

IN THE BEGINNING THERE WERE THE WEIRD: A PHONOTACTIC NOVELTY PREFERENCE IN ADULT WORD LEARNING

Lamia Haddad Johnston & Vsevolod Kapatsinski

University of Oregon, USA

lhj@uoregon.edu; vkapatsi@uoregon.edu

ABSTRACT

It has been argued that words that contain difficult-to-pronounce sound sequences may be avoided in production, causing words with difficult phonotactics to drop out of the language at a disproportionate rate. We argue that there is also an opposing pressure favoring phonetically unusual words. We show that, at least for adults, word learning is more successful for words with unfamiliar phonetic properties to the listener. After a ten minute ambiguous training session where two novel objects were presented with an audio recording of nonce words, subjects were tested on their memory of the creatures' "names". The results show a preference for words that contain illegal word-initial consonant clusters over words that obey the subjects' native language phonotactics.

Keywords: phonotactics, lexicalization, cross-situational word learning

1. INTRODUCTION

Constraints on permissible sounds and sound sequences have been argued to influence the selection of words for production [2, 4, 5, 12, 15, 16]. When faced with a choice between two near-synonymous words, both children and adults are argued to make their choice in part based on phonotactics, choosing the more phonotactically-optimal word. This claim is supported by evidence from language change, where words with universally dispreferred phonotactics are replaced by unrelated words at a relatively high rate (see [4, 12]). It is also supported by data from first language acquisition, where children tend not to attempt words that contain sounds that are difficult for them to produce, as documented in [2, 15, 16]. There is currently no evidence for an influence of phonotactics on word *choice* in adult speech production. However, there appear to be phonological influences on the production vs. omission of the complementizer *that*, (avoiding *that that* [21] and avoiding stress lapses [10])

supporting an effect of phonology on "upstream" processing stages. These results support the prediction that phonotactically suboptimal words should be more likely to drop out of use.

However, phonotactic "strangeness" may also be helpful. In particular, Storkel, et al. [20] propose that word learning involves recognizing that a novel word is heard, which triggers learning. Recognizing that a new word is heard is easier when the word is phonotactically unusual, thus phonotactically-illegal words may actually be easier to learn. Storkel, et al. [20] exposed adults to 8 novel words for novel objects in the context of a story and tested how well the words were learned using a picture-naming task. She found that, consistently with her prediction, the subjects learned a higher proportion of phonotactically improbable words than phonotactically probable ones. Importantly, the picture naming task involves production pressure to avoid phonotactically improbable words, yet the subjects were found to prefer to produce words that are phonotactically improbable. This result is counter to findings with children [2, 15, 16, 18], although cf. also [19].

The present study provides further support for the low-phonotactic-probability advantage in word learning by showing that even phonotactically illegal words are preferred in adult word learning, at least in the early stages, as predicted by Storkel, et al. [20]. We use the cross-situational word learning paradigm [22]. In this paradigm, every training trial consists of multiple pictures paired with multiple spoken words. Thus, within a trial, the mappings between words and referents are ambiguous. However, across trials, the mappings can be disambiguated because the same word always co-occurs with the same referent. Unlike in [20], where the novel words were embedded in English sentences, making it particularly important to pay attention to cues of unfamiliarity, all words in the present study are novel and need to be learned, thus cues of unfamiliarity could be argued to be less important.

We presented adult listeners with the word-picture mappings for a very short time, such that the mappings were not yet fully learned. Some of the pictures had names that were phonotactically legal in English, while others had names that were phonotactically illegal. The subjects' guesses at the pictures' names were then examined to determine whether they had a bias in favor of legal vs. illegal names. A bias in favor of illegal names is found.

2. METHODS

16 native English speakers participated in a short training session where they were instructed to learn the names of novel creatures. The creatures were 2D monochromatic shape combinations created in Powerpoint and were either small or large. They were presented in random pairs to each subject accompanied by the audio of nonce words that were intentionally ambiguous in which creature they applied to. The audio stimuli were monosyllabic words, half of which contained word initial consonant clusters that do not exist in English ([bn], [pn], [tn], [fn], [bz], [ks], [kt], [sr], [tl], [gd], [bd], [nd], [lb], [kt]), while the other half obeyed English phonotactics (starting with [bl], [br], [pl], [pr], [tr], [dr], [sw], [fr], [sk], [sp], [st]). For each of the illegal words, there was a corresponding legal word that shared the rime. Additionally, each set of illegal and legal words were divided in half again for vowel type—half contained the high front vowel [i] and half, the low vowel [a]. This allowed us to test the effect of size sound symbolism, which we do not focus on in the present paper. All of the words were produced by a native speaker of Russian (cf. [3]), since Russian tolerates the clusters that were phonotactically illegal in English. There were 32 distinct words, with each word presented 6 times during training.

