



Methodological Issues in a Real-Time Study of Glaswegian Vowels: Automation and Comparability

Brian José (bjose.ling.phd@gmail.com) and Jane Stuart-Smith (jane.stuart-smith@glasgow.ac.uk)



88th Annual Meeting of the Linguistic Society of America – Minneapolis, Minnesota – 03 January 2014



Figure 1: Map of Scotland

BACKGROUND AND INTRODUCTION:

- Goals:
 - To contribute to...
 - ...recent (and not so recent) studies of language variation in Glasgow, Scotland e.g., Lawson, E. & Stuart-Smith 1999; Lawson, R. 2011; Macafee 1983, 1988; Macaulay 1977; Stuart-Smith 1999; Stuart-Smith, Timmins, & Tweedie 2006, 2007; Stuart-Smith, Pryce, Timmins, & Gunter 2013; among others
 - ...an ever-growing body of real-time sociolinguistic research e.g., Blake & Josey 2003; Blondeau 2001; Gregersen 2009; José 2010; Labov, Rosenfelder, & Fruehwald 2013; Rickford & Price 2013; Sankoff 2005, 2006; Sankoff & Blondeau 2007; Trudgill 1988; van Hofwegen & Wolfram 2010; Wagner 2012; among others
 - ...discussions over the increasing use of automated procedures in sociolinguistics e.g., Labov et al 2013
 - ...ongoing questions about data comparability in sociolinguistics e.g., Bailey & Tillery 2004; Chambers & Trudgill 1980; Gregersen, Beck Nielsen, & Thøgersen 2009; Labov 1981; Sankoff 2005; Tillery & Bailey 2003; among others

THE CURRENT FOCUS:

- The six bimoraic monophthongs of Scottish English (Scobbie et al 1999)
 - FLEECE /i/ FACE /e/ BOOT /u/
 - CAT /a/ COT /ɔ/ COAT /o/
 - A cross-sectional (i.e., trend) study
- | Table 1: Thirty-one speaker sub-sample | | Age and Sex of Speakers | | | |
|--|-------|-------------------------|---|-------------------|---|
| | | "Old" 67-90 yrs | | "Young" 10-17 yrs | |
| Decade of Recordings | 1970s | M | F | M | F |
| 1970s | 4 | 3 | 4 | 4 | |
| 2000s | 4 | 4 | 4 | 4 | |
- Recording types
 - Sociolinguistic interviews (70-O, 70-Y)
 - Oral history interviews (70-O, 00-O)
 - Peer pair conversations (00-Y)

METHODS:

- LaBB-CAT database (Fromont & Hay 2012)
 - Storage of time-aligned orthographic transcriptions
 - Forced-alignment of phonemic transcriptions and audio recordings
 - A searchable database: automatic vowel extraction
 - Automatic formant measurements via integration with Praat
 - 3 measurement points stipulated: 25% ~ 50% ~ 75%
- Quality control
 - Likely formant tracking errors
 - Removal of clear and obvious mathematical outliers
 - Likely sub-optimal forced alignments (and/or formant tracking errors)
 - Removal of arguably non-monophthongal formant measures
- Statistical modeling (main effects only)
 - Linear mixed effects modeling in R
 - Fixed factors: {Group (70-O, 00-O, etc) or Vowel}*, Sex, following POA; (DCA category)
 - Random factors: Speaker, Word

- Comparability coding: One approach among other possibilities
 - LANCHART's Discourse Context Analysis (DCA) framework, modified

LANCHART version: 6 categories	Glasgow version: 4 categories
• S = Type of speech event	⇒ • P = Participant structure + recording type
• A = Activity type	⇒ • A = Activity type
• I = Type of interaction structure	⇒ —x—
• M = Type of macro speech act	⇒ —x—
• G = Type of speech genre	⇒ • G = Type of speech genre
• U = Enunciation (Danish <i>udsigelse</i>)	⇒ • C = Citations

