procedure.

In a pretest phase, infants were familiarized to this procedure using classical music excerpts. An image would appear on the TV, and then a light would flash on the left or right. When the infant oriented towards the light, music began to play. As long as infants continued to look, the music continued, with the trial ending if infants looked away from the flashing light for more than two seconds, or if 30 seconds of look-time had been accrued. This was followed by two test trials from either the same excerpt as the familiarization or a new excerpt.

Then the experiment began. In the same manner as in the pretest, infants were exposed to the familiarization stimuli for two minutes, followed by a test phase in which they heard the two types of test items (CVcvcv:c or cvcvCV:C) on alternating sides for three trials for each stimulus type, resulting in 6 test trials total. Side of presentation and item type were randomized.

2.4 Results and Discussion

Infants, like the adults in Experiment 1a, preferred the cvcvCV:C items over CVcvcv:c items (mean difference = 1.65s, $t(15) = 2.28$, $p < 0.02$), see Figure 2. This preference is in line with the typologies of the world's languages, and is in contrast with English-speaking infants' natural preferences for initial stress (Juszyk, Cutler, & Ridanz, 1993; Houston, Santelman & Juszyk, 2004).

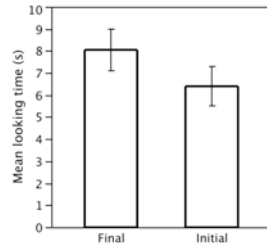


Figure 2: Mean looking time in seconds by infants to cvcvCV:C and CVcvcv:c items (see text).

3. Experiment 2

In Experiment 2 we tested the hypothesis that in the context of an obstruent-devoicing rule, our subjects would prefer to posit an absolute-final phonological rule, given ambiguous data that would support either an absolute-final or relatively-final rule. This prediction stems from the phonological rules in the world, which allow for absolute-final rules that operate over obstruents, but not relative-final rules.

We tested an absolute-final obstruent-devoicing rule that occurs in languages like Russian, and its (unattested) relative-final counterpart:

- (A) Voiced obstruents (e.g. /b, d, g/) are disallowed in absolute-final position:
 as a rule: $[- \text{sonorant}] \rightarrow [- \text{voice}] / _ \#$
 as a constraint: $*[- \text{sonorant}, + \text{voice}] \#$
- (B) Voiced obstruents are disallowed in relatively-final position:
 as a rule: $[- \text{sonorant}] \rightarrow [- \text{voice}] / _ [+ \text{sonorant}]^* \#$
 as a constraint: $*[- \text{sonorant}, + \text{voice}] [+ \text{sonorant}]^* \#$

We hypothesized that all subjects would learn some form of the voicing restriction and would come to disprefer words with final voiced obstruents (e.g. ‘otab’); this dispreference is consistent with both possible generalizations. Given that, we would then be able to see subjects’ response to words with crucially relative-finally voiced obstruents (e.g. ‘odam’). If subjects posited the relatively-final process B (the unattested rule), these words would be in violation of the generalization since the relatively-final obstruent in the word (which happens to be medial) is voiced. Alternately, if subjects posited A (attested in languages such as Russian), only the words ending with obstruents are under the jurisdiction of the rule, and thus medially voiced obstruents such as ‘odam’ should be judged as belonging to the language.

3.1. Experiment 2a

3.2. Materials

The materials were recorded by a native English speaker (the first author) using Prorec (<http://www.phon.ucl.ac.uk/resource/prorec/>). They were then analyzed, edited and normalized using Praat to ensure that all stimuli were of comparable length and volume.

Familiarization materials consisted of 200 bisyllabic tokens of the form VCVC, using the vowels /o, a, i, e/ and the consonants /p, t, k, m, n, b, d, g/ exhaustively. There were four types of items:

- i) words with medial and final nasals e.g. **onam**,
- ii) words with medial nasals and final voiceless obstruents e.g. **onap**,
- iii) words with medial and final voiceless obstruents, e.g. **otap**, and
- iv) words with medial voiced and final voiceless obstruents, e.g. **odap**.

All of these words were fully compatible with either voicing restriction, above. Nasals (being sonorants) are not informative for these rules, and were simply added for phonetic variety. As in Experiment 1, certain forms were withheld from the training so that they could be used in the test phase.

