Optimizing Language Variation Analysis: Language Variation Suite

Olga Scrivner (Indiana University)
Manuel Díaz-Campos (Indiana University)
Rafael Orozco (Louisiana State University)

With advances in statistical computing and methods, there has been increasing interest in the use of advanced statistical models in sociolinguistic research. Recent variationist studies have successfully applied mixed-effects modeling, conditional and random tree analyses, and Bayesian modeling (Johnson, 2009; Tagliamonte and Baayen, 2012; Levshina, forthcoming; among others). These advanced methods enable sociolinguists to handle imbalanced data, measure individual and group variation, estimate significance, and rank variables according to their significance (Strobl et al., 2009); their implementation, however, requires some programming skills or access to statistical tools that are not always freely available. The shift in current practices in sociolinguistic data analysis has also motivated methodological debates concerning the advantages and disadvantages of traditional and novel methods. As Díaz-Campos and Dickinson (forthcoming) point out, there is a need in sociolinguistics to test new tools "for their comparability and reliability in the study of language, variation and change."

To address these issues, we have developed a user-friendly interactive application *Language Variation Suite*[1] that implements state-of-the-art statistical methods (Scrivner and Díaz-Campos, 2016). Built as a Shiny app with various statistical and graphical R packages, LVS provides access to advanced statistical methods to a broader audience, as its use does not require installation and programming skills. In the format of hands-on session, we will introduce participants to the LVS application, its structure and functions. In addition, we will discuss various statistical approaches and present novel methods of visualizing sociolinguistic data.

***Participants are encouraged to bring a laptop to work through provided sample files and instructions.

References

Díaz-Campos, Manuel and Stephanie Dickinson. Forthcoming. "Using Statistics as a Tool in the Analysis of Sociolinguistic Variation."

Johnson, Ezra Daniel. 2009. "Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis." *Language and Linguistics Compass* 3: 359–383.

Levshina, Natalia. Forthcoming. "When variables align: A Bayesian multinomial mixed-effects model of English permissive constructions." *Cognitive Linguistics*.

Scrivner, Olga and Manuel Díaz-Campos. 2016. "Language Variation Suite: A theoretical and methodological contribution for social and linguistic data analysis." *Linguistic Society of America*.

Strobl, Carolin, James Malley, and Gerhard Tutz. 2009. "An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests." *Psychological methods* 14: 323–348.

Tagliamonte, Sali and R. Harald Baayen. 2012. "Models, forests and trees of York English: Was/were variation as a case study for statistical practice." *Language Variation and Change* 24: 135–78.

[1] LVS: https://languagevariationsuite.shinyapps.io/Pages/