Frequency, Similarity, and Exemplars in Phonology

Stefan Frisch
Speech Research Laboratory
Psychology Department, Bloomington, Indiana, 47405
Indiana University
safrisch@indiana.edu

In this poster, I present evidence from a dictionary study of the phonotactics of the Arabic verbal roots that type frequency, phonological similarity, phonological processing, and the distribution of lexical exemplars all play a role in shaping the phonology of Arabic. The results support a model of phonological competence which uses probabilitistic constraints influenced by functional, cognitive factors.

The Arabic verbal roots are traditionally represented as morphemes consisting of abstract consonant sequences. These sequences are interleaved with morphemes consisting of vowels sequences according to morphosyntactic templates (see McCarthy 1979 for a treatment in autosegmental phonological representation). For example, the verbal root /ktb/ 'write' can be interleaved with vowel sequence to form the active (/aa/, [katab] 'to write') or the passive (/ui/, [kutib] 'to be written'). Previous research has noted that homorganic consonant pairs (e.g. /ds/) are under-represented, and that repeated sequences of identical consonants (e.g. /dd/) are not found (Greenberg 1950, McCarthy 1994, Pierrehumbert 1993).

I studied the 2676 triconsonantal verbal roots in Wehr's dictionary of Modern Standard Arabic (Cohen 1979). model of the lexicon which predicts that the expected number of occurrences of a root type is determined by the independent combination of consonants in first, second, and third position was fit to the data. The following linear regression was used:

$$O(C_1C_2C_3) = A \cdot 2676 \cdot P(C_1) \cdot P(C_2) \cdot P(C_3)$$

where $O(C_1C_2C_3)$ is the number of observed roots of the form $C_1C_2C_3$, $P(C_1)$ is the marginal probability of C_1, $P(C_2)$ is the marginal probability of C_2, $P(C_3)$ is the marginal probability of C_3. This model predicts 85% of the variation in the data.

Pierrehumbert (1993) proposed that the phonotactics of Arabic are subject to a single gradient constraint, in which homorganic consonants are restricted as a function of their similarity. Identical consonants are maximally similar, and are thus not found. Less similar consonant pairs occur more frequently. Using a similarity metric based on the phonological classification of phonemes into natural classes proposed in Broe (1993), similarity of phoneme pairs was computed for the consonants of Arabic. This similarity metric is quantitatively superior to traditional metrics of similarity based on feature counting (Frisch 1996). The model including a function of similarity accounts for 99.5% of the variation in the data.

Two additional sub-patterns in the data have been discovered. First, there is an effect of temporal order on occurrence. Roots which contain homorganic consonant pairs in C_1C_2 position are found less frequently than roots which contain analogous pairs in C_2C_3 position. I claim this is an effect of interference in phonological processing of the word. When processing C_2C_3, C_1 acts as interference, making the similarity comparison less reliable, and application of the constraint less strict. Second, I have found that roots which violate the constraint but have high expected frequency are disproportionately found, even when expected frequency is included in the model (as above). I propose that the existence of lexical exemplars which violate the constraint act as templates for other words to be formed which violate the constraint in an analogous way. Since high expected frequency forms are more likely to occur, they are more likely to act as exemplars, and become disproportionately represented among violations.

Overall, the data support a model of phonology which permits functional factors like type frequency, similarity, lexical processing, and lexical storage to influence phonological competence.

Acknowlegements

This work supported by NIH Training Grant DC00012 to Indiana University and NSF Grant BNS 9022484 to Northwestern University.

References

Broe, M. (1993). Specification theory: the treatment of redundancy in generative phonology. Unpublished Ph.D. dissertation, University of Edinburgh.

Frisch, S. (1996) Similarity and frequency in phonology. Unpublished Ph.D dissertation, Northwestern University.

Greenberg, J. (1950). The patterning of root morphemes in Semitic. *Word* 6: 162-181.

McCarthy, J. (1979). *Formal problems in Semitic phonology and morphology*. New York: Garland.

McCarthy, J. (1994). The phonetics and phonology of Semitic pharyngeals. In P. Keating (ed.), *Papers in laboratory phonology III*. Cambridge: Cambridge University Press.

Pierrehumbert, J. (1993). Dissimilarity in the Arabic verbal roots. *Proceedings of NELS 23*. Amherst: GLSA Publications.