Lexical Frequency and Syntactic Variation: A Test of a Linguistic Hypothesis
Robert Bayley, Cory Holland and Kristen Ware, University of California, Davis  
Session 3 - Frequency

The role of frequency in language variation has received a great deal of attention in recent years, particularly in phonology and second language acquisition (e.g. Bybee 2001, 2002; Ellis 2002). Recently, Erker and Guy (2010, in press) extended the analysis of frequency effects to morphosyntactic variation.

Based on data from a sample of Dominican and Mexican speakers from the Otheguy and Zentella (2012) corpus of New York City Spanish, they examined the role of frequency in variation between null and overt subject personal pronouns (e.g. yo/Ø pienso, “I think”). They hypothesized that lexical frequency, operationalized as verb forms that constitute one percent or more of the tokens in the data, would either activate or amplify the effects of other well-established conditioning factors such as switch reference and person and number, and they presented results that appeared to confirm their hypothesis.

This paper provides a further test of Erker and Guy’s hypothesis based on a larger data set of Mexican immigrant and Mexican American Spanish collected from 30 speakers in California and Texas. Based on multivariate analysis with Rbrul (Johnson 2012) of more than 8,600 tokens, we show that frequent verb forms, operationalized in the same manner as in Erker and Guy’s study, have only a small effect on overt pronoun use and that effect is opposite to what Erker and Guy predict. Separate multivariate analyses of frequent and non-frequent verb forms show that fewer constraints reach significance in the analysis with frequent verb forms only than in the analysis with non-frequent forms only. Moreover, in cases where linguistic conditioning factors reach significance in both analyses, the effects are stronger in the analysis with non-frequent than with frequent verb forms. Finally, when all tokens are combined in a single analysis, non-frequent forms are significantly more likely to co-occur with overt pronouns than are frequent forms, although the percentage of SPP use is very slightly higher among frequent forms. The mismatch between the factor weights and percentages for frequency in the combined analysis may be explained by the fact that frequent forms include few first person plural forms, which strongly disfavor SPP use, and by the fact that stative verbs, which favor overt SPP use, are overrepresented among frequent forms. The paper concludes that the rich array of linguistic constraints identified here and elsewhere (see Flores-Ferrán 2007 for a review) provides a better explanation for the observed patterning of SPP variation in Spanish than an analysis involving frequency. Finally, the results presented here suggest that the hypothesis that frequency activates and/or amplifies overt SPP use is untenable.