Collocation frequency and the reduction of Carioca Portuguese para
Michael Gradoville, Indiana University Bloomington

Although the reduction of Portuguese para 'to, for' to pra (and several contracted forms) is certainly a well-known phenomenon, it is not particularly well understood. Many authors (e.g. Perini 2002, Thomas 1969) assert that the reduced forms are used (almost) categorically in speech, although no quantitative evidence has been presented to support these assertions. This study examines the effect of collocation frequency as well as that of phonetic factors on the reduction of para in the educated spoken Portuguese of Rio de Janeiro. Collocation frequency has been shown to affect the reduction of English know (Bybee & Scheibman 1999; Scheibman 2000), American English word-boundary palatalization (Bush 2001), Spanish word-boundary vowel sequence reduction (Alba 2006), and Spanish para reduction to pa (Díaz-Campos, Fafulas, & Gradoville 2012). Such effects demonstrate that language speakers store the more frequent collocations in memory, thereby supporting a usage-based model of linguistic representation (Bybee 2001, 2006, 2010).

Employing variationist methodology, this study uses data from the Projeto da Norma Urbana Oral Culta do Rio de Janeiro (Educated Oral Urban Norm Project of Rio de Janeiro; Callou & Lopes 1992, w/o date). The interview data from this corpus constitute 1,050,000 words from 159 different speakers. Those speakers exhibiting categorical behavior have been excluded. The data were coded for the dependent variable, namely whether the form was unreduced para or one of the reduced forms (pra, pro, pros, pras).

The independent variables considered in this study include the frequency of the two-word collocation with the preceding word (e.g. vai para, só para), the frequency of the two-word collocation with the following word (e.g. para mim, para fazer), the following sound (vowel, coronal consonant, dorsal consonant, non-lingual consonant), and the following syllable stress. The data were submitted for analysis using the statistical program GoldVarb X (Sankoff, Tagliamonte, & Smith 2005).

Of the 7415 tokens included in the analysis, some 83.3% are reduced. Both collocation frequency variables were selected as significant in predicting the reduction of para. For both variables, there is a steady trend of increased reduction, the higher the frequency of the collocation. Major exceptions to this trend are para que 'so that' and para + DEFINITE ARTICLE, which despite their high frequency have exceptionally low rates of reduction. The following sound is also statistically significant with non-lingual consonants favoring reduction, coronal and dorsal consonants disfavoring, and vowels most disfavoring. Despite an overall higher rate of reduction when para is followed by a stressed syllable, following syllable stress was not selected as significant.

The results from this study coincide with previous studies of the effect of collocation frequency on phonological variation and thus support the assertion that frequent collocations are stored as chunks in memory with all of the phonetic detail intact, providing evidence against models of language that would assert that grammar and the lexicon are distinct units.