Numbers in the grammar?
Evidence from experiments with English relativizer and complementizer optionality

Robin Melnick, San Joe State University
Tom Wasow, Stanford University
Florian Jaeger, University of Rochester

Bresnan et al. (2007) develops a corpus-based model of multivariate probabilistically driven variation with the English dative alternation, and Bresnan (2006) subsequently (the work being chronologically later, though it came to publication earlier) shows that subject intuitions substantially correlate with the predictions of the corpus-derived model, offering evidence that knowledge of probabilistic elements is part of speaker competence. The present work extends these explorations to the highly gradient phenomena of English relativizer and complementizer (i.e. 'that') optionality, e.g. in 'We hope (that-COMP) the talk (that-REL) we give will be interesting.' While prior work offers evidence that speakers make use of knowledge of certain types of distributional information in then making probabilistic judgments, in the present work, the corpus-derived models for 'that'-optionality phenomena (Jaeger 2006, 2009) have as their most significant factors the specific percentage frequency that each matrix verb or relativized head noun has been previously seen to trigger an embedded clause. That subject judgments in the current experiments then significantly correlate with model predictions offers evidence that speaker competence may include such rather specific frequency information.

The current work employed several different experimental methodologies in probing these data, and some time in the talk will be devoted to comparative merits of various methods.

For background, the talk will begin with a brief review of some of the material underlying such explorations in probabilistic syntactic variation: syntactic factors contributing to variation, statistical procedures used in multivariate analysis (including logistic regression), and touching on the existing competence vs. usage debate.