Because the languages consistent with relatively final devoicing (B) are a subset of those consistent with absolutely final devoicing (A) (as words like **odam** are consistent with A but violate B), the testing stimuli cannot be as straightforward as those employed in the stress case (where the stress ended up on the first vowel if the final generalization was inapplicable). Therefore, some forms of the same types used in the familiarization had to be present in the test cases as well (though the same exact items were never used in both familiarization and test). This was necessary to assess whether subjects did prefer absolutely final voiceless obstruents to voiced ones (e.g. was *onap* preferred to *onab*). Note that this generalization is not part of English phonology (modulo questions of vowel length signaling the voicing status of the final consonant), and so it was necessary to test whether subjects learned anything from the familiarization phase. Provided that the subjects succeeded in differentiating the status of voiced and voiceless obstruents in absolute final position, then we could test their opinion of medial obstruents. The tokens with medial voiced obstruents would wrest apart whether subjects were positing a relatively-final or absolutely-final generalization. Note also that the differentiation of voiced and voiceless items in absolute-final position is consistent with either hypothesis (both exclude forms such as **onab* and **otab*) and thus demonstrating this knowledge in the subjects is a prerequisite for their further differentiation of medially voiced items as acceptable in the language or not.

Therefore the test items fell into three groups: (I) voiced versus voiceless obstruents in absolute final position with medial nasals (e.g. *onab* vs. *onap*), (II) voiced versus voiceless obstruents in absolute final position with medial voiceless obstruents (e.g. *otab* vs. *otap*), and (III) voiced versus voiceless obstruents in medial (here equivalent to relative final but not absolute final) position with final nasals (e.g. *odam* vs. *otam*). All subjects should differentiate between

voiced and voiceless items in cases (I) and (II), the crucial question is whether they also differentiate between voiced and voiceless items in case (III).

3.3. Participants/Procedure

Participants were 37 undergraduates from the University of Maryland. All gave informed signed, informed consent. The procedure was identical to experiment 1a, using the new materials.

3.4. Results and Discussion

As predicted by both hypotheses, participants were significantly less likely to respond “yes” to items with voiced obstruents in absolute-final position with both medial nasals (type I, Figure 3, matched pairs $t(36) = 7.61$, $p < 0.0001$) and with medial voiceless obstruents (type II, Figure 4, matched pairs $t(36) = 8.60$, $p < 0.0001$).

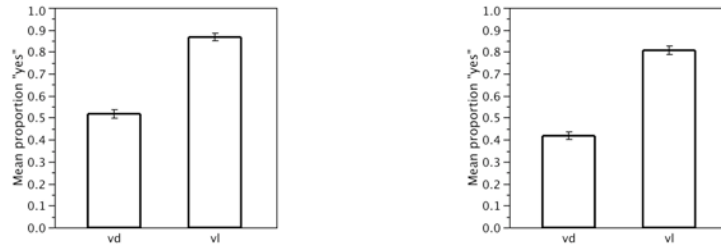


Figure 3 (left): Mean proportion of “yes” responses to words with medial nasals and obstruents in final position (e.g. onab vs. onap)

Figure 4 (right): Mean proportion of “yes” responses to words with medial voiceless obstruents and obstruents in final position (e.g. otab vs. otap).

These findings are consistent with either voicing restriction. The overall acceptance rates for both voiced and voiceless items with medial voiceless obstruents were slightly lower than with medial nasals, but this difference was not statistically significant (matched pairs $t(36) = 1.39$, $p > 0.17$, n.s.). Therefore, we will combine these two conditions (medial nasals and medial voiceless obstruents) in the further analysis.

In the case of the medial obstruents, there was a small but significant preference for medial voiceless obstruents over voiced ones (Figure 5, matched pairs $t(36) = 2.33$, $p < 0.025$). To compare the findings in the two positions (final and medial), Figure 6 shows the mean *differences* in acceptability between voiceless and voiced obstruents:

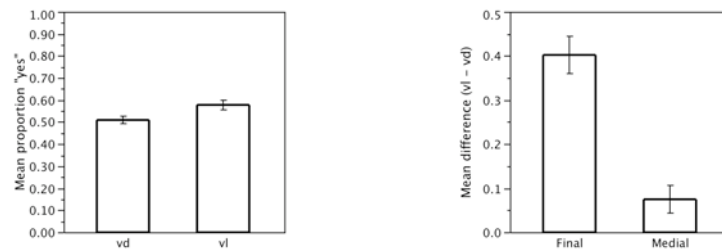


Figure 5 (left): Mean proportion of “yes” responses to words with medial obstruents and final nasals (e.g. odam vs. otam).

