

Constraints, community, coherence:

Do sociolects exist?

Gregory R. Guy

New York University

# Coherence vs. Diversity

- Speech communities appear to be coherent: speakers who share a language communicate efficiently
- Communities, and individuals, are also highly diverse in linguistic experience and practice

# Speakers and communities

- Coherent lects: languages, dialects, sociolects, ethnolects, etc.
  - each distinctive variety is identified by a cluster of linguistic features
- Identity, performance, agency
  - each speaker constructs identity and performs style by purposeful choice of linguistic features

# The speech community model

Speech communities are defined by:

- high internal density of communication
- shared linguistic features
- shared norms for language use

Shared characteristics co-occur in usage, make the community coherent.

# The individual agency model

By their purposeful choices to use particular linguistic forms, speakers:

- Construct and perform social identities
- Create social meaning
- Do styling, stance-taking

Chosen forms may differ between speakers, or discourses, permitting incoherence

# Speakers' choices constitute *bricolage*

- Speakers draw from “a range of existing resources to construct new meanings or new twists on old meanings”

(Eckert 2004)

# Historical roots

- Analogous issues arise in dialectology and diachronic linguistics
- Dialects: Do isoglosses bundle?
- Diachrony:
  - “sound laws” vs. each word/feature has its own history
  - Family trees vs. ‘wave’ models of change; areal phenomena

# The coherent view: reification of language varieties

- Linguistic varieties ('lects') are commonly treated (in popular usage and by linguists) as if they are identifiable and coherent entities
  - languages
  - dialects
  - ethnolects
  - class/status-based varieties
  - styles/registers

# Each variety is typically associated with multiple variables

- NYC English (cf. Labov 1966)
  - Coda /r/ deletion; raised /æh,oh/; th-stopping
- African American English
  - Invariant ‘be’, remote past ‘been’, etc.
- Popular Brazilian Portuguese
  - Non-agreement in NP and VP
  - Coda /s/ deletion, vowel denasalization

# Coherence and covariation

- For each identifiable lect, the set of associated variables co-occur, to collectively define the variety
- The variables are the individual bricks that together build the structure of the lect – the coherent ‘unified whole’

## Against coherence:

### Identity construction and *bricolage*

- Each linguistic feature may have distinct and unique social indexicalities
- Speakers assemble feature clusters for individual purposes, constructing personal identities and styles
- Clusters of features are ephemeral, and social groups of speakers are not necessarily linguistically coherent

# Speech communities and accommodation

- Speech communities (SCs) are networks of communicative networks
- SCs have relatively high internal density of communication and shared norms
- Speakers accommodate to interlocutors
- Therefore, networks of speakers should be linguistically similar/coherent

# Speech communities

- The speech community model accounts well for groups of speakers that talk more to each other than to outsiders
- Hence, communities defined by
  - Geography (dialects)
  - Ethnicity (ethnolects)
  - Social class

# The limits of the SC model

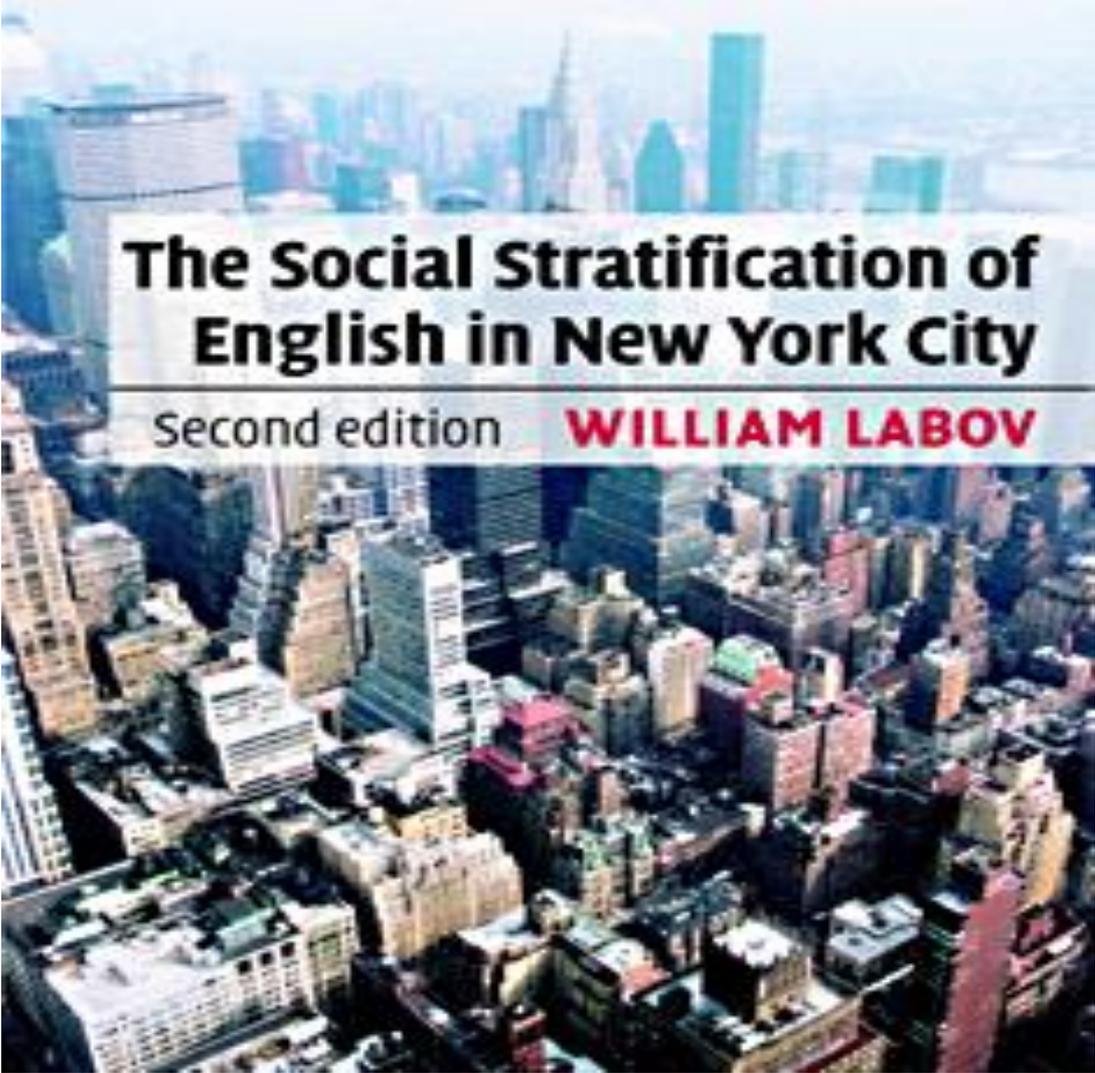
- The speech community model is less adequate for modeling language varieties associated with speakers who are not linguistically isolated from others
- Thus, varieties associated with
  - Gender
  - Sexual orientation
  - Other social clusters: nerds, hip-hoppers, adolescents, communities of practice

