Probabilistic variation in a comparative perspective: the grammar of varieties of English

We report on an ongoing project that applies the Probabilistic Grammar framework (e.g. Bresnan 2007) to variation within and across a range of dialects and varieties of English. Through state-of-the-art quantitative analyses of syntactic alternations in corpus data plus supplementary rating task experiments, the project seeks to explore the extent to which language users' grammatical knowledge differs across dialects and varieties of English. In treating variation as a "core explanandum" (Adger and Trousdale 2007, 274) of linguistic theory, the project contributes to the development of usage-based theoretical linguistics by adopting a variational and large-scale comparative perspective.

On the methodological plane, the project explores variability in the hidden – though cognitively 'real' – probabilistic constraints that fuel variation within and across speech communities. Such constraints, e.g. the tendency to place long constituents after short constituents (Wasow and Arnold 2003), are not necessarily tied to surface material but to subtle stochastic generalizations about language usage, which – according to experimental evidence (Bresnan and Ford 2010) – language users implicitly know about. Thus, rather than simply describing probabilistic variation in corpus data, the project aims to illuminate aspects of the linguistic *knowledge* that language users with different dialect backgrounds implicitly command. This objective is accomplished by investigating the probabilistic effects of various constraints on linguistic choice-making as a function of regional variety

The case study which we present to highlight the descriptive and theoretical benefits of the approach explores three patterns of syntactic variation in English that are sensitive to a range of constraints: the genitive alternation (the *president's speech* versus *the speech of the president*), the dative alternation (*Tom sent Mary a letter* versus *Tom sent a letter to Mary*), and particle placement (*Tom looked the word up* versus *Tom looked up the word*). These alternations we study in three regional varieties of English covered in the International Corpus of English: British English (Europe), Indian English (Asia), and Canadian English (North America). Using advanced statistical methods, such as mixed-effects logistic regression modeling and conditional random forest analysis, we will be specifically interested in (1) the extent to which we see a core grammar that is explanatory across different varieties, (2) the degree to which individual probabilistic constraints are stable as opposed to malleable, and (3) the degree to which the alternations under study exhibit cross-constructional parallelisms.

References
Adger, David, and Graeme Trousdale (2007): "Variation in English Syntax: Theoretical Implications." *English Language and Linguistics* 11: 261–78.
Bresnan, Joan. (2007): "Is Syntactic Knowledge Probabilistic? Experiments with the English Dative Alternation." In *Roots: Linguistics in Search of Its Evidential Base*, edited by Sam Featherston and Wolfgang Sternefeld, 75–96. Berlin, New York: Mouton de Gruyter.
Bresnan, Joan, and Marilyn Ford (2010): "Predicting Syntax: Processing Dative Constructions in American and Australian Varieties of English." *Language* 86 (1): 186–213.
Wasow, Thomas, and Jennifer Arnold (2003): "Post-Verbal Constituent Ordering in English." In *Determinants Of Grammatical Variation in English*, edited by Günter Rohdenburg and Britta Mondorf, 119–53. Berlin, New York: Mouton de Gruyter.