Vowel Features in P-base

San Duanmu & Yaoyun Shi
University of Michigan, duanmu@umich.edu

Some recent studies have offered frequency rankings of distinctive features using large-scale corpora. For example, Mielke (2008) offers feature rankings based on natural classes in P-base and Clements (2009) offers feature rankings based on phonemic inventories in UPSID. In this study, we use 627 languages in P-base and examine frequency rankings of distinctive features based on *minimal phonemic contrast*, with a focus on eight vowel features [high, low, back, tense, round, nasal, creaky, murmur].

A minimal phonemic contrast is found when a pair of phonemes differs in one feature. For example, [i]-[e] differ in [high], [i]-[ɪ] differ in [tense], and [i]-[y] differ in [round]. If two phonemes differ in two (or more) features, such as [i]-[u] (differing in [back] and [round]), they do not constitute a minimal contrast and no feature is counted. In other words, our study counts features that must be assumed.

A difficulty in a contrast-based study is to interpret phonetic symbols that are ambiguous. For example, when a language has [a]-[ɑ] (and no [æ]), [a] is likely to be a front vowel, but when a language has [æ]-[a] (and no [ɑ]), [a] is likely to be a back vowel, as is the case in American English. In P-base, there are 589 [a]'s and only 32 [ɑ]'s, apparently because of typographic convenience, rather than a true preference for [-back] over [+back] in low vowels (it is worth noting that P-base gives [a] and [ɑ] the same set of feature values). Such cases require conditional interpretation of minimal contrastive pairs (e.g. [a]-[ɑ] for [back] if there is no [æ]).

We scan each of the 627 languages with over 100 minimal contrastive pairs (each feature has multiple contrastive pairs, e.g. for [high] we have [i]-[e], [u]-[o], [ɪ]-[ɛ], …). Our result shows that [high] is by far the most frequent contrastive vowel feature, occurring 1,088 times, while [low] is found only 231 times. In addition, there is a wide range of frequency differences among contrastive pairs of the same feature. For example, for the feature [high], [i]-[e] occurs 476 times, [ɪ]-[ɛ] occur 59 times, while [ɯ]-[ɤ], [ɯ]-[ə], and [ɯ]-[ʌ] occur a total of merely 15 times. Our results will be compared with those of Mielke (2008) and Clements (2009).

Clements, G. N. 2009. The role of features in phonological inventories. In *Contemporary views on architecture and representations in phonology*, ed. Eric Raimy and Charles E. Cairns, 19-68. Cambridge, MA: MIT Press.
Mielke, Jeff. 2008. *The Emergence of Distinctive Features*. Oxford University Press.