

Investigating phonotactics, lexical analogy, and sound symbolism using xenolinguistics: A novel word-picture matching paradigm

Vsevolod Kapatsinski (vkapatsi@uoregon.edu)

Department of Linguistics, 1290 University of Oregon,
Eugene, OR 97401 USA

Lamia H. Johnston (lhj@uoregon.edu)

Department of Linguistics, 1290 University of Oregon,
Eugene, OR 97401 USA

Abstract

All human languages have restrictions on sound sequences, called *phonotactic constraints*. Knowledge of phonotactic constraints is typically tested using pseudoword rating tasks, e.g., an English speaker might be asked to rate acceptability or wordlikeness of the phonotactically illegal /bnɪk/ and the phonotactically legal /bɪk/. We introduce a new method of testing knowledge of phonotactic constraints. Instead of asking subjects to rate pseudowords, we ask them to assign pseudowords to pictures of novel objects. The set of available pseudowords is larger than the set of pictures and includes both legal and illegal pseudowords. We find legal pseudowords to be less likely to be left unassigned to pictures than illegal pseudowords. Thus, the listeners show knowledge of the phonotactics of English. We suggest that the present method has important advantages over rating tasks: it is a more direct measurement of the influence of phonotactics on the lexicon, and it allows the experimenter to detect influences of sound symbolism and lexical analogy and separate them from the influence of phonotactics.

Keywords: phonology; phonotactics; sound symbolism; analogy; acceptability

Introduction

The grammars of all languages contain restrictions on possible sound sequences, called *phonotactic constraints*. For instance, despite /bnɪk/ and /bɪk/ not being actual English words, /bɪk/ obeys the phonotactic constraints of English but /bnɪk/ does not because there are no word-initial stop+nasal sequences in English. Native English speakers would also rate /bnɪk/ as being less acceptable than /bɪk/, showing that they have knowledge of the phonotactic constraints of their language (Chomsky & Halle 1965).

The phonotactic constraints are thought to place restrictions on the way the lexicon of the language can develop in the future, such that newly coined or adopted words are likely to also obey the phonotactics of the language. If a word does not obey the phonotactics of a language into which it is borrowed, it often changes to fit the phonotactics. One way this change can happen is through misperception (Ohala 1981). Berent et al. (2007), Dupoux et al. (1999), and Pitt (1998) have documented that phonotactically illegal sequences are often perceived as similar legal sequences, e.g., English listeners often perceive

natural productions of /bnɪk/ by speakers of Russian, for whom the /bn/ cluster is phonotactically legal, as having a vowel between /b/ and /n/. Thus a word like /bnɪk/ is likely to be misperceived by English speakers as /bənɪk/ and borrowed into English as /bənɪk/.

An additional, and much more controversial, way in which phonotactic constraints can influence the development of a language is by militating against the adoption or retention of phonotactically illegal words. Thus, phonotactically illegal words may be less likely to be borrowed and retained in the language than phonotactically legal words. An intriguing piece of evidence for this influence of phonotactics is provided by Berg (1998:230-233) who examines the probability of Old English words surviving into Modern English depending on the phonotactics of the initial cluster in Modern English. He finds that 803/968 (83%) words containing a phonotactically legal cluster (/kr/, or /sn/) have survived, compared to 555/774 (72%) for words containing now illegal clusters (/kn/, /gn/, and /wr/, $\chi^2(1)=31.1$, $p<.001$). He argues that “a word may pass out of the system because of phonological problems” (Berg 1998:231), suggesting that phonotactic constraints may not only force illegal words to change but also force illegal words out. A plausible mechanism for this effect is suggested by Martin (2007), who provides simulation data from neural networks showing that, as long as sublexical-to-lexical feedback is assumed, words that are phonotactically suboptimal are less likely to be selected for production than more well-formed competitors.