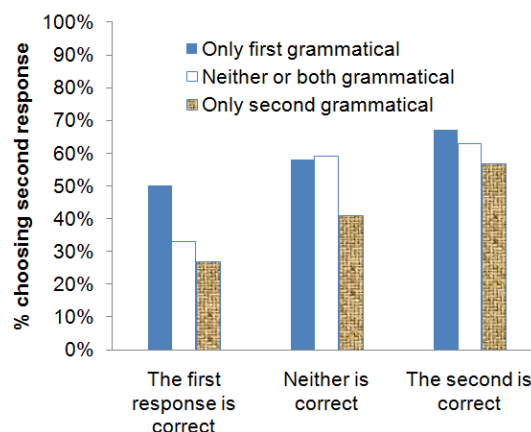
Following the training session, subjects were tested on the names of the creatures. The subjects would see one creature and hear two words, both of which were paired with the creature during training. On two thirds of the trials, one of the words was a word that had been *consistently* paired with the creature (on all 6 occasions the creature was experienced) and thus was the creature's name. Half of the trials were fillers for the purposes of examining the effect of phonotactics. On these trials, either both words were phonotactically legal, or both were illegal. On the experimental trials, one of the words was illegal while the other one was legal. The legal word was the name of the

creature half of the time. In addition, two thirds of the trials could be decided on the basis of sound symbolism: these were trials in which one of the words had the vowel [i] while the other had the vowel [a]. Phonotactic legality, sound symbolism and correctness were fully crossed factors. There were trials on which one could decide only on the basis of sound symbolism, trials on which one could decide only on the basis of legality, trials on which one could only decide on the basis of form-picture pairing ("lexical correctness") and trials on which more than one factor (e.g., phonotactic legality and lexical correctness) could be used as the basis for deciding. The subjects responded by button press. There were 48 trials total.

3. RESULTS

We analyzed the results of trials on which the two words differ on phonotactic legality (the filled bars in Figure 1) using a logistic mixed effects model [1], as implemented in R [14] with subjects and words as random effects and phonotactic legality, sound symbolism, and lexical correctness as fixed effects. There was no significant effect of sound symbolism. There was a significant main effect of phonotactic legality ($z = 2.63$, $p = .009$) and an interaction between phonotactic legality and lexical correctness ($z = 2.09$, $p = .037$). As shown in Figure 1, phonotactically illegal words were chosen more often than phonotactically legal words, although this effect was weaker for cases in which the second word was the correct response based on word-picture matching during training (due to the fact that when the first word is correct one does not have to listen to the second word).

Figure 1: The effects of phonotactics and lexical access on word choice.



4. DISCUSSION

The results show a sizeable effect of phonotactic legality on word choice: phonotactically illegal words were chosen more often than phonotactically legal words. This supports the hypothesis that words with unusual phonological characteristics are easier to map onto referents (as proposed in [20]).

Phonotactically unusual words are difficult to produce, which leads to their avoidance in children's productions [2, 15, 16] and may cause them to be especially vulnerable to being replaced in language change [4, 11]. However, they are also more noticeable, which may explain why they are overrepresented in parental reports of words children produce [19]. They attract attention, which may make them easier to learn to map onto referents [20].

Given that being phonotactically unusual has both advantages and disadvantages, we may expect it to help or hurt learning depending on the learner and where one is on the learning curve. It is well known that familiarity preferences give way to novelty preferences during the timecourse of learning and development [7, 23]. This may make adults more likely to prefer the phonotactically unusual compared to children. In addition, child learners may be more likely to be hurt by phonotactic "weirdness": their speech production systems are yet immature and the production costs of phonological weirdness may be too high to bear. Adults, on the other hand, have much more precise control over their speech production system [17] and thus may be helped by phonotactic weirdness. While the present study minimized production costs by asking for a button-press response, there is evidence that phonotactic weirdness helps word learning in adults even when they are required to respond by pronouncing the word [20], although the phonotactic constraints involved are relatively weak and gradient in character. In a follow-up experiment to the present study, we attempted to change the direction of the effect of phonotactic weirdness by requiring subjects to type in their chosen words after making the button press under strong time pressure (a 2-second deadline). We have thus far been unsuccessful: adults continue to prefer phonotactically illegal words even when asked to type them in. Future work should examine stronger constraints, like the prohibition against onset velar nasals in English.

Phonotactic weirdness may be particularly helpful before the word-meaning mappings are firmly established. Suppose that there are two words competing for lexical selection for production. According to models of speech production (e.g., [5]), there are multiple possible influences on word choice in this situation including the feed-forward activation from semantics and the feedback activation from sublexical phonological units. Other things being equal, feedback favors the phonotactically legal. However, if the meaning-form mappings are not yet completely acquired and they are being strengthened faster (per number of exposures) for phonotactically unusual words, then phonotactically unusual words may be more likely to be selected until the strengths of form-meaning mappings to both of the competing words reach ceiling, and the speaker knows that there are two equally faithful ways of expressing the same meaning.

The present study shows that even phonotactically illegal words can be preferred over phonotactically legal ones. A possible counterargument is that phonotactically illegal words are likely to be misperceived as acoustically similar phonotactically legal words [3]. Thus, the subjects may be unaware that the illegal words are illegal. They may simply be more distinctive because English speakers are likely to perceptually repair initial illegal clusters by inserting epenthetic schwas (bnik → bənik) [3]. However, this is unlikely to explain the data. In a separate experiment, we asked native English speakers to assign the same stimuli to creatures with the constraint of only assigning names that could be English. Unlike in the present study, the subjects avoided the phonotactically illegal names [8]. Thus, the subjects know that the illegal words used in the present experiment are phonotactically illegal, or at least unlikely to be English.