# More limits

- The speech community model also does not provide a simple account of:
  - Stylistic variation (different usages in the same community and the same individual; do these cohere?)
  - Linguistic change (produces incoherence at the community level)

# Speech style

- Are speech styles coherent?
- Does use of 'Casual style' imply simultaneous use of all 'casual' variants?

An aerial photograph of New York City, showing a dense urban landscape with numerous skyscrapers and buildings. The image is slightly hazy, giving it a soft, atmospheric quality. The text is overlaid on the upper portion of the image.

# **The Social Stratification of English in New York City**

Second edition

**WILLIAM LABOV**

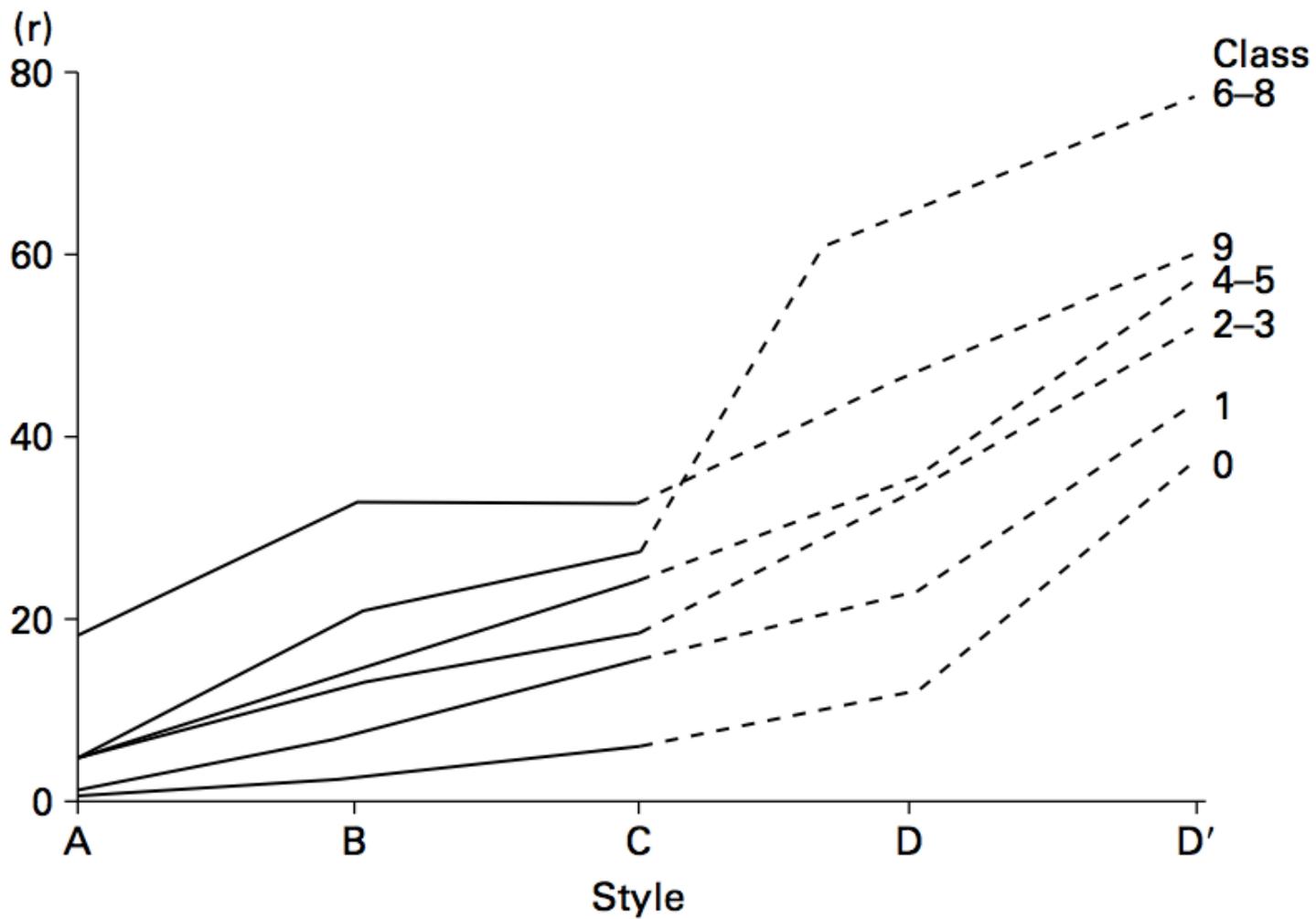
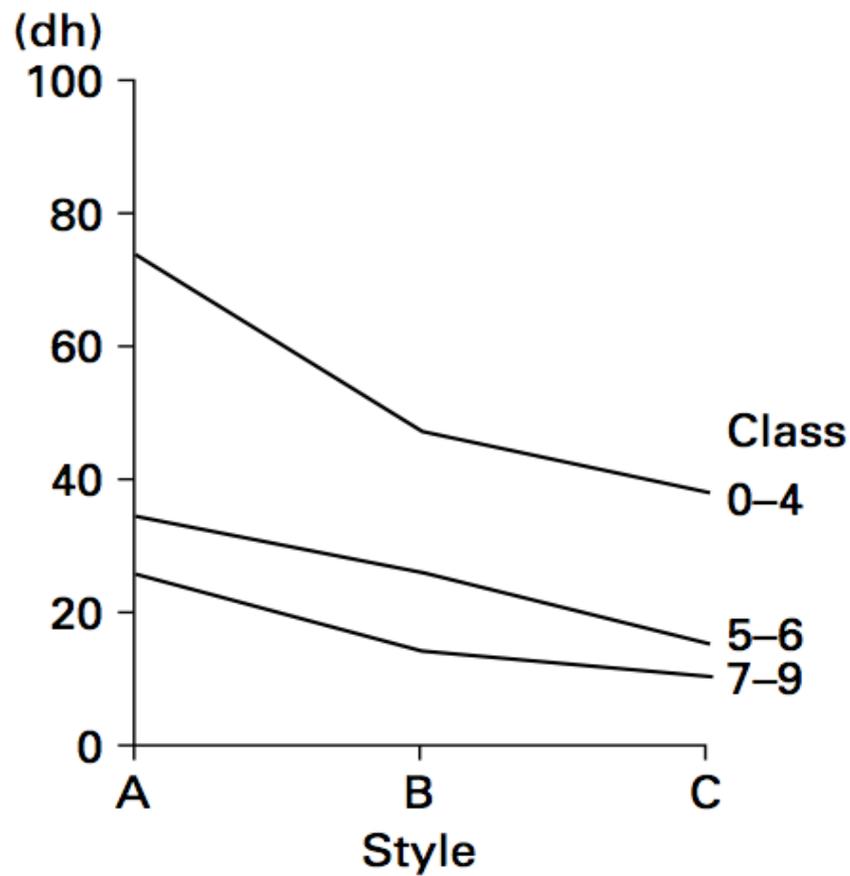
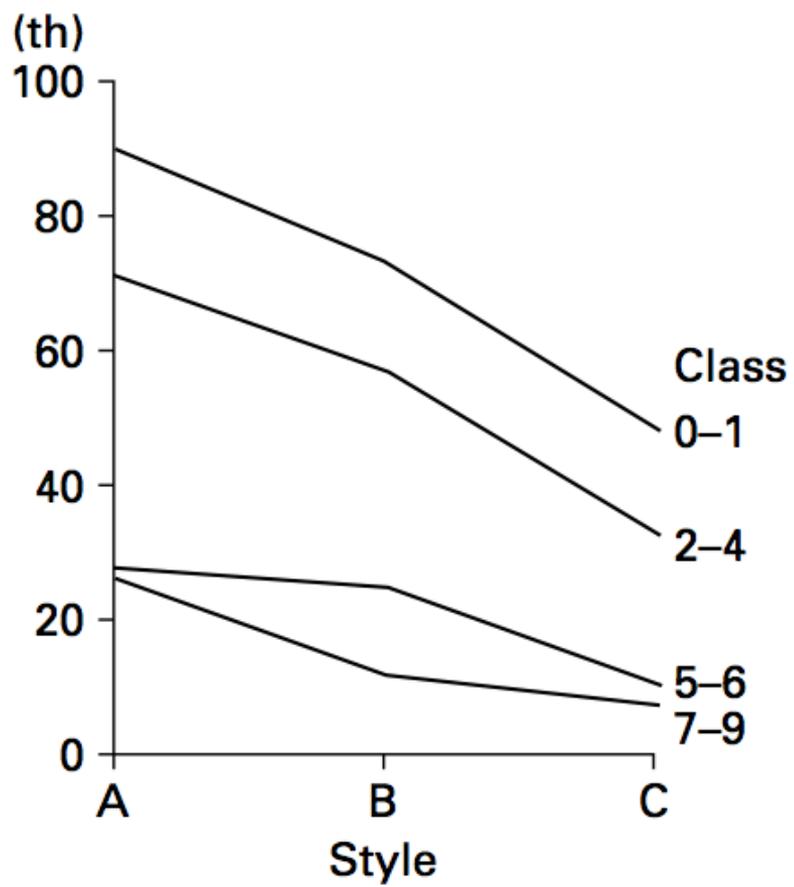


