

Modeling Frequency Effects in Mandarin Zero-onset Variation

Shuo Zhang
Georgetown University

Coetzee and Kawahara (2013) proposed an augmented model of noisy Harmonic Grammar [2] to incorporate the contribution of usage frequency to the application of variable processes. In this model, the influence of frequency is incorporated by scaling the weights of faithfulness constraints up or down for words of different frequencies. Even though the superiority of this model has been demonstrated in accounting for reductive processes such as English t/d deletion, C&K(2013) left open the evaluation of its application to an augmentation phonological process such as epenthesis. In this paper, I investigate this possibility with the case of two-way variation (augmentation vs. reduction) in Beijing Mandarin Chinese (BMC) zero onsets. In BMC, coda consonants are prohibited from re-syllabifying as the onset of the following onset-less syllable. Instead, one of the two processes occurs: a "zero" onset segment, such as the glottal stop can be inserted [1], or the consonant coda of the first syllable may be deleted. A production task eliciting /C#V/ sequences in a word list was designed to determine whether there is any frequency effects on the distribution of the two processes. 9 subjects who are native speakers of BMC completed the semi-spontaneous speech-style production task, yielding 837 target tokens for statistical analysis. Word frequency bins are established by querying the online interface of Word List with Accumulated Word Frequency in Sinica Corpus 3.0. Results show a positive correlation between frequency and deletion rate ($r = .29$). Moreover, lower frequency words showed a greater amount of variation than higher frequency words. These results suggest that the low contrastive salience of the epenthesis segment makes the epenthesis rate behaving similarly to the faithful candidates in C&K(2013). Next, I apply C&K(2013)'s computational learning model on the BMC data, incorporating the frequency effects using a Noisy Harmonic Grammar with weighted scaling factor. The model yields significant improvement (86.1% squared error decrease across frequency bins) over the original NHG model on accounting for the Mandarin zero-onset data, while its behavior shows sensitivity to the binning strategy in data preprocessing. I discuss the implication of these results on modeling frequency biases in phonological variation.

Reference

- [1] Chao, Yuan-Ren. (1968). *A Grammar of Spoken Chinese*. Berkeley and LA: UC Press.
- [2] Coetzee & Kawahara. (2013). Frequency Biases in Phonological Variation. *Natural Language and Linguistic Theory*, 31: 47-89.
- [3] Li, Fangkui. (1966). The Zero Initial and the Zero Syllabic. *Language*, 42:300-2.