This raises a second issue, which is whether the words learned in the present study (as well as [20]) are incorporated into the native lexicon or segregated into a separate lexical module. Previous studies with similar tasks [9, 11] support the latter possibility: at least on the first day of training, phonological neighbors residing in the native lexicon do not measurably interfere with recognition of the newly-learned words. However, development of distinct lexical sub-modules may also occur within a language, as suggested by the

existence of distinct lexical strata obeying different phonotactic generalizations in many languages, including English (e.g., [6, 13]).

Thus, it is unclear whether the weirdness advantage would manifest itself in both first-language and second language learning, or if it is uniquely relevant to second-language learning. Phonotactic weirdness may be especially helpful for second-language word learning because words that are phonotactically unusual for the first language of the learner are distinctively foreign. However, the weirdness advantage may also plausibly be present in first-language learning, and especially likely to outcompete production pressure for adults and bilinguals. If this hypothesis is correct, it would provide an explanation for the development of lexical strata, and the resulting diachronic weakening of phonotactic constraints.

5. CONCLUSION

In a word learning task, phonotactically illegal words are more memorable, despite the same words being unlikely to be assigned to the same creatures in a name-invention task [8]. These results suggests a “weirdness advantage” in the early stages of word learning where word-meaning mappings are learned faster for words that are phonotactically unusual. This advantage may compete with production pressure against phonotactically-illegal words, with the two competing pressures having different weights for adults and children, bilinguals and monolinguals. Speakers who have more flexible articulatory systems may then favor phonotactically unusual words and be primarily responsible for such words entering their native language.

6. REFERENCES

- [1] Baayen, R.H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- [2] Becker, M. 2010. *Target Selection in Error Selective Learning*. Paper presented at *Computational Modeling of Sound Pattern Acquisition* Edmonton, Uni. of Alberta.
- [3] 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.
- [4] Berg, T. 1998. *Linguistic Structure and Change: An Explanation from Language Processing*. Oxford: Oxford University Press.
- [5] Dell, G. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93, 283-321.
- [6] Giegerich, H.J. 1999. *Lexical Strata in English: Morphological Causes, Phonological Effects*. Cambridge: Cambridge University Press.
- [7] Hunter, M.A., Ames, E.W. 1988. A multifactor model of infant preferences for novel and familiar stimuli. In Lipsitt, L.P. (ed.), *Advances in Child Development and Behavior*. New York: Academic Press.
- [8] Kapatsinski, V., Johnston, L.H. 2010. Investigating phonotactics using xenolinguistics: A novel word-picture matching paradigm. *Proc. of the Annual Conference of the Cognitive Science Society* 32, 2010-2015.
- [9] Leach, L., Samuel, A. 2007. Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology* 55, 306-53.
- [10] Lee, M.W., Gibbons, J. 2007. Rhythmic alternation and the optional complementiser in English: New evidence of phonological influence in phonological encoding. *Cognition* 105, 446-456.
- [11] Magnuson, J.S., Tanenhaus, M.K., Aslin, R.N., Dahan, D. 2003. The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General* 132, 202-227.
- [12] Martin, A. 2007. *The Evolving Lexicon*. Ph.D. Dissertation, UCLA.
- [13] Mohanan, K.P. 1986. *The Theory of Lexical Phonology*. Dordrecht: Reidel.
- [14] R Development Core Team. 2009. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available from: <http://www.R-project.org>.
- [15] Schwarz, R.G., Leonard, L. 1982. Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language* 9, 319-336.
- [16] Schwarz, R.G., Leonard, L., Loeb, D., Swanson, L. 1987. Attempted sounds are sometimes not: An expanded view of phonological selection and avoidance. *Journal of Child Language* 14, 411-418.
- [17] Smith, A., Goffman, L. 1998. Stability and patterning of speech movement sequences in children and adults. *Journal of Speech, Language, and Hearing Research* 41, 18-30.
- [18] Storkel, H.L. 2001. Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, & Hearing Research* 44, 1321-37.
- [19] Storkel, H.L. 2009. Developmental differences in the effects of phonological, lexical, and semantic influences on word learning by infants. *Journal of Child Language* 36, 291-321.
- [20] Storkel, H.L., Armbruster, J., Hogan, T.P. 2006. Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, & Hearing Research* 49, 1175-1192.
- [21] Walter, M.A., Jaeger, T.F. 2008. Constraints on English that-drop: A strong lexical OCP effect. *Chicago Linguistic Society* 41, 505-519.
- [22] Yu, C., Smith, L.B. 2007. Rapid word learning under uncertainty using cross-situational statistics. *Psychological Science* 18, 414-420.
- [23] Yurovsky, D., Hidaka, S., Yu, C., Smith, L.B. 2010. Linking learning to looking: Habituation and association in infant statistical language learning. *Proc. of the Annual Meeting of the Cognitive Science Society* 32, 1589-1594.