On the survival of Polish clusters in extemporaneous speech

Polish possesses a rich phonotactic inventory. The complexity of the phonotactic system is visible in terms of 1) cluster size (up to 6 consonants), 2) the number of cluster types (over 2400), and 3) the phonological content of the cluster (sonority rises, reversals and plateaus). A considerable subset of the clusters may be triggered by morphology (Dressler and Dziubalska-Kołaczyk 2006). Cluster compliance with sonority (a traditional diagnostic of phonotactic markedness) amounts to 75%. Moreover, approximately 85% of word types contain at least one cluster. This means that running speech is highly saturated with clusters.

The aim of this contribution is to examine the production of clusters in extemporaneous speech of Polish speakers. The study focuses on two aspects of phonotactics in particular: the role of morphology and markedness in cluster production. It is predicted that morphologically-driven clusters might be reduced less frequently than phonologically motivated ones as they serve a morphological function. Markedness is measured by means of the Net Auditory Distance Principle (henceforth NAD, Dziubalska-Kołaczyk 2014). The NAD Principle formulates well-formedness conditions for clusters of various lengths and word positions. Cluster quality (preferred or dispreferred) is based on three criteria of consonant description: manner and place of articulation as well as the distinction between an obstruent and a sonorant in a sequence. The condition for a preferred word-initial or -final CC cluster states that the distance between the two neighbouring consonants (C_1C_2) must be greater than (or equal to) the distance between the vowel and the neighbouring consonant. Clusters which violate the NAD condition(s) are considered dispreferred and predicted to undergo simplification more frequently.

The empirical data were extracted from *The Dictionary of phonetic variability of the contemporary Polish language* (Madelska 2005), which comprises phonetic transcriptions of words from conversations with 30 Polish speakers. Clusters were tagged according criteria such as cluster position in a word, size, cluster goodness as measured by NAD, the presence of a morphological boundary in a cluster, the influence of the following vowel (the presence of homorganic environment), word length, word stress and word frequency. A mixed effects logistic regression model was fitted to the dataset with the *lme4* package (Bates et al. 2015) in R (R Core Team 2018) to examine factors affecting cluster rendition with word and cluster as random effects, and cluster position, size, NAD status, morphological boundary, word length, homorganic environment and word frequency as fixed effects. The analysis confirmed statistical significance of all the variables. Word-initial clusters are modified more frequently than word-medial but less frequently than word-final sequences. Cluster modification rate increases alongside cluster length and word frequency. Preferred clusters undergo modification less frequently. Homorganicity of the (CC)CV sequence is strongly disfavoured. Cluster modification rate is proportional to word length. Unexpectedly, morphonotactic clusters are simplified more frequently than phonological clusters. It is hoped that the present study will constitute a valuable addition to the description of spoken phonotactics of Polish.

Word count: 482;

Keywords: phonotactics, spontaneous speech, Polish

References

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using lme4.”, *Journal of Statistical Software* 67 (1): 1-48.

Dressler, Wolfgang U. and Katarzyna Dziubalska-Kołaczyk. 2006b. “Proposing Morphonotactics”, *Rivista di Linguistica* 18 (2): 249-266.

Dziubalska-Kołaczyk Katarzyna. 2014. “Explaining phonotactics using NAD”, *Language Sciences* 46: 6-17.

Madelska, Liliana. 2005. Słownik wariantywności fonetycznej współczesnej polszczyzny [A dictionary of phonetic variability in contemporary Polish]. Kraków: Collegium Columbinum.

R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. (https://www.R-project.org) (date of access: 12 Nov. 2018).