Figure 7.11 Re-defined class stratification of (r): six class groups



# The incoherence of linguistic change

- If speech communities are coherent, why do they ever show language change?
- Contact with outsiders could trigger change ‘from above’: introducing new interlocutors, patterns of accommodation and convergence
- But ‘change from below’, -- innovation led by younger speakers -- is disruptive to community coherence, and constitutes anti-accommodation to established community patterns

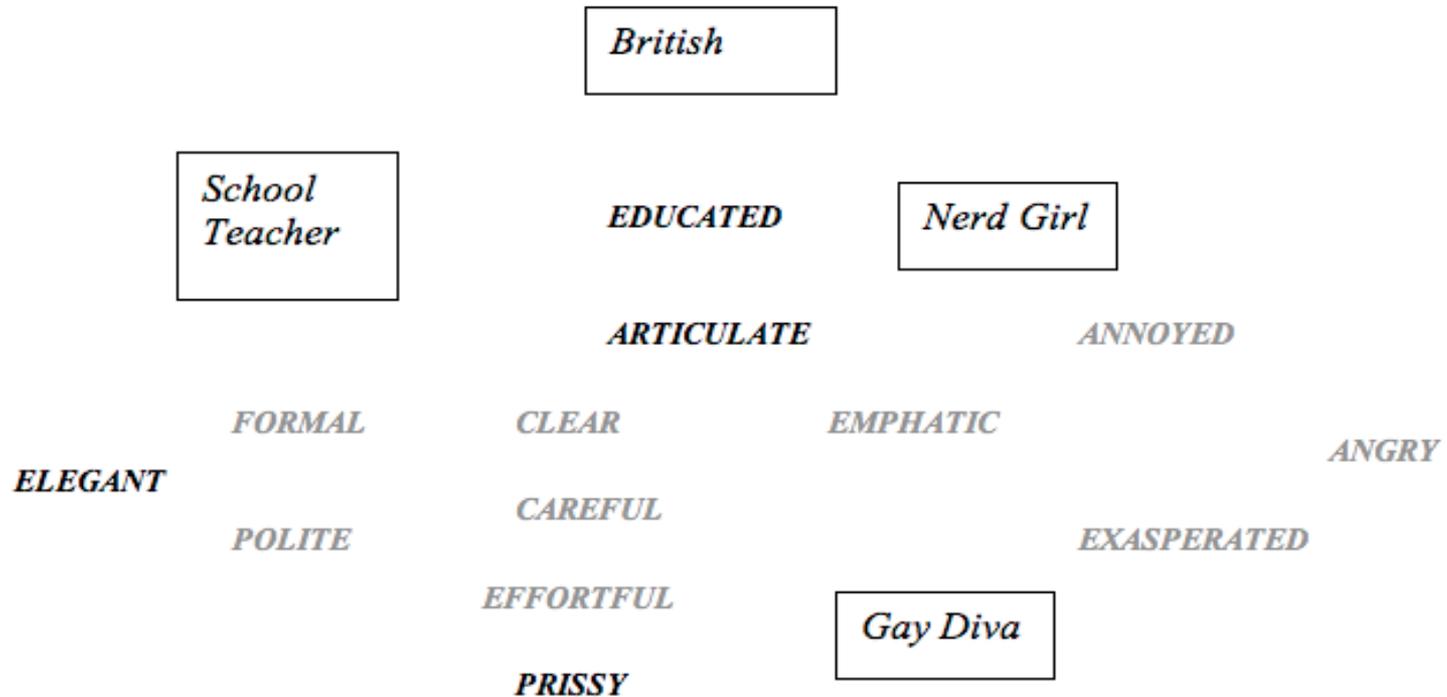
# Hence, identity construction

- The anti-coherence model(s) thus focus on aspects of linguistic usage associated with innovation, stylistic practice, stance-taking, identity formation
- Emphasize individual agency and the unique indexicality of each variable.

# Variables have complex and idiosyncratic indexicalities

- May separately or simultaneously index characteristics associated with locality, class, ethnicity, gender, age, innovation, style, stance, etc.
- Variables do not necessarily cluster on any of these dimensions
- Speaker agency means they can select features for personal, even ephemeral purposes

# Indexical field for /t/ release in American English (Eckert 2008)



# An empirical approach:

Do variables cluster, correlate, co-occur?

Dialects:

- Do most speakers from a place use most or all of the features associated with the local dialect?

Ethnolects:

- Do most speakers of a given ethnicity use most/all of the features associated with that ethnolect?

Social class:

- Are the socially stratified variables in a speech community correlated?
- Does use of one prestige variant imply use of other prestige variants?