Knowledge of phonotactics is typically tested using rating tasks (for recent representative examples, see Bailey & Hahn 2001, Coleman & Pierrehumbert 1997, Frisch et al. 2000, in press, Shademan 2005, Treiman et al. 2000), involving a metalinguistic judgment of ‘acceptability’, ‘grammaticality’, ‘goodness’, ‘wordlikeness’ etc. However, judgment tasks offer at best an indirect way to gauge the hypothesized effect of phonotactics on lexical selection. One goal of the present paper is to develop a more direct method for examining the potential influence of knowledge

of phonotactics on lexical choice experimentally (Berg 1998, Martin 2007).

Phonotactic constraints are not the only influence on lexical selection. Two other potential factors are sound symbolism (e.g., Sapir 1929, Ultan 1978 vs. Diffloth 1994) and lexical analogy (e.g., Bailey & Hahn 2001, Shademan 2005). A word containing a consonant cluster that is never observed in English may nonetheless be selected (and receive high ratings in a judgment task) if it is sufficiently phonologically similar to an existing English word. In addition, words that contain sounds that iconically represent some aspects of their referents may be especially likely to enter the lexicon. In the present study, we focus on size symbolism, where high vowels like [i] symbolize small creatures while low vowels like [a] symbolize large ones (Sapir 1929, Ultan 1978).

Methods

40 native English speakers were recruited from the Psychology/Linguistics human subjects pool and participated for course credit. All reported being native English speakers. Each subject was presented with a Microsoft Powerpoint file containing instruction slides followed by experimental slides.

The instructions asked the subject to imagine oneself in the distant future, arriving on an unknown planet (called Terra Enigmatica) and discovering the remains of an Earth colony that was established by speakers of both English and Wilkipaengo (the language name was invented, so as to avoid the influence of knowledge regarding non-English phonotactics). The rest of the story, shown in (3), explained the importance of matching names to creatures and stressed that the lists ‘inadvertently’ included non-English names that should not be assigned to creatures.

- (3) It appears that the colony was established by speakers of both English and Wilkipaengo. Before disappearing, the colonists recorded an archive of messages.

Listening to the English, you notice some unfamiliar words. The words appear to be names for creatures common to Terra Enigmatica. According to the recordings, some creatures are benign while others are extremely dangerous and may be responsible for wiping out the entire colony!

Now you need to match the creatures you’ve encountered to the names given to them by the English-speaking colonists.

You are not interested in the Wilkipaengo names that seem to have somehow crept into your lists.

The backstory was designed to avoid the speakers treating the nonsense words as loanwords from another language,

since languages often have more tolerance of phonotactic violations in borrowings than in the native vocabulary (e.g., McCauley 1968, Pierrehumbert 2006, Schutze 2005). We also wanted to avoid asking speakers to ‘name’ the creatures believing that such an instruction would unleash the subjects’ creativity and perhaps lead them to choose the strangest-sounding words to match the strangeness of the novel creatures (although see Martin 2007 for corpus data showing that even names of characters of role-playing games produced (largely) by English speakers tend to obey the phonotactics of English). Thus, the backstory is designed to suggest to the speakers that the words to be assigned to creatures should be ordinary English words that speakers of English would be using in speech. In Schutze’s (2005) terms, we are after the “dictionary scenario” where the word is assumed to be unknown to the subject but to be a regular English word that could be found in a big enough dictionary of the right variety of the language. An important goal for future work is to determine the extent to which subjects’ behavior in the task is influenced by instructions.

The experimental slides, which followed the instruction slides, are exemplified by Figure 1.



Figure 1: An experimental slide containing draggable and playable sound files and creature animations.

When a subject came to an experimental slide, s/he clicked on ‘Play animations’, which played all creature animations simultaneously. The animations were made using Electronic Arts’ Spore™ and featured movement and animal sounds. After playing the animations, the subject would double click on the sound files of pseudowords on the left and drag the desired sound files onto the creatures they name using the computer mouse. This procedure avoids presenting subjects with orthography (see Clopper & Pisoni 2007 for a related free classification paradigm for acoustic stimuli). The subjects could listen to the sound files as much as they wanted to and could also replay creature animations if desired. They were instructed to make sure that they listened to all the words on a slide before proceeding to the next one.

