PHONOLOGICAL GRAMMARS EVOLVE TO PRESERVE INFORMATION AT WORD BEGINNINGS

ANDREW WEDEL [*1], ADAM USSISHKIN[1], ADAM KING[1], JAYCIE MARTIN[1], JOHN GEARY[1]

[*]Corresponding Author: wedel@email.arizona.edu
[1]Department of Linguistics, University of Arizona, Tucson, USA

1. Introduction

Listeners identify words incrementally at the sub-lexical level, continually updating hypotheses about the identity of the word as its phonetic signal unfolds. Because information is integrated incrementally, phonetic cues earlier in the word contribute more information on average to lexical identification than phonetic cues later in the word (e.g. Van Son & Pols 2003). For example, the initial [v] in the English word *vacuum* is highly informative, while the final [m] contributes less, because *vacuum* is already likely given the preceding signal [vækju-]. Recent work suggests that lexicons evolve to take advantage of this inherent bias, preferentially allocating more informative segments toward word-beginnings where they can contribute more to lexical access (King 2017; King & Wedel in prep; Meylan & Griffiths 2017). The phonetic form of words is not static however: all languages have a set of *phonological rules* which predictably modify pronunciation in context. As an example, all stop consonants in German are devoiced word-finally, such that 'Hund' *dog* is pronounced 'Hun[t]'.

Evidence suggests that phonological rules develop in a language when there are consistent biases on how speakers pronounce sounds in context. Because phonetic cues are on average most informative at word beginnings, and speakers pronounce informative phonetic cues more carefully than less informative cues (e.g., Aylett & Turk 2004; Wedel, Nelson & Sharp, to appear), we predict that languages should be less likely to evolve phonological rules which reduce lexical information at the beginnings of words. Here, we investigate this question through a statistical analysis of a cross-linguistic dataset of phonological rules.

2. Methods and Results

We assembled a genetically and areally-diverse dataset of phonological rules coding whether the rule modifies the beginning or end of a word. The dataset contains 266 rules from 50 languages (Figure 1). We dichotomously coded each rule in the dataset for whether the rule is phonemically neutralizing (i.e., whether it potentially creates homophones). Two patterns appear in this dataset: (i) there are significantly more rules overall that modify the ends of words, and (ii), this end-bias is significantly stronger for neutralizing rules (Figure 2). Both patterns are statistically significant in a mixed-effects regression model including Language, Family and Area as random intercepts. We further investigated two potential confounds that could explain this data: (i) a well-known bias toward rules that modify syllable-final consonants, and (ii) the fact that suffixing-dominant languages are more common; given that phonological rules often arise at stem-affix boundaries, this could potentially account for an apparent end-bias in a sample of languages. However, we show that neither of these alternative explanations can account for the apparent bias for contrast-reducing rules to target the ends of words.

These results provide the first statistical evidence supporting the hypothesis that languages evolve phonological grammars which preferentially preserve initial lexical material, which tends to be most informative in communication. More broadly, this finding contributes to the increasingly sophisticated body of evidence that language structures evolve under conflicting biases toward accurate transmission of meaning and effort reduction (e.g., Zipf 1949; Lindblom 1990; Piantadosi et al. 2011; Wedel et al. 2013; Futrell et al. 2015).

Figure 1. Areal language distribution   Figure 2. Rule type by word edge
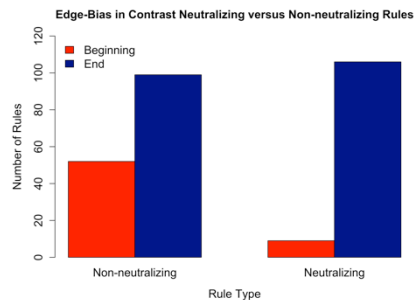
References

Aylett, M. & Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47: 31-56.

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112: 10336-41.

King, A. (2017). Word Informativity Affects Segmental Informativity. Unpublished Masters Thesis. University of Arizona, USA.

King, A. & Wedel, A. (in prep). Greater word informativity predicts greater word-initial segment informativity.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. *Speech production and speech modelling*, ed. by W.J. Hardcastle & A. Marchal, 403-39. Dordrecht: Kluwer.

Meylan, S. C. & Griffiths, T. L. (2017). Word forms - not just their lengths - are optimized for efficient communication. Unpublished manuscript.

Wedel, A., Nelson, N.R., & Sharp R. (to appear). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language*.

Piantadosi, S. T, Tily H. & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences* 108: 3526-29.

Van Son, R. J. J. H. & Louis C. W. Pols. (2003). How efficient is speech? *Proceedings of the Institute of Phonetic Sciences* 25: 171-84.

Wedel, A., Kaplan, A. & Jackson, J. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128: 179-86.

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Addison-Wesley.