# Correlations: the logical possibilities

- Multiple sociolinguistic variables could correlate tightly, loosely, or not at all

Complete absence of correlation, 9 lects

Perfect correlation; 3 lects

Values of variable B

High    Mid    Low

Values of	High	hh	hm	hl	hh	
variable A	Mid	mh	mm	ml		mm
	Low	lh	lm	ll		ll

# Caution: structural vs. social correlation

- Some variables may be correlated for reasons of linguistic structure
  - e.g. vocalic chain shifts; parametrically linked syntactic variables
- Structural correlations of variants do not prove social coherence

# Subject pronoun expression in Spanish: Dialect differences in the effect of specificity of reference with 2<sup>nd</sup> sg. *tú*

	San Juan, PR			Madrid, Spain		
	% overt pro.	factor wt.	N	% overt pro.	factor wt.	N
[+specific]	48%	.51	145	40%	.72	58
[-specific]	69%	.72	188	19%	.50	150

- Source: Cameron 1993, p325

# Contemporary practice

- Coherence is often assumed in SC studies
- Strict correlations are sometimes claimed (e.g. creoles: basilectal vs acrolectal variants)
- Non-correlation is assumed in studies of identity construction, *bricolage*, etc.
- But the issue is not often empirically tested (exceptions: e.g., Horvath & Sankoff on Australian English)
- Much sociolinguistic analysis looks at one variable at a time

# Empirical testing of coherence: Horvath & Sankoff 1987

- A classic study looking at multiple variables, inferring the social groupings from the clustering of variants, rather than defining the social groups *a priori*, by social criteria.

BARBARA HORVATH AND DAVID SANKOFF

*THE SYDNEY SPEECH COMMUNITY*

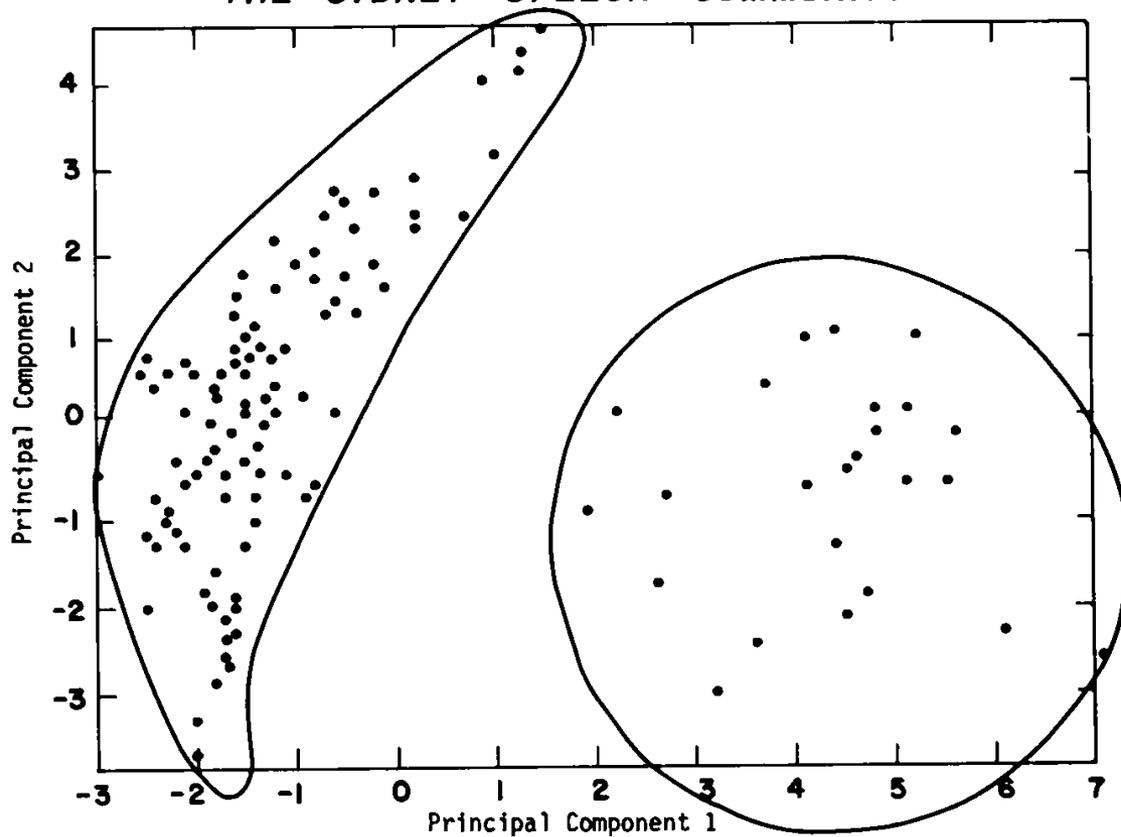
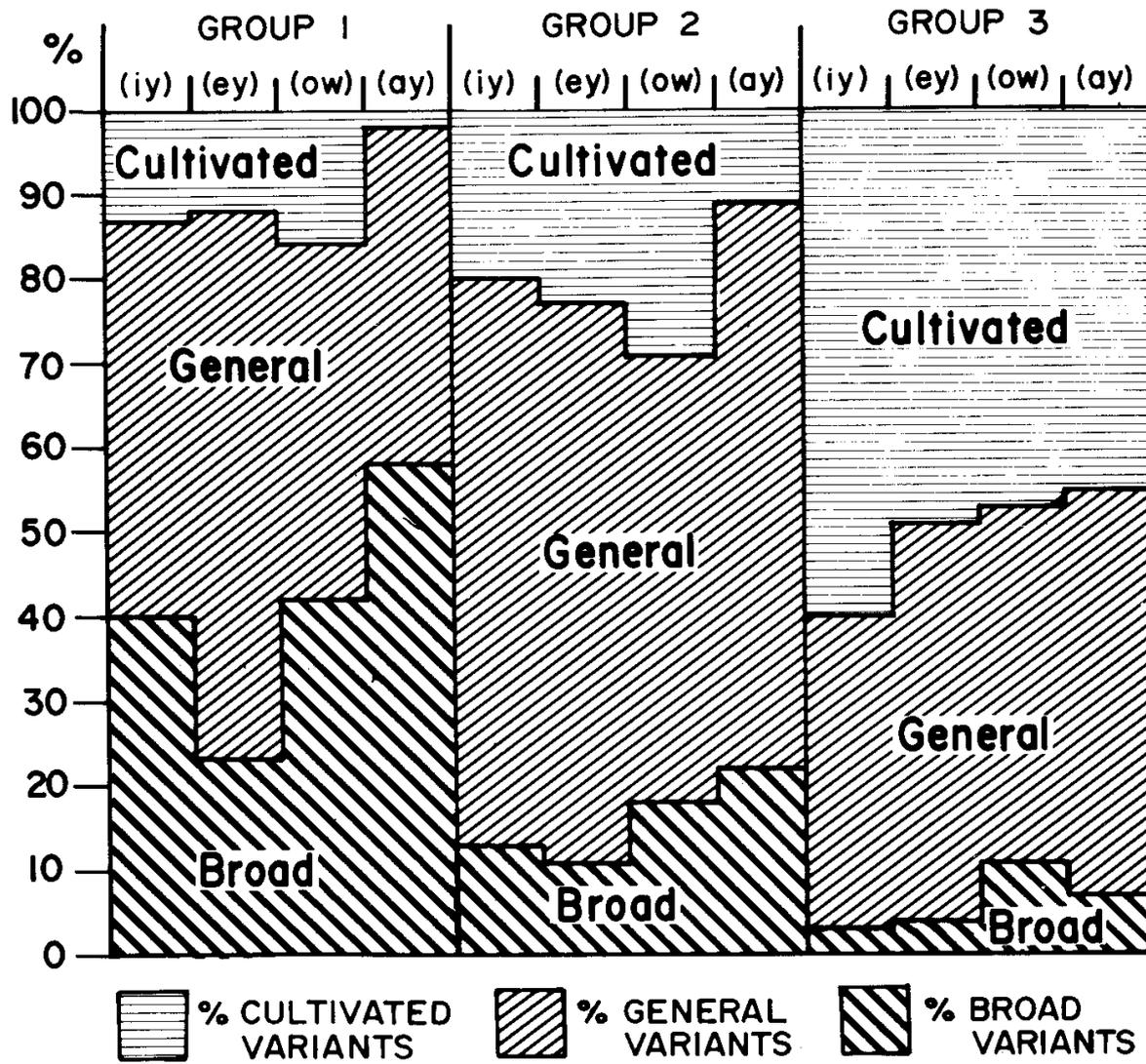