Figure 6 (right): Mean difference in proportion of “yes” responses between words with voiced or voiceless obstruents in final or medial position.

It is obvious from Figure 6 that the size of the voicing effect differed significantly between the two positions, being about four times larger in final position (matched pairs $t(36) = 6.04$, $p < 0.0001$), strongly indicating that there are two separate effects. Thus, we take these results to show that subjects did *not* disprefer medial voiced obstruents to the same extent that they dispreferred absolutely-final voiced obstruents, and thus that the two positions were subject to different generalizations for the participants. That is, the participants did not choose generalization (B), which would have applied equally to both medial and final obstruents.

Nevertheless, the participants did not particularly like obstruents in medial position generally. There are several potential reasons for this dispreference. The first might be a simple reticence on the part of the participants to accept items with any voiced obstruents as belonging to the language at all. That is, they could be making an inventory-based position-independent stochastic generalization. This makes sense given that only a quarter of the training words had voiced consonants at all. Secondly, in the training set, any word with a final nasal also had a medial nasal. If the subjects picked up on this regularity (which they very well may have, as nasal harmony processes exist in languages such as Guraní, Lunt 1973), they should have dispreferred any words that ended with final nasals but which lacked medial nasals. This seems in line with our data, which shows lowered acceptance of both items with final nasals and medial obstruents. It should also be noted that subjects had a general bias towards accepting words as belonging to the language, often termed a ‘yes’ bias.

It is important to reiterate the difference in effect size between the first set of contrasts in absolute final position (which confirmed that subjects had indeed learned a voicing generalization) and the second set (which unintentionally conflated our test of absolute and relative devoicing with a nasality rule). We conclude that this data offers no support for subjects’ dispreference of medial voiced obstruents, in line with the absolute-devoicing hypothesis. This in turn again reflects the attested patterns in the languages of the world.

4. Experiment 2b

4.1. Materials/Participants/Procedure

Materials were identical to Experiment 2a. Participants were 16 eight-month old infants (8 female, age range: 7;25-8;17) recruited from the area around College Park, Maryland.

The procedure was identical to experiment 1b, using the new materials. But here, as with the adults, since there were more than two comparisons we were interested in, there were several conditions required. We began with the condition that would demonstrate that infants differentiate between final-devoiced and final-voiced items (types I and II in the materials).

4.2 Results and Discussion

Contrary to our predictions, we found that infants were unable to differentiate between items ending in voiced and voiceless consonants (Figure 7, matched pairs $t(15) = 0.20$, $p > t$ -test: $p > 0.42$, n.s.).

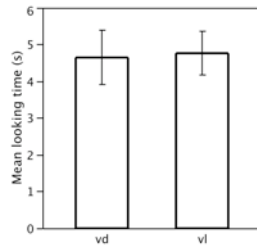


Figure 7: Mean looking time in seconds for items with absolute final voiced and voiceless obstruents.

This perhaps makes sense given the subtle nature of the acoustic cues for voicing, as opposed to the very salient stress and length differences between test items in Experiment 1. Moreover, this finding is consistent with other work (Zamuner, 2006) which has shown that even in language settings in which the infants will ultimately make meaningful differences based on voicing, such as Dutch, infants have not yet learned to perceive this rather fine-grained acoustic difference, even by 16 months of age.

We are currently doing follow-up work with a different absolute-final rule (modeled after the Korean process $s \rightarrow t / _ \#$) which will allow us to both clear up the adult results by eliminating the presence of nasals, and provide infants with a distinction that they should be able to make at 8 months.

5. General Discussion

The results of these studies support the idea that phonological processes are intrinsically connected to their environments. That is, whether a process applies

in absolute or relative final position depends on the type of process involved. Stress generalizations prefer relative finality, voicing generalizations prefer absolute finality. What's particularly intriguing about this set of linguistic facts is that it didn't have to be this way: *a priori*, there doesn't seem to be any reason that the attested rule is more 'learnable', it just happens to be what we find in the world, even though the reverse rule (e.g. Rule B in each of our experiments) is just as easy to formulate in all of our formal linguistic systems. That is, there are no obvious reasons that stress rules should apply finality relatively while obstruent devoicing rules should apply finality absolutely, yet this is precisely what we see in the languages of the world as well as in our infant and adult learners.