There were six experimental slides, each containing six animated creatures and twelve sound files of pseudowords. Six of the pseudowords on each slide began with a consonant cluster that is phonotactically illegal in English

while six began with either a single consonant or a legal consonant cluster. Consonant cluster legality was fully crossed with vowel identity such that half of the words with legal clusters contained one vowel, and half another vowel. The vowels contained in words differed across slides, with two slides featuring [i] and [a], two featuring [u] and [æ], and two featuring [ou] and [eɪ]. The order in which vowel pairs were presented was counterbalanced: half the subjects were exposed to each of the slide sequences in (4).

(4) $i/a \rightarrow u/\text{æ} \rightarrow eɪ/ou \rightarrow u/\text{æ} \rightarrow i/a \rightarrow eɪ/ou$
 $eɪ/ou \rightarrow u/\text{æ} \rightarrow i/a \rightarrow eɪ/ou \rightarrow u/\text{æ} \rightarrow i/a$

There were two matched sets of pseudowords such that for each phonotactically illegal pseudoword there was a legal pseudoword that differed from the illegal counterpart only in the onset. All pseudowords had a (C)CVC structure. The legal and illegal counterparts were never presented to the same subject. Rather, they appeared in the same positions on the same slides but for different subjects. This was done to avoid presenting minimal pairs differing only in the (legality of) the onset and thus perhaps drawing abnormal degree of attention to phonotactics. Half of the subjects assigned to each vowel sequence order received each pseudoword set. The mappings between legal and illegal clusters are shown in (5) with numbers of word pairs exemplifying a mapping in parentheses.

(5) bd/bl (9), bn/bl (3), bn/br (1), bw/kw (1), bz/sp (3), bz/sk (1), bz/bl (1), dg/dw (3), dg/dr (5), fn/fl (3), fn/fr (1), gd/gl (3), gd/gr (3), kp/kw (6), ks/sk (3), lb/bl (4), lb/w (3), sr/fl (1), nd/dr (1), nd/pl (1), pn/pl (2), pt/pr (2), pw/pl (2), sr/sw (3), sr/tr (3), tk/tw (2), tn/tw (2)

Results and Discussion

The effect of phonotactic legality is shown in Figure 2. Phonotactically legal words were significantly more likely to be assigned to creatures than the corresponding phonotactically illegal words (by items, $t(71)=8.05$, $p<1/10^{11}$; by subjects, $t(39)=5.57$, $p<1/10^5$).

The legal/illegal pairs in which the illegal pseudoword was (unexpectedly) used less often than the legal one are drVC/dgVC ($n=4$), dwæf/dgæf, fneɪk/freɪk, kwum/kpum, and twɪs/tnɪs.

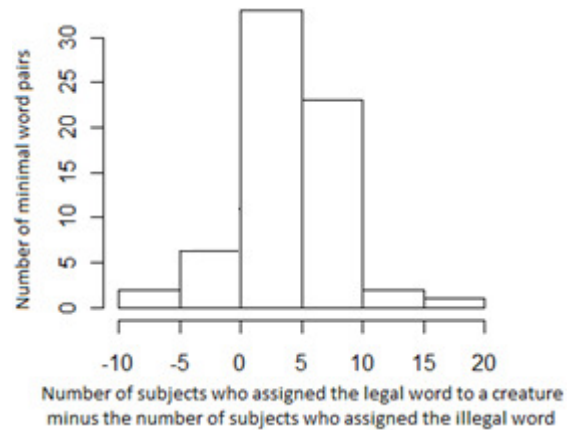


Figure 2: The effect of phonotactic legality on a word's frequency of being assigned to any creature (maximum possible difference = 20; pairs with no difference in popularity between legal and illegal words ($n=4$) not shown).

It is important to distinguish between underuse of legal clusters, which could then be argued to have been perceived as illegal by the subjects, and overuse of illegal clusters. Figures 3 and 4 show that in the present case we are dealing primarily with underuse of the legal clusters [dr] and [Cw] (all words beginning with these clusters are shown as darkened blocks in Figure 3) rather than overuse of the corresponding illegal words.