FIGURE 1: The Sydney speech community.

# VOWEL VARIATION WITHIN CRESCENT GROUPS



# New data

- Four studies examining speech communities in which multiple variables are present, some phonological, some syntactic
- All investigating whether speakers tend to use multiple variables in similar ways
- Distinct sociolinguistic processes:
  - social stratification
  - dialect contact and convergence
  - language contact and assimilation
  - change in progress

# Studies of covariation

- Brazilian Portuguese: socially stratified variables (Guy 2013-RJ; Oushiro & Guy 2013-SP)
- NYC Spanish: dialect and language contact and convergence (Erker 2012)
- NYC English: Change in progress (Becker 2010)

# Studies of shared constraint effects

- Becker on NYCE
- Guy on Brazilian Portuguese
- Guy on US and NZ English
- Lim on Singapore English
- Forrest on English *-ing*

# Brazilian Portuguese: the variables

- Two syntactic variables in both studies:
  - Verbal agreement (3<sup>rd</sup> plural marking)  
Eles disse/disseram. ‘They said(sg/pl)’
  - Nominal agreement (NP number marking)  
os leão/leões ‘the(pl) lion(sg/pl)’

# Phonological variables

## Rio (Guy study)

- Denasalization of unstressed final vowels  
vagem~vage 'green bean'
- -S deletion (targets coda sibilants)  
menos~meno 'less'

