Examining chain shifts through machine prediction

The principles of chain shifts, especially following Labov (1994), require that vowels be structurally interconnected, so that changes in one vowel cause changes in another vowel. This paper presents predictive modelling by conditional inference tree as a new technique for examining structural interconnections among vowels that are proposed to be participating in a chain shift.

Conditional inference trees have been used periodically in research on language variation and change (e.g., Gordon et al. 2004 and Tagliamonte & Baayen 2012). Software packages build trees by splitting data for a dependent variable on the basis of a predictor variable with the lowest *p*-value, and then recursively re-splitting each branch until all significant predictors are exhausted.

Besides identifying significant splits in a dataset, conditional inference trees can also be used for machine learning and predictive modelling. Packages create a "forest" by generating multiple trees from random samples of a dataset. They return weighted averages for predictors across all the trees, and these weighted averages provide a model of the data that can be used to make predictions about new datapoints.

I show that this predictive modelling provides a new technique for examining structural relationships among vowels. Models can estimate what the mean F1 and F2 of Vowel A will be on the basis of measurements of Vowel B in a training dataset. Then, in a test dataset, measurements for Vowel B can be used to estimate what the measurements of Vowel A should be. Those estimates can be compared to observed measurements to check the validity of the model.

As a case study of this technique, I examine data from an American English variety which has been identified as undergoing a chain shift referred to as the "Canadian" or "California Vowel Shift" (see Becker *et al*. fc.). I built a training set of vowel measurements from 63 speakers and a test set from 21 speakers, and used the {partykit} package in R to build trees and predictive models (Hothorn, Hornik & Zeileis 2006; Hothorn & Zeileis 2015).

The models confirm that the vowels are structurally interrelated, consistent with the concept of a chain shift. Furthermore, these structural interrelationships tend to exist between adjacent vowels rather than across the vowel system, suggesting a sequencing of adjustments in the vowel system rather than the entire system moving in lockstep. The models also show that structural interrelationships cannot fully account for observed changes, and that macro-level social factors should be included in models. This shows that the chain shift is indeed a consequence of structural factors, but also a consequence of social factors.

At an immediate level, then, this project sheds new light on a the structural (and social) causes of sound changes in a particular chain shift. More generally, though, it demonstrates predictive modelling as a powerful approach to test claims about relationships among vowels in a chain shift, and provides a useful tool for sociolinguists working from increasingly massive corpora of vowel measurements.

References

Becker, Kara (ed.). Forthcoming. *The North American Shift: Uniting the Canadian Vowel Shift, the California Vowel Shift, and short front vowel rotations across North America* [Publications of the American Dialect Society 104]. Durham, NC: Duke University Press.

Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret Maclagan, Andrea Sudbury & Peter Trudgill. 2004. *New Zealand English: Its origins and evolution*. Cambridge: Cambridge University Press.

Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15.3. 651-674.

Hothorn, Torsten & Achim Zeileis. 2015. partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research* 16. 3905-3909. URL: http://jmlr.org/papers/v16/hothorn15a.html

Labov, William 1994. *Principles of linguistic change*, vol. 1, *Internal factors*. Oxford: Blackwell.

Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests and trees of York English: *was/were* variation as a case study for statistical practice.