This finding is also consistent with recent computational approaches to phonological acquisition, in particular Heinz (2007), who posits three kinds of procedures for phonological learning: neighborhood-distinct learning for stress patterns, n-gram learning for local patterns and precedence learning for long-distance patterns. N-gram learners must be star-free (McNaughton and Papert 1971) which corresponds to the essential omission of the Kleene star operator in absolutely final generalizations, and induces a class of strictly local generalizations. Stress patterns, since they can count modulo 2 or modulo 3 (binary and ternary feet) cannot be counter-free (nor star-free) and must belong to a more powerful set of learning mechanisms. While this corresponds closely to our experimental findings separating stress and voicing, and provides a mechanism which excludes relatively final generalizations from consideration in some cases (e.g. voicing) it fails to provide a clear reason why absolute final generalizations would be dispreferred in stress learning, as n-gram learning falls within the neighborhood distinct class. Nevertheless, the convergence between computational modeling and experiments reported here is encouraging.

Even the 8-month olds in our stress experiment were aware of this seemingly arbitrary distinction. This is especially impressive given that the eight months of exposure to English that these infants have had leads them to have the opposite stress preference in their native language (Juszyk, Cutler, and Redanz, 1993; Houston, Santelmann and Juszyk, 2004), to the one that they have demonstrated here. This suggests that firstly, our familiarization had an effect on the infants in the study, and that secondly, something in the language acquisition mechanisms seems hardwired to lead infants (and even adults) to prefer phonological patterns that stem from phonological rules that occur in the world, even when subjects only have minutes of exposure to the linguistic system in question!

Moreover, as mentioned earlier, what learners in these studies (and in the world for that matter) must do is not only an issue of inducing phonological patterns from primary data. Understanding the variation in application of relative and absolute finality first requires noticing and internalizing several phonological patterns, and then abstracting over the data to realize that certain rules only apply in certain types of phonological environments. To these authors, this seems easier to posit as an innate constraint on the system, rather

than to posit that some sort of 2nd order learning is occurring directly from the input.

6. Conclusion

In conclusion, we have shown through a series of Artificial Language Learning experiments that subjects prefer novel stimuli that follow cross-linguistically attested rules. This in turn suggests that learning preferences do follow typological preferences.

References

- Berent, Iris, Lennertz, Tracy, Jun, Jongho, Moreno, Miguel, & Smolensky, Paul. (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences*, 105, 5321-5325.
- Boersma Paul, Weenink David. Praat: doing phonetics by computer (version 5.0.27) 2008; Computer program.
- Chomsky, Noam and Halle, Morris. (1968) *The Sound Pattern of English*. New York: Harper and Row.
- Gerken, L., & Boltt, A. (2008). Three Exemplars Allow at Least Some Linguistic Generalizations: Implications for Generalization Mechanisms and Constraints. *Language Learning and Development*, 3, 228-248.
- Heinz, Jeffrey. (2007) Inductive learning of phonotactic patterns. PhD dissertation, UCLA.
- Houston, Derek, Santelmann, Lynn & Jusczyk, Peter. (2004). English-learning infants' segmentation of trisyllabic words from fluent speech. *Language and Cognitive Processes*, 19, 97-136.
- Jusczyk, Peter, Cutler, Anne, & Redanz, Nancy. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, 64(3), 675-687.
- Kleene, Stephen. (1956) Representation of events in nerve nets and finite automata. In Claude Shannon and John McCarthy (eds) *Automata Studies*. Princeton: Princeton University Press. 3-42.
- Lunt, Horace. (1973) Remarks on nasality: The case of Guraní. In Stephen Anderson and Paul Kiparsky (eds) *A Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston. 131-139.
- McNaughton, Robert and Papert, Seymour. (1971) *Counter-Free Automata*. Cambridge MA: MIT Press.
- Sebeok, Thomas & Ingemann, Frances. (1961). An Eastern Cheremis Manual: Phonology, Grammar, Texts, and Glossary. *Indiana University and Mouton*, 5.
- Seidl, Amanda., & Buckley, Eugene. (2005). On the learning of arbitrary phonological rules. *Language Learning and Development*, 3-4, 289-316.