## São Paulo (Oushiro study)

- R-retroflexion
- Diphthongal eN

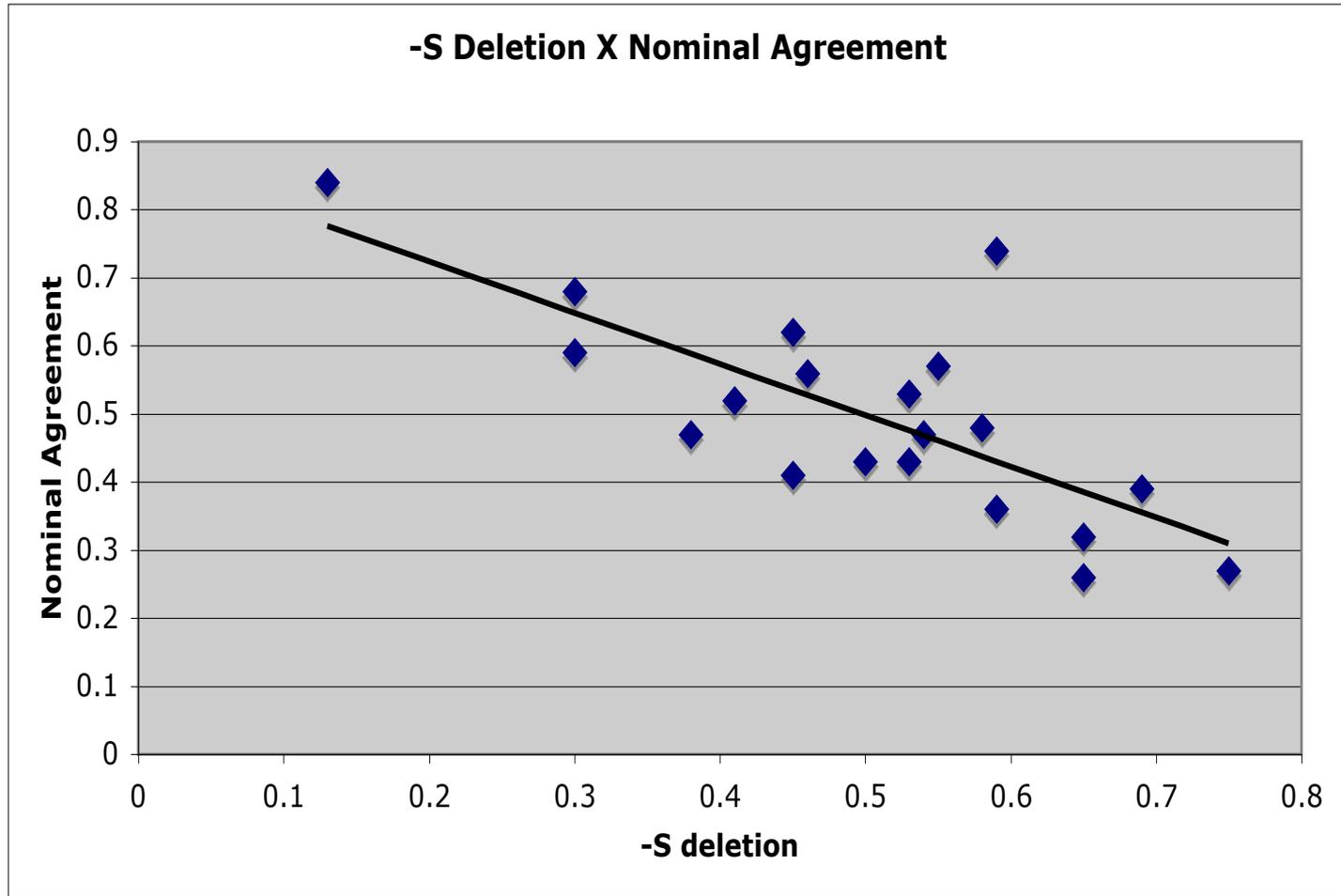
# Correlations among 4 sociolinguistic variables in PBP (RJ)

<b>Denas</b>			
<b>-0.450*</b>	<b>VerbAgr</b>		
0.256	-0.371	<b>Sdel</b>	
<b>-0.436*</b>	<b>0.592**</b>	<b>-0.740***</b>	<b>NomAgr</b>

