Modeling the Speech Community Through Multiple Variables: Trees, Networks and Clades

James A. Walker  
York University

Michael Dunn  
Uppsala University

Aymeric Daval-Markussen  
Aarhus University

Miriam Meyerhoff  
Victoria University of Wellington

Introduction

- Variationist sociolinguistics has been criticized for its focus on individual variables in isolation rather than taking into consideration the multiple variables used by members of the speech community to construct a 'lect'.
- Recent work has addressed the question of 'co-variation' by examining the interaction between multiple variables using statistical tests of correlation (e.g. Guy 2014; Tagliamonte & Waters 2011), though the results of such studies have proven inconsistent.
- Moreover, use of overall rates of occurrence in measuring correlation does not take into consideration effects of language-internal conditioning.
- Thus, correlation may not be able to tap into the linguistic system underlying the variation.

- We address the question of co-variation by taking a different statistical approach through the examination of an interrelated set of grammatical variables in a potentially polylectal speech community.

St. Vincent and the Grenadines

Sociolinguistic interviews with 18 speakers from 3 villages.

10,422 verb-phrase tokens.

Each token represents an utterance TYPE, a particular configuration of 'formal' and 'functional' characteristics:

- Formal:
  - Tense/aspect marker
  - Auxiliary/modal
  - Form of copula
  - Form of lexical verb
  - Form of negation
  - Following grammatical category
- Functional:
  - Lexical verb
  - Temporal reference
  - Aspect
  - Grammatical person
  - Subject type
  - Sentence type
  - Clause type

Measuring distance between utterances

\[ d = \frac{1}{2} \sum_{i} d_{ij} \]

Where \( d_{ij} \) is the distance between utterances \( i \) and \( j \).

We could draw a tree using a neighbour-joining algorithm ...

But not everything fits neatly into a tree ...

Conclusion

- We have used neighbour-joining algorithms implemented in phylogenetic software to model the relationship between speakers in three ways: trees, cladograms and neighbour networks.
- The most consistent grouping of speakers according to the distribution of utterance types in their sociolinguistic interviews is according to the village they come from on Bequia.
- These results are consistent across models.
- These results are also consistent with those of previous studies of individual variables (see Meyerhoff & Walker 2013)

REFERENCES