Computationally Assessing the Usefulness of Distinctive Features for Phonotactic Learning

Nicole Mirea  
nimirea@u.northwestern.edu  
Klinton Bicknell  
kbicknell@northwestern.edu  
Linguistics Department  
Northwestern University

Most models of phonotactic learning represent phonetic segments as bundles of universal distinctive features. But how helpful are these distinctive features for learning phonotactic generalizations? More specifically, can phonotactic learning occur by inferring natural class membership from distributional characteristics of phonetic segments alone, or is it necessary to specify this membership through the use of predetermined, potentially universal distinctive features, a priori of phonotactic acquisition? To answer these questions, we compare the performance of two computational models of phonotactic learning: one that has access to phonetic information in the form of phonological distinctive features at the start of the learning process, and one that does not.

The models we use are recurrent neural networks with long short-term memory units, which have previously obtained great success in language modeling at both the word [1] and character [2] levels. These networks incrementally predict the next phonetic segment in a word, using all previous phonetic segments in the word as input. Both models represent each unique segment in the segment inventory as a vector of numbers which changes with exposure to more data. Our experimental manipulation lies in how these segment vectors are initialized. For the model with access to natural classes, this is a vector of 34 distinctive features, along with 34 binary dimensions that capture whether each feature is defined or not. The model thus treats segments with more distinctive features in common as more similar to each other. In contrast, the model in the control condition initially represents segments as vectors of random numbers, which bear no phonetically grounded structural relationships among each other at the outset of training.

Our models are trained and tested on non-overlapping subsets of the English WOLEX dictionary corpus. Performance was assessed through per-word log likelihood—a measure of how likely each model judges extant words in the language that it has not seen before. We find that the model assigns significantly higher probability ($p < .05$) to these words when each segment is initialized as a random vector, meaning that it more accurately captures the phonotactic generalizations present in the training set.

In addition, we perform hierarchical clustering on phonetic segments based on the similarities between their vector representations achieved as a result of training, to determine whether the best-performing models come to encode phonetic information despite lacking initial access to it. We find mixed evidence for this: although the clusters formed using the best-performing (randomly initialized) model are qualitatively somewhat phonetically interpretable (see Fig. 1), the extent to which individual vector dimensions correspond to phonetic features is unclear. This suggests that the model is learning some other phonetic similarity structure from distributional data alone.

In summary, it seems that phonotactic acquisition can be accomplished without external, prior knowledge of phonetic distinctive features; indeed, according to our results, this knowledge may be a hindrance rather than a help. Although inference of segment-level phonotactic patterns may still benefit from access to a finer-grained phonetic specification of the speech stream, a predetermined encoding of this input in terms of distinctive features does not appear to be required for this purpose.

---

1Features were derived from Futrell, Albright, Graff, et al. [3], a model for generating phones out of their component features.
Figure 1: Dendrogram from trained vectors of best performing (randomly initialized) model in this study—the model that assigned the highest probability to the unseen words in the English WOLEX corpus. Although there is some distinction between vowels and other manner classes, these categories are not well-separated (agglomerative coefficient = 0.16).

References

