

Learning the Sonority Sequencing Principle in English and Korean

Jongho Jun & Nayoung Park
(Seoul National University)

The present study explores how speakers learn the Sonority Sequencing Principle (SSP), a well-known universal phonotactic generalization. Specifically, we attempt to model English and Korean speakers' gradient acceptability of unattested onset clusters based on the lexicons of their native languages.

Multiple behavioral studies on unattested onset sequences in English and Korean show that the SSP is part of native speakers' phonological knowledge (Berent et al. 2007, Daland et al. 2011 for English; Berent et al. 2008, Sung 2016 for Korean). Further, the results of learning simulations performed by Hayes (2011) and Daland et al. (2011) suggest that the SSP can be learned from the lexicons of English and Korean. However, in these previous studies, only a very small set of unattested consonant clusters were employed in testing the predictions of the learned grammars, and thus it is not yet clear whether the same predictions would be made for the rest of the set of all possible consonant clusters. In addition, no previous studies on the SSP performed learning simulations on a lexicon consisting of real, not hypothetical, Korean words, and thus it is not yet clear that the SSP can in fact be learned from the Korean lexicon.

Employing an expanded set of testing data and real word lexicons of Korean (Kwuklip kwuke yenkwuwen 1999) and English (the CMU Pronouncing dictionary), we conducted two sets of learning simulations of Maximum Entropy (maxent) phonotactic grammars (Hayes & Wilson 2008). The two sets are minimally different in that the Across-The-Board (ATB) generalizations (including, crucially, generalization from attested coda-onset $C_1][C_2$ sequences to unattested onset $[C_1C_2$ sequences) are possible in the first set, whereas a *limited* generalization, in particular, the generalization only from attested onset clusters, is allowed in the second set. The two sets of simulations are no different with respect to all the other learner settings and the training data which consist of syllabified words.

The simulation results for both Korean and English show higher correlations, between the sonority and the predicted well-formedness scores of the learned grammar, for the model with the limited generalization than the model with the ATB generalization, indicating that the limited model is more closely related with sonority in both Korean and English than the ATB model. This suggests that in order to correctly model the SSP based on the lexicon, the generalization ability of a learning model should properly be limited so that it cannot generalize from attested coda-onset (and complex coda) clusters to unattested onset clusters.

References

- Berent, Iris, Tracy Lennertz, Jongho Jun, Miguel A. Moreno & Paul Smolensky (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences* 105. 5321–5325.
- Berent, Iris, Donca Steriade, Tracy Lennertz & Vered Vaknin (2007). What we know about what we have never heard: evidence from perceptual illusions. *Cognition* 104. 591–630.
- Daland, Robert et al. (2011) Explaining sonority projection. In *Phonology* 28: 197-234.
- Hayes, Bruce (2011) Interpreting sonority-projection experiments: the role of phonotactic modeling. *Proceedings of the 2011 ICPHS*, Hong Kong, pp. 835-838.
- Kwuklip kwuke yenkwuwen [National Institute of the Korean Language] (ed.) (1999). *Standard Korean dictionary*. [Pyocwun kwuke taysacen.] Seoul: Doosan Dong-A.

Sung, Eunkyung (2016) Perception of onset clusters by English and Korean listeners: Universal markedness and L2 phonotactic knowledge. *Studies in Phonetics, Phonology and Morphology* 22.3, 477-497.