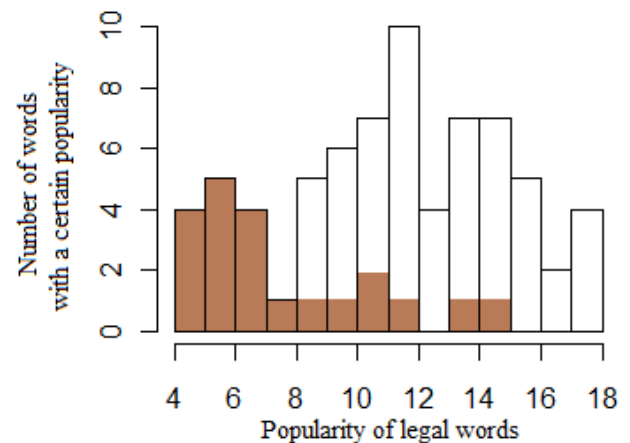


Figure 3: The distribution of popularities of legal words with the legal words beginning with /dr/ or /Cw/ shown darkened

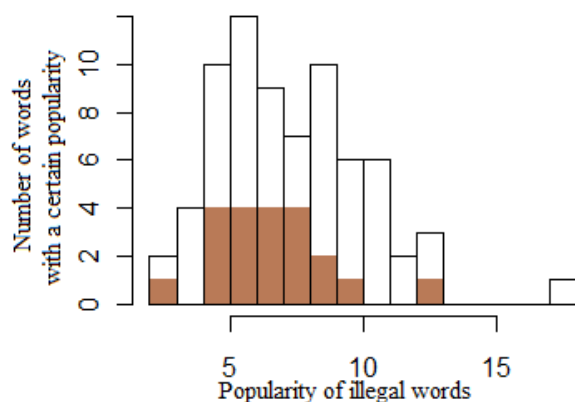


Figure 4: The distribution of popularities of illegal words. Darkened blocks represent illegal words that are minimal pairs for the legal words in Figure 3 (differing in onset cluster)

The underuse of /dr/ onsets may be due to the speaker's strong affrication of /d/ in these clusters, possibly resulting in the cluster being perceived as the phonotactically illegal cluster /dʒr/ by listeners who produce less affrication of /d/ in /dr/ (cf. Ohala 1981). The lack of preference for Cw over illegal clusters may be due to the legal clusters having a very low type frequency in English, which makes these clusters, though legal, marginal (for effects of type frequency on acceptability ratings, see Bailey & Hahn 2001, Coleman & Pierrehumbert 1997, Frisch et al. 2000, in press, Treiman et al. 2000).

Finally, the strong preference for /fneik/ over /freik/ (the former is used by 8 more subjects than the latter and is the most popular illegal word in the present study: the clear outlier in Figure 4) is likely to be an effect of lexical analogy to the word 'snake'. To assess possible effects of lexical analogy and sound symbolism, we tested whether some words might be preferentially paired with certain creatures by cross-tabulating sound files and the creatures they are paired with and looking for cells with values that are significantly higher than expected under the null hypothesis. We tested three different null hypotheses: 1) subjects are randomly pairing words with creatures within a slide (which produces a 1/12 change of assigning a word to a creature), 2) subjects randomly pair phonotactically legal words with creatures within a slide, and 3) for each slide, subjects choose a set of words to assign to creatures, and then randomly match the words within the set with creatures on the slide. With any of the three null hypotheses, there were three words that were paired with particular pictures more often than would be expected if the null hypothesis were true. The words were /fneik/, /blun/, and /blut/ (assigned to their preferred creatures 43%, 42%, and 37% of the time they were assigned to *any* creatures; $p=.0005$, $p=.0003$, $p=.0006$ respectively according to the binomial test with null hypothesis 3; the Bonferroni-adjusted critical p value is $.05/72=.0007$). The preferred creature-word

pairings are shown in Figure 5. The likely explanation for these preferred assignments is lexical analogy to the words 'snake' [sneik], 'bloom' [blum], and 'blue' [blu] respectively: the creatures in question are the only snake-like, bloom-like, and blue creatures on their slides.

