

Toward a Graded Model of English Phonology

Brent Vander Wyk (bcv@andrew.cmu.edu)

Department of Psychology and the Center for the Neural Basis of Cognition, Carnegie Mellon University
5000 Forbes Ave., Pittsburgh, PA 15213 USA

James L. McClelland (jlm@cnbc.cmu.edu)

Department of Psychology and the Center for the Neural Basis of Cognition, Carnegie Mellon University
5000 Forbes Ave., Pittsburgh, PA 15213 USA

Introduction

Traditional linguistics seeks to specify the universal and absolute properties of phonology and produce a set of inviolable rules. These rules serve to make binary distinctions between allowable forms and disallowed ones. In a thoughtful analysis by Harris (1994) such a set is described. However, this approach has difficulty accounting for graded differences in frequency among phonological forms that do not violate the rules, except to acknowledge certain preferences. For example, post-vocalic stops are more frequent in the context of a short vowel than a long vowel, though they occur with both.

The alternative approach taken in this investigation is to use a set of graded constraints to determine the frequency of a phonological form. This single mechanism captures both the binary and graded patterns assuming that the degree of concordance with the constraints is what determines the frequency. The less concordant a form is with the constraints the lower its frequency, and an unattested form is one that is extremely discordant. The approach taken here is similar to a graded version of Optimality theory (Boersma, 2000; Prince & Smolensky, 1993). In the following model, concordance with constraints is motivated by the observation that more complicated forms tend to be less frequently used than simpler forms, perhaps because as more phonetic material is added to a syllable there is an overall compression that makes articulation and perception difficult.

Graded Model

Phonotactic constraints differ in word internal and word-final contexts, and weaken both across morphological boundaries and from onset to rhyme. Rather than deal with all the sources of complexity at the outset of the investigation we limit our analysis to the rhymes of monosyllabic monomorphemic words. Furthermore, only those rhymes that contain a stop and follow principles of sonority sequencing (Harris, 1994) were considered. Vowels were categorized as either long or short with further distinctions ignored. So, rhymes under consideration have the following form: a long or short vowel followed by an optional liquid, optional nasal, or optional coronal fricative followed by a requisite stop followed by an optional coronal fricative or optional coronal stop. This yields a set of 64 possible rhymes, consisting of a vowel plus up to two additional phonemes, of which 38 are realized in English.

Each rhyme was characterized as a set of complexity-adding features, such as the presence of a fricative after the stop. The frequency of occurrence, measured as the average number of words that use the rhyme per vowel, was then predicted by a linear function which starts with a positive baseline and assesses a weighted penalty for each of the features. The weight was adjusted so that the model would predict the average number of words that use each rhyme type per vowel. Only those forms that occur in the language were allowed to contribute to weight adjustment.

The features used in the model and their final weighted penalty values were: presence of a long vowel (-5.46), stop voicing (-4.79), features to indicate whether the stop is labial (-4.75) or back (-3.22), the presence of a pre-stop homorganic nasal (-10.06), a pre-stop liquid (-13.96), a pre-stop coronal fricative (-12.93), a post stop coronal fricative (-12.80), and a post-stop coronal stop (-15.93). The positive baseline was 21.21. In terms of predicting which terms are attested in English, the model predicts a positive frequency of occurrence for 32 of the 38 that do occur, and a zero frequency of occurrence for 21 the 26 forms that do not. Among the forms that actually occur in English the model predicts 89% of the variance in the frequency per vowel.

Discussion

Previous linguistic work characterized phonology using a set of rules remaining largely unconcerned with graded patterns in the language. The goal for this investigation was to formulate a simple graded constraint model that could account for the binary distinction of which forms are attested in English and which are not, as well as account for the variations in frequencies among occurring forms. Based on the results from this simple model, continued development seems justified.

References

- Boersma, P. (2000). *Learning a grammar in Functional Phonology*: Oxford University Press.
- Harris, J. (1994). *English Sound Structure*. Oxford: Blackwell.
- Prince, A., & Smolensky, P. (1993). *Optimality theory: Constraint interaction in generative grammar*. Technical Report TR2. New Brunswick, NJ: University Center for Cognitive Science, Rutgers.