Quantitative Analysis with Rbrul and R

While many sociolinguists continue to perform quantitative analysis with GoldVarb, that software has been shown to be slow, limited and often incapable of providing accurate results. Today’s sociolinguists want to extend GoldVarb’s multiple regression methodology to cover data sets where independent variables are continuous numbers as well as categorical factors, and where dependent variables are either the binary choices familiar from GoldVarb or the numeric measurements obtained in sociophonetics or psycholinguistics. Researchers want to model interactions between independent variables and perhaps most importantly they want to use hierarchical (mixed--effects) models to test the venerable if dubious assumption that individual speakers and words behave identically with respect to linguistic variables. Rbrul is a program that works within the free statistical software environment R to make these improvements available with a user--friendly interface. The first section of this hands--on workshop will teach participants to analyze sociolinguistic data using Rbrul, while highlighting the similarities and differences between it and other software. For example, GoldVarb presents logistic regression results in idiosyncratic units called factor weights; all other packages report them in log--odds. Rbrul bridges the gap by presenting results in both formats, and the workshop will explain the difference. Another example is the way regression coefficients are reported for factors: sum contrasts (GoldVarb) vs. treatment contrasts (the default in most regression work). As with factor weights and log--odds, familiarity with both contrast types will allow sociolinguists to better understand work conducted in allied fields, and also enable them to make their own results more comprehensible to researchers in other disciplines. The idea of quantitative analysis within sociolinguistics has for too long almost been treated as equivalent to multiple regression carried out in GoldVarb. Since this single method should neither be the starting point nor the ending point of an analysis, the second part of the workshop will branch out from multiple regression, exploring a selection of the many other data analysis tools available in R.

In particular, the workshop will give an introduction to the graphical techniques R provides for exploratory data analysis and display. Although designed as a transitional tool Rbrul has enjoyed some success so far, being used by sociolinguists mainly in the US and UK, but also by some in Europe and in other countries.