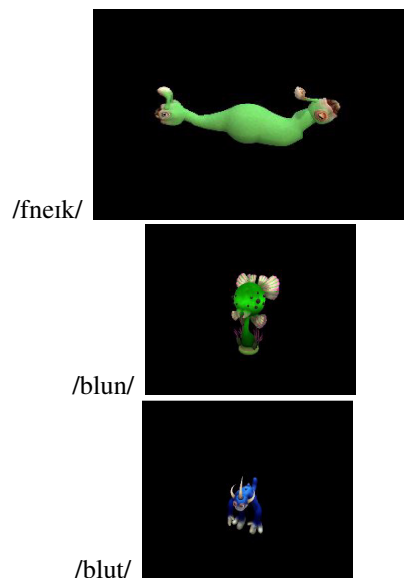


Figure 5: Non-random word-creature pairings.

Schutze (2005) objects that the "dictionary scenario" (exemplified by our backstory) is inappropriate for use in nonce probe tests of grammatical knowledge because of being particularly subject to effects of lexical analogy. The present findings confirm the presence of lexical analogy effects in the scenario. However, we do not believe this invalidates the use of the "dictionary scenario" in the present paradigm even if one believes in grammar as a cognitive module that is separate from the lexicon (Schutze 2005). Unlike in rating tasks, lexical analogy effects can be detected (and factored out) in the present task by searching for non-random picture-word co-occurrences. In order to examine possible differences between rating tasks and word-picture matching, we have conducted a wordlikeness rating task where "1" meant "not at all like English words" and "5" meant "very much like typical English words". The same pseudowords were used but no pictures were presented. We observed that [fneik] received the highest ratings out of all phonotactically illegal pseudowords. Given the results of the picture-matching task, we would argue that this result is due to lexical analogy to the word /sneik/. We would not have been able to infer this based on the rating data alone, leaving the effect unexplained.

The use of pictures in the present experiment may discourage the use of phonological analogy to existing words that are phonologically similar to the experimental pseudowords but not semantically similar to any of the pictures of the slide, e.g., the pseudoword /glog/ could be

rated highly wordlike on analogy with /grog/ or /log/ but the existence of /grog/ and /log/ might not lead the subjects to assign /glog/ to a creature because /grog/ and /log/ are not names for animals (or features of animals). This hypothesis remains to be tested.

Both rating tasks and the present paradigm are limited by the fact that phonotactically illegal sound sequences are often misperceived as phonetically similar legal sequences (Berent et al. 2007, Dupoux et al. 1999, Pitt 1998). Furthermore, as Berent et al. (2007) show, phonotactically illegal sequences are not equal in how likely they are to be misperceived. In particular, typologically marked onsets with falling sonority like /lg/ are more likely to be misperceived by English speakers than onsets with flat sonority like /bd/, which are less likely to be misperceived than clusters with rising sonority like [bn] or [pw]. While we might have expected that English listeners would judge words beginning with /lg/ to be particularly unnatural and would be unlikely to assign them to objects, the finding that such clusters are most likely to be misperceived as legal sound sequences (e.g., /læg/) throws a wrench into this expectation. Thus, it is a priori unclear whether illegal clusters strongly violating sonority sequencing should be assigned to creatures more often or less often than illegal clusters that do not violate sonority sequencing (as much).

The breakdown of onsets by sonority is shown in Figure 6. Assuming that [s] is extrasyllabic, the optimality of the sonority sequence in the onset rises from left to right.

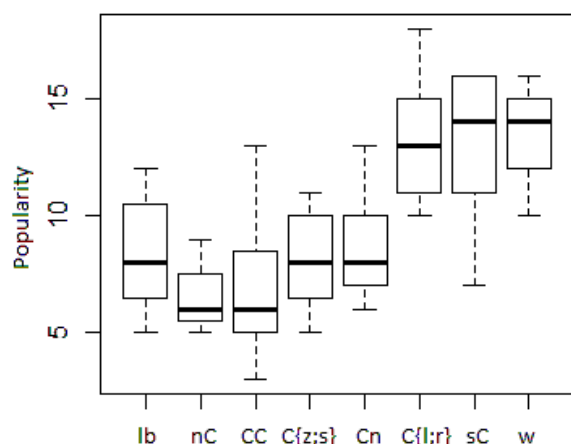


Figure 6: Frequency of being assigned to any creature as a function of sonority (C="obstruent"). This figure does not include /dr/ clusters.

