Variation in Tagalog Tapping: Word Structure and Frequency

Kie Ross Zuraw

University of California at Los Angeles

Intervocalic tapping in Tagalog, whereby /d/ becomes [ɾ] (spelled *r*), applies variably at prefix-stem boundaries (Schachter & Otanes 1972): *ma-**d**ilím* 'dark' (root *dilím*), *ma-**r**alitâ* 'poor' (*dalitâ*). Variation can occur within a morphological construction, as in the preceding examples; among derivatives of a stem, as in *ma-**d**ayà* 'be taken in', *pagpapa-**r**ayâ* 'forbearance' (*dayà*); and between pronunciations of the same word, as in *pa-**r**ayà* ~ *pa-**d**ayà* 'let oneself be fooled' (examples from English 1987).

This paper considers the influence of word structure on tapping, and of frequency on word structure. Results on alternations at morpheme boundaries in other languages (e.g., Baroni 1999 for Italian, Hay 2003 for English) predict tapping to occur when speakers treat prefix and stem as tightly integrated. I follow Hay in using frequency ratios to predict integration: Hay finds that when the frequency of a morphologically complex word (*maralita*), is high compared to the frequency of its stem (*dalita*), the word is more likely to be treated as a unit, according to various measures. (Hay explains frequency-ratio effects in terms of processing.)

This study takes frequency and tapping data from a written corpus of 20 million Tagalog words. Although typographical errors and unclear orthographic intentions pose problems, the written corpus is easier to assemble and work with than an audio corpus. To minimize hand-checking, only words from English's dictionary were used. Words were searched in tapped and untapped form, i.e., frequencies for both *madalita* and *maralita* were retrieved. A word was usable if it had total frequency (tapped + untapped) ≥15, and could be assigned to one of three categories: non-tapping (tapped variant has frequency ≤2, or untapped variant's frequency is ≥50 times tapped variant's), tapping (the converse), or varying (neither of the other two criteria applies, and each variant has frequency ≥4).

This procedure yielded 236 usable words. The log ratio of prefixed-form frequency to stem frequency was calculated for each word. As expected, the "non-tapping" group had significantly lower frequency ratios (p <.005 by one-tailed t-test) than the "tapping" group, with the "varying" group lying in between and not significantly different from the other two.

The analysis proposed is that tapping's domain is the phonological word, and frequency ratio contributes to the probability of a phonological-word boundary between a prefix and its stem: *ma-*$(dilim)_{PWd}$ vs. $(ma\text{-}ralita)_{PWd}$. Preliminary results on other Tagalog alternations are consistent with this analysis.

Intervocalic tapping has been described as obligatory at stem-suffix boundaries (no exceptions in English 1987), as in *pa-tawa**r**-in* 'forgiveness' (root *tawad*), suggesting that suffixes are obligatorily incorporated into the phonological word. The corpus, however, contains

exceptions--the mean percentage of untapped tokens per word was 6.3%. While some exceptions probably represent errors, their distribution is too regular for all to be dismissed: there is a weak but significant correlation between affixed-frequency/stem-frequency ratio and tapping rate (.316, p <.005). The interpretation adopted here is that the requirement that a suffix be incorporated into the stem's phonological word does not absolutely override the influence of frequency ratio.