Table 2: Adaptation of DCA Coding Categories and Values Retained after Recoding

DCA-inspired category	values in this speaker sample ultimately recoded as...
P: Participant Structure	⇒ —x—
A: Activity Type	⇒ (i) Conversation with Participant; (ii) Other
G: Speech Genre	⇒ (i) Narrative; (ii) Specific Acct; (iii) General Acct; (iv) Other
C: Citations	⇒ (i) yes; (ii) no/other

RESULTS:

- I. Automation: Glaswegian Vowels in Real Time

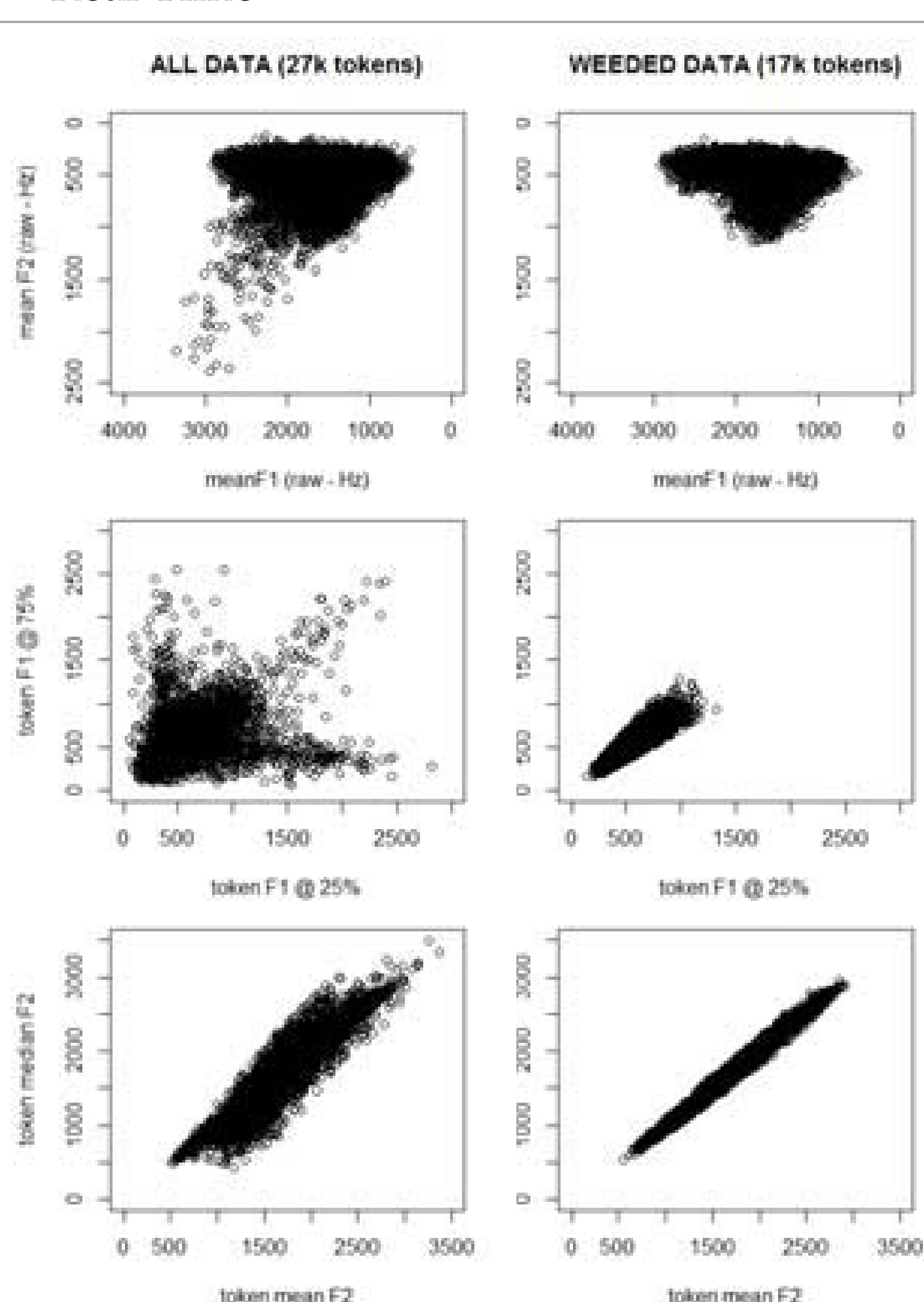


Figure 2: The effects of suspicious token removal

- Selected References:
 - Gregersen, Frans; Beck Nielsen, Søren; Thøgersen, Jacob (2009). Stepping into the same river twice: On the discourse context analysis in the LANCHART project. *Acta Linguistica Hafniensia* 41: 30-63.