There is a statistically significant difference between C{z;s;n} and C{l;r} ($W=349$, $p=.00001$). However, there is only a trend for {n;C}C clusters to be used less than C{z;s;n} clusters ($W=137$, $p=.034$, which would not reach $p_{critical}$ with the Bonferroni correction), and /lb/ clusters are assigned to creatures numerically more often than clusters that should be more acceptable according to sonority

sequencing. The effect of sonority on acceptability of illegal clusters is thus ambiguous and requires perception data for interpretation.

In future work, it appears important to supplement data from picture-word matching with data on how the stimuli are perceived by the same subjects. We expect that subjects who often misperceive an illegal cluster as a related legal sequence should be more likely to assign words containing the cluster to pictures of novel objects. Nonetheless, the presence of the effect of phonotactic legality in the present data as well as in rating studies of phonotactics (Bailey & Hahn 2001, Coleman & Pierrehumbert 1997, Frisch et al. 2000, in press, Treiman et al. 2000) shows that the perceptual mechanism of repairing phonotactically illegal sequences does not succeed in repairing the sequence 100% of the time, leaving room for speakers to choose between borrowing or retaining phonotactically legal and illegal pseudowords, thus repairing phonotactic violations on the lexical level (Berg 1998). The imperfection of perceptual repair is what allows rating studies as well as the present method to assess knowledge of phonotactics.

Following the completion of all experimental slides, we asked subjects to review all creature animations and rate the creatures' size and cuteness. Subjective and objective (height, width, area, thickness) measures of the size of a creature, the height of F0 in the creatures' vocalizations, and ratings of creature cuteness did not correlate with the presence or absence of any segments or segment features in the words subjects assigned to the creature (all $p>.1$). Thus, size sound symbolism did not seem to play an important role in this experiment. We hypothesized that this may be due to the presence of many dimensions other than size in the visual stimuli. Figure 7 presents the results of an ongoing follow-up study. Thus far 7 subjects have been asked to name ten (5 big, 5 small) monochromatic 2-dimensional creature pictures using 20 words (half phonotactically illegal, half containing [i] or [u], half containing [a] or [au]). As Figure 7 shows, words with high vowels tended to be assigned to small creatures while words with low vowels tended to be assigned to large creatures ($\chi^2(1)=8.21$, $p=.004$). Thus, size sound symbolism effects may be observed in the present task when size is a salient dimension of variation for the presented objects.

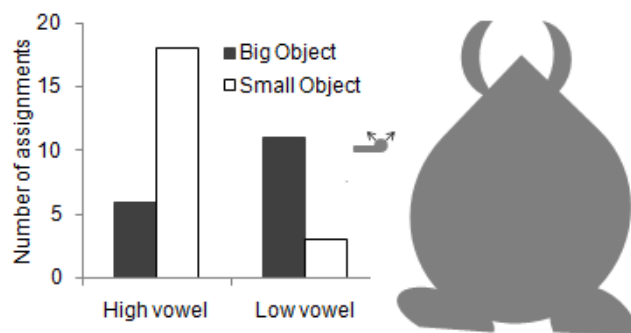


Figure 7: An effect of size sound symbolism with simpler creatures.

Conclusion

Asking subjects to match a set of pseudowords with a smaller set of novel objects provides a new way to assess the subjects' knowledge of phonotactics. This method provides important advantages over the traditional method of assessing knowledge of phonotactics (acceptability or wordlikeness ratings). First, the proposed method is a much more direct way of assessing the influence of phonotactics on lexical selection (found to operate in historical data by Berg 1998 and Martin 2007). Second, the method facilitates separating out and investigating the effects of lexical analogy and may restrict the occurrence of lexical analogy to words that are semantically related to the pictures, although analogies based on such words may be more likely in the present task than in rating. The method may also be profitably used to examine the effect of sound symbolism and how it competes with phonotactics.

