Graph-based phonological classes predict sonority projection

Max Nelson
University of Massachusetts Amherst

Introduction - The present work proposes a phonotactic model in which phones are represented by membership in learned distributional classes as opposed to traditional linguistic feature specification. This represents an implementation of an extreme version of the hypothesis that phonological representations are emergent (Mielke, 2008) which forces features to be both emergent and distributional (Mayer, 2020). Phonological classes are induced from the lexicon by constructing and partitioning phone and phonological environment similarity graphs. The resulting class systems are used to construct Maximum Entropy grammars which predict sonority projection as well or better than a feature-based baseline (Hayes and Wilson, 2008) in three test cases. Additionally, the use of a learned class system is shown to limit the constraint space, serving as an alternative to constraint induction.

Graph-based class learning - This work proposes and tests algorithms for learning phonological classes based on the idea that the distributional statistics of phones can be represented by graphs (Goldsmith and Xanthos, 2008). These algorithms work by constructing a bipartite graph in which one part consists of phones, one part consists of phonological environments, and weighted edges between them represent cooccurrence. This graph is then projected onto either the set of phoneme or environment nodes to yield a similarity graph. The three proposed algorithms represent different methods of partitioning a similarity graph into overlapping subgraphs and then using the resulting partitions to generate a phonological class system.

Sonority Sequencing - The Sonority Sequencing Principle (SSP) is a phonotactic constraint that states that syllables are well-formed when they increase monotonically in sonority from the left edge, peak at the nucleus, and then decrease monotonically towards the right edge (Clements, 1990). Sonority projection is the phenomenon by which speakers have been observed to gradiently generalize the SSP to novel sequences, in such a way that a sequence is better formed the steeper it rises in sonority from its left edge to its nucleus. Sonority projection has been argued to be a result of universal, innate bias towards SSP-conforming clusters (Berent et al., 2007), to be an epiphenomenon of statistical learning (Daland et al., 2011), and to be a combination of the two (Jarosz and Rysling, 2017).

Methods - The proposed distributional class learning algorithms are used to learn class systems in three languages in which sonority projection has been observed. The resulting phonological class systems are then used to define a constraint set for a Maximum Entropy phonotactic model (Hayes and Wilson, 2008) by creating constraints for bigrams over induced classes and their complements, which is fit to a syllabified lexicon. Results are aggregated across five random initializations for each model. As a point of comparision the Hayes-Wilson phonotactic learner (HW), which uses traditional linguistic features, is also tested.

English - In English, lexical statistics reflect the SSP. Daland et al. (2011) run an experiment in which participants' well-formedness judgements of novel forms are shown to be gradiently sensitive to SSP. They then show that the HW predicts human judgements in a way that captures the influence of SSP. The proposed algorithms and HW were fit to the English onset lexicon from Hayes and Wilson (2008) and then used to compute Harmony scores for the unattested onsets in Daland et al. (2011)'s experiment. A linear regression was then fit predicting human judgements from model Harmony scores and sonority profile and an

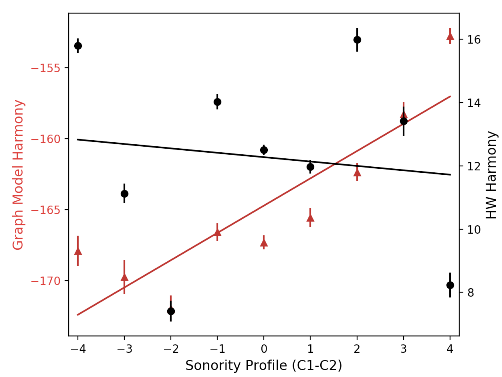
ANOVA was used to determine whether the inclusion of sonority profile yields a significantly better fit. The best model fit with learned distributional classes was shown to predict human judgements with a higher correlation coefficient than the best HW model ($r = 0.75$ vs $r = 0.62$). Neither correlation was significantly improved by the addition of sonority profile as a predictor - indicating that, in English, both the graph-based algorithms and HW are able to capture the effects of sonority profile.

Polish - In Polish, lexical statistics contradict the SSP. Jarosz and Rysling (2017) run an experiment in which Polish speakers' well-formedness judgements of novel forms are shown to be gradiently sensitive to SSP in a way that is not consistent with their lexical statistics. They also show that HW is able to predict human judgements well, but does not do so in a way that captures the effects of SSP. The proposed algorithms and HW were evaluated in the same way as in English: they were fit to the Polish lexicon compiled by Jarosz (2017) and their predictions evaluated against the test items from Jarosz and Rysing (2017)'s experiments.

The best model fit using learned distributional classes was shown to predict human judgements with a slightly lower correlation than the best HW model ($r = 0.57$ vs $r = 0.62$). However, the HW model's correlation was significantly improved by the addition of sonority profile ($p = 0.02$), replicating the results of Jarosz and Rysling (2017). The distributional class model's correlation was not significantly improved by the addition of sonority profile ($p = 0.13$), suggesting that this model captures the effects of sonority on human judgements.

Korean - In Korean, there are no complex onsets, with the possible exception of consonant-glide sequences. Berent et al. (2007) show evidence through perceptual epenthesis that Korean speakers are gradiently sensitive to sonority profile. Hayes (2011) shows that HW predicts this behavior in a toy language similar to Korean, on the condition that it has a feature set with stringent sonority representations.

The proposed algorithms and HW were fit to a Korean lexicon from Park (2021), and then used to make predictions on a test set generated by systematically combining all Korean consonants (with the exception of [ŋ]) with all vowels. Because the experiments of Berent et al. (2007) are not directly modelable with the types of MaxEnt grammars used, the success of the models is evaluated qualitatively by assessing the extent to which model judgements increase monotonically with sonority profile. In Korean, the distributional class models predict a pattern whereby well-formedness increases near-monotonically as a function of sonority profile, the HW models predictions have no meaningful correlation with sonority.



Harmony of complex onsets by sonority profile for HW and a graph-based model trained on Korean.

Discussion and Conclusions - Unlike the HW model, the class learning models here are able to use the exhaustive logically possible n-gram constraint set because the class learning algorithm yields a more restricted class system. For example, the best English model learns an average of 9.4 classes, while the feature set used by HW specified 485 classes. These results suggest that a model in which phonological representations are learned from distributional statistics may have both empirical and computational advantages over one in which phones are represented by sets of universal, innate features.