

## **A multivariate approach to measuring change**

In a variationist perspective, language change is usually perceived as a kind of competition between variants, that is, one variant becomes more frequently used at the expense of another. Yet, it is also known that very often this cannot be adequately described as a simple, gradual replacement; rather, the determinants of the variation also change (cf., e.g. Poplack & Malvar, 2007). In order to describe a changing variation, then, researchers can track frequency developments for each factor individually (i.e. take a univariate approach), or they can compare the full picture of the variation at different points in time (i.e. compare multivariate models). The former approach can measure change but fails to capture the effect of a factor at a given point in time given the import of other factors; the latter approach can assess the effect of each factor at each point in time, but cannot assess the significance of changes in these effects.

I propose a relatively simple (and hence easily interpretable) way of combining these approaches by using the factor 'time' as a moderator variable (cf. Jaccard, 2001: 12), that is, each predictor of the (synchronic) variation is tested for its interaction with 'time'. A significant interaction means that the effect of a predictor is changing. This way, change can be measured making use of the advantages of multivariate analysis, namely that the influence of all factors is taken into account and the effects are weighed against each other.

This multivariate approach to measuring change is shown in a study of the changing variation between semi-modals (*going to*, *HAVE got to*, *want to*) and their contracted variants (*gonna*, *gotta*, *wanna*) in American English, using speech-purposed data from the Corpus of Historical American English (Davies, 2010-; for detail regarding the study see Lorenz, 2013). A logistic regression model is used, comprising numeric independent variables (string frequency, sentence length) as well as categorical ones (types of preceding and following item, type of genre), while 'time' is incorporated as a binary variable (i.e. two time periods).

It is shown that this model of change can adequately capture the complex dynamics of a changing variation over two time periods, but may reach its limits when a more detailed time measure is applied. A general limitation is that a sufficiently large amount of data is needed to usefully apply interactions; also, categorical factors with many levels pose difficulties to interpretation.

## References

- Davies, Mark. (2010-). *The Corpus of Historical American English*. <http://corpus.byu.edu/coha>
- Jaccard, James. (2001). *Interaction Effects in Logistic Regression*. Thousand Oaks: Sage.
- Lorenz, David. (2013). *Contractions of English Semi-Modal: The Emancipating Effect of Frequency*. NIHIN Studies 2013. Freiburg: Rombach.
- Poplack, Shana, and Elisabete Malvar. (2007). Elucidating the transition period in linguistic change: The expression of futurity in Brazilian Portuguese. *Probus*, 19, 121-169.