 - Labov, William; Rosenfelder, Ingrid; Fruehwald, Josef (2013). One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language* 89: 30-65.
- Acknowledgements:
 - Robert Fromont for expert technical support with LaBB-CAT
 - Frans Gregersen and his LANCHART team for DCA guidance
 - Ludger Evers & Tereza Neocleous for helpful statistical advice
 - Ellen Gallagher for invaluable assistance with data handling

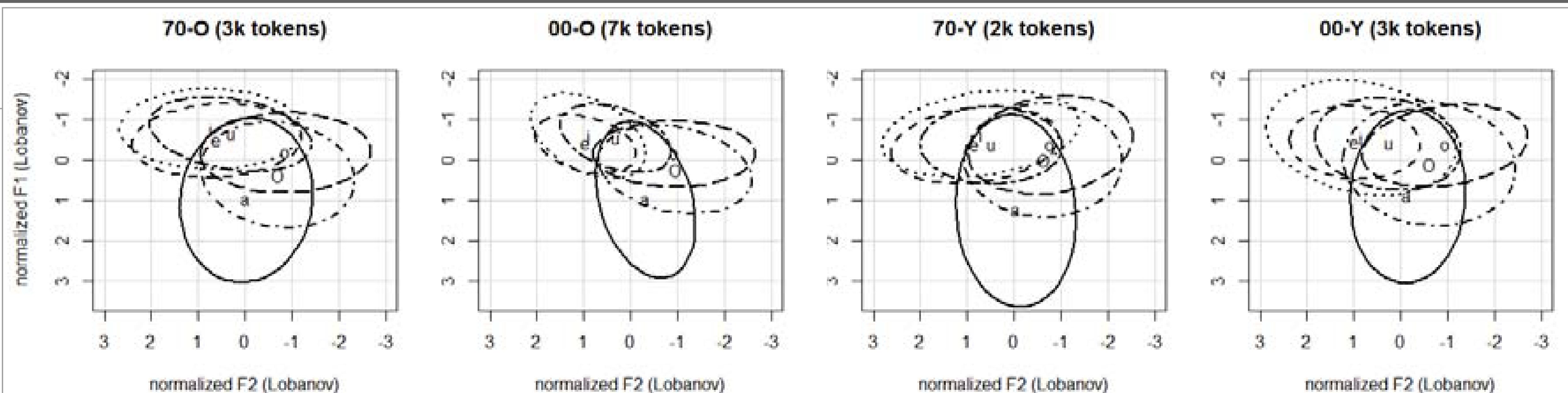


Figure 3: Structural Relationships between Vowels at 4 Points in 'Time': 70-O, 00-O, 70-Y, 00-Y
FLEECE /i/ & FACE /e/; COT /ɔ/ (=O) & COAT /o/ not statistically distinguished on F2 in any 'time' period
F1 distinctions among FACE /e/, BOOT /u/, & COAT /o/ lost in 00-Y