This task does share some shortcomings with rating tasks. First, it requires somewhat accurate perception of illegal clusters. Given the evidence that phonotactically illegal sequences are often misperceived as similar legal sequences (e.g., Berent et al. 2007, Dupoux et al. 1999, Pitt 1998), the present task should ideally be followed by an assessment of the same subjects' perception of the stimuli. This might be accomplished using discrimination, transcription or identification tasks, or testing for the presence/absence of identity priming between the similar-sounding legal and illegal sound sequences (Berent et al. 2007, Dupoux et al. 1999, Pitt 1998). Second, the present instructions still require subjects to explicitly judge whether or not the presented words could be words of English. Future work should investigate the importance of this instruction.

Finally, the principal disadvantage of the present task compared to rating is that subjects perform the task much more slowly than a comparable rating task (the subjects in the word-picture matching version of the present task took on average 15 minutes to go through the 72 words, while a rating task using the same words took only 3 minutes). A possible way to reduce the time demands is to present fewer words and pictures per slide, thus simplifying the decision. The principal potential disadvantage of such a move is a reduction in the possibilities for detecting effects of lexical analogy due to an even more restricted set of referents to be assigned to the words.

References

- Bailey, T., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory & Language*, 44, 568–91.
- Berent, I., Steriade, D., Lennertz, T., & Vaknin V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, 104, 591–630.
- Berg, T. (1998). *Linguistic structure and change: An explanation from language processing*. Oxford: Oxford University Press.
- Chomsky, N., & Halle, M. (1965). Some controversial questions in phonological theory. *Journal of Linguistics*, 1, 97–138.
- Clopper, C., & Pisoni, D. B. (2007). Free classification of regional dialects of American English. *Journal of Phonetics*, 35, 421–38.
- Coleman, J. S., & Pierrehumbert, J. (1997). Stochastic phonological grammars and acceptability. In *Third Meeting of the ACL Special Interest Group in Computational Phonology* (pp. 49–56). Somerset, NJ: Association for Computational Linguistics.
- Diffloth, G. (1994). i: big, a: small. In L. Hinton, J. Nichols, & J. J. Ohala, eds. *Sound symbolism*. Cambridge: Cambridge University Press.
- Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception & Performance*, 25, 1568–78.
- Frisch, S. A., Brea-Spahn, M. R., & Orellana, C. I. (In press). Metalinguistic judgments of phonotactics by bilinguals. *Laboratory Phonology*, 11.
- Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory & Language*, 42, 481–96.
- Martin, A. (2007). *The evolving lexicon*. PhD Dissertation, UCLA.
- McCauley, J. (1968). *The Phonological Component of a Grammar of Japanese*. The Hague: Mouton.
- Ohala, J. J. (1981). The listener as a source of sound change. *Chicago Linguistic Society*, 17-2, 178–203.
- Pierrehumbert, J. (2006). The statistical basis of an unnatural alternation. *Laboratory Phonology*, 8, 81–107.
- Pitt, M. A. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & Psychophysics*, 60, 941–51.
- Sapir, E. (1929). A study in experimental symbolism. *Journal of Experimental Psychology*, 12, 225–39.
- Schutze, C. T. (2005). Thinking about what we are asking speakers to do. In S. Kepser & M. Reis, eds. *Linguistic evidence: Empirical, theoretical, and computational perspectives*. Berlin: Mouton de Gruyter.
- Shademan, S. (2005). Is phonotactic knowledge grammatical knowledge? *Proceedings of the West Coast Conference on Formal Linguistics*, 25, 371–9.
- Treiman, R., Kessler, B., Knewasser, S., Tincoff, R. & Bowman, M. (2000). English speakers' sensitivity to phonotactic patterns. *Laboratory Phonology*, 5, 269–82.
- Ullman, R. (1978). Size sound symbolism. In J. H. Greenberg, C. R. Ferguson, & E. A. Moravcsik, eds. *Universals of human language*, vol. 2. Stanford: Stanford University Press.