

Syntactic variation, probabilistic indigenization, and World Englishes

We report on an ongoing project combining variationist methods (e.g. Tagliamonte 2001) with the Probabilistic Grammar framework (e.g. Bresnan et al. 2007) to study syntactic variation within and across nine varieties of English from around the world: British, New Zealand, Canadian, Irish, Indian, Singaporean, Hong Kong, Philippines, and Jamaican English. Our main interest lies in the extent to which language users' grammatical knowledge displays stability and/or fluidity across these typologically and regionally diverse speech communities. In this context, we assess —among other things —the extent to which seemingly universal principles (such as end-weight) are amenable to regional differences.

Our case study explores three well-known patterns of syntactic variation in English, drawing on data from the International Corpus of English (ICE): the genitive alternation (*Marys speech* versus *the speech of Mary*), 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*). In total, our data comprise 8-10k observations of each construction, which were semi-automatically extracted from the ICE corpora: N = 10594 genitives; N = 8549 datives; N = 8072 particle verbs. We carefully circumscribe the variable contexts to identify variants in the corpora, annotating for numerous factors conditioning the choice of constructional variant. We then use multivariate techniques such as generalized linear mixed-models (Pinheiro and Bates 2000) to investigate variable effects in the conditioning factors that constrain syntactic choices. The effects of such conditioning factors, e.g. the tendency for longer constituents to follow shorter ones, can be seen as stochastic generalizations about language usage, which —according to experimental evidence (Bresnan and Ford 2010) —language users implicitly know about. Thus, we aim to illuminate the variability in the linguistic knowledge that language users with differing English backgrounds implicitly command.

Our results indicate that the direction of the influence of individual factors is largely consistent across all nine varieties, but that a few cross-varietal differences in the strength of their influence nevertheless emerge. Such differences are more likely to turn up in outer circle (non-native) varieties, e.g. Hong Kong or Indian English, than inner circle varieties, e.g. Canadian English, but there are few consistent patterns with respect to which factors differ in specific varieties. We also find that functional cognitive pressures that are universal (such as end-weight) and operative in all three word order alternations differ in their effect size among some varieties - but not consistently in the same varieties across the alternations.

The results of our study thus reflect the reshaping of stochastic patterns of internal linguistic variation due to shifting usage frequencies among speakers of post-colonial varieties, a process we refer to as probabilistic indigenization. In addition, we argue that our findings support a probabilistic model of linguistic knowledge which is mostly shaped by general, higher-level cognitive factors but also by surface level, community-specific usage norms. Our results thus carry theoretical implications for functional approaches to language structure and raise new questions for usage-based models of grammar.

References

- Bresnan, J., A. Cueni, T. Nikitina, and H. Baayen (2007). Predicting the Dative Alternation. pp. 69–94. Amsterdam: Royal Netherlands Academy of Science.
- Bresnan, J. and M. Ford (2010). Predicting Syntax: Processing dative constructions in American and Australian Varieties of English. *Language* 86(1), 168–213.
- Pinheiro, J. C. and D. M. Bates (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.
- Tagliamonte, S. (2001). Comparative sociolinguistics. In J. Chambers, P. Trudgill, and N. Schilling-Estes (Eds.), *Handbook of Language Variation and Change*, pp. 729–763. Malden and Oxford: Blackwell.