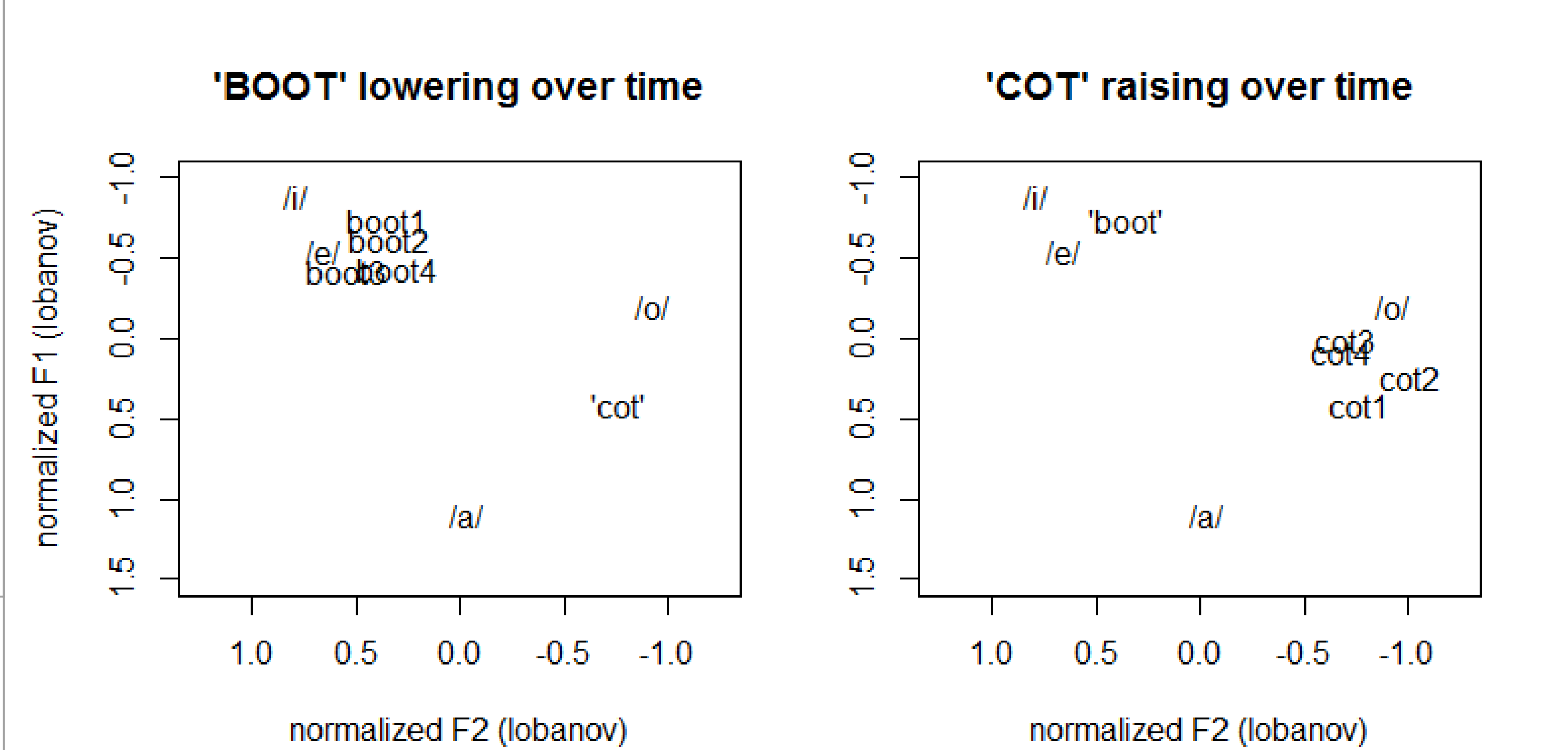


Figure 4: The Repositioning of Two (Individual) Vowels over 'Time'
BOOT /u/ exhibits lowering between the older and younger speakers
COT /ɔ/ (=O) exhibits (i) raising between 70-O & 70-Y and (ii) fronting between 00-O & 70-Y (after some backing between 70-O & 00-O)

- II. Comparability: the DCA framework
 - DCA-P: Interviews are only P-LILS, P-LIMS; Conversations only P-0I2S, P-0IMS
 - DCA-A: Most tokens (>95%) = A-ConP, A-Mono (Not clearly separated cat.s, here) A-Mono frequent in 00-O (~60%) Much less A-Mono in -Y (~4% ~11%)
 - DCA-G: G-GnAc, G-Narr, & G-SpAc are the only consistently robust cat.s ('though less so for -Y than for -O speakers) G-SpBx also relatively frequent in 00-O (~3%) & 70-Y (~10%), G-Refl in 70s (~3%), G-Gssp in 00-Y (~5%)
 - DCA-C: 1638 tokens (of ~17k) Prejudicially removed from data
 - Next Steps: Examine DCA influence on F1, F2; ...