Significance: \*p<.05, \*\*p<.01, \*\*\*p<.005

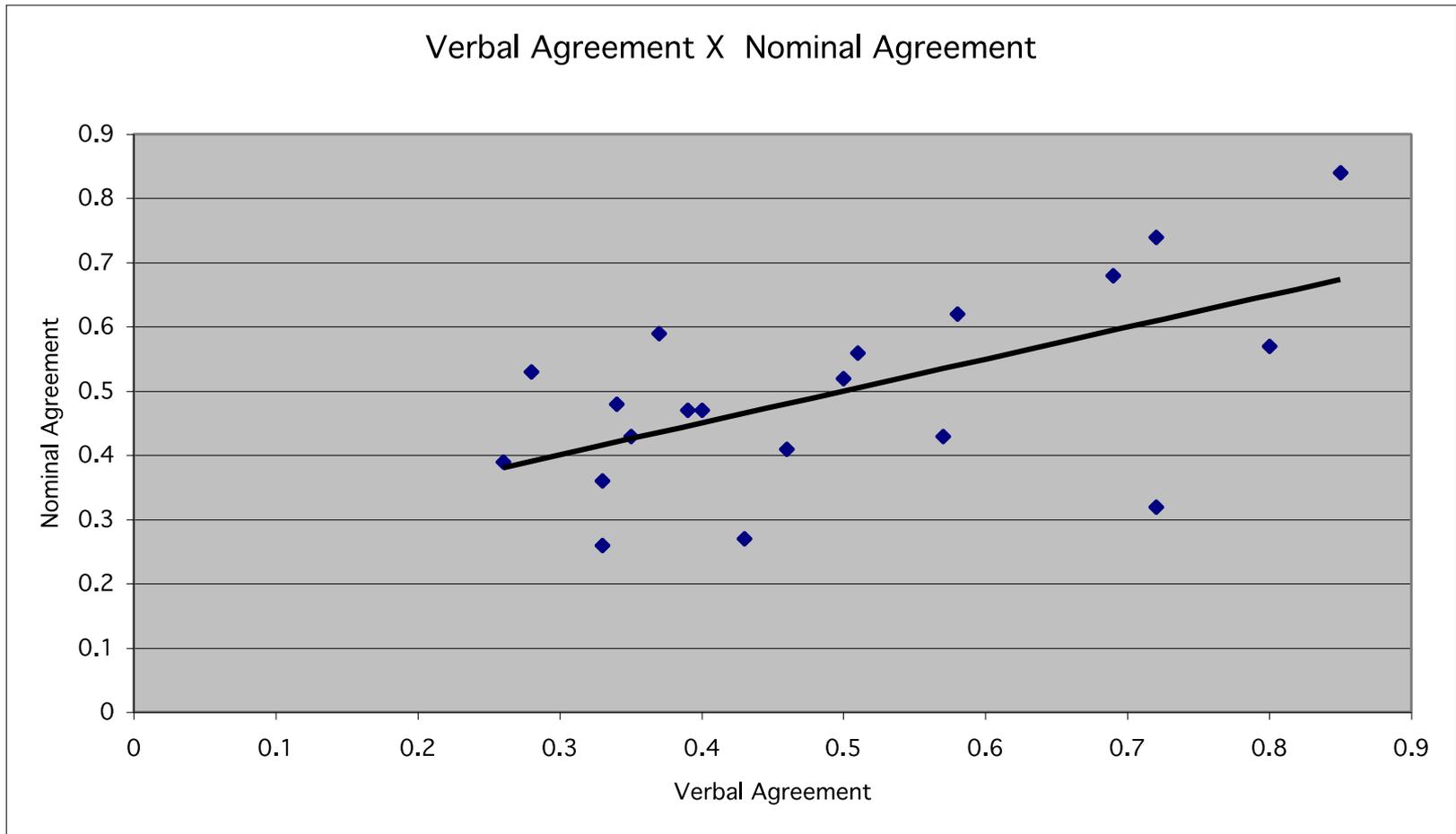


# -S deletion and Nominal Agreement



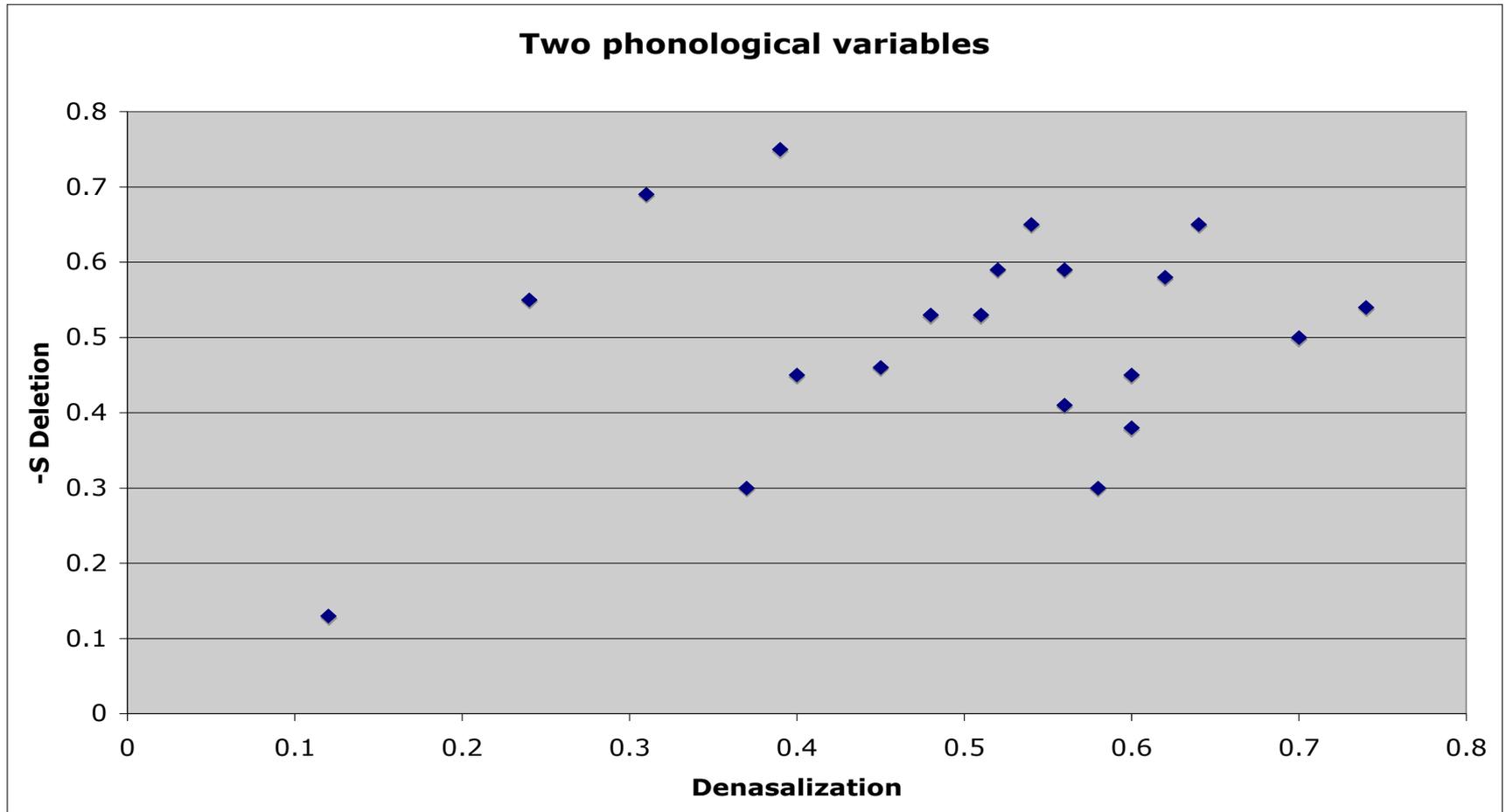
$r = -.74, p < .005$

# Verbal and Nominal Agreement



$r = .59, p < .01$

# Denasalization by -S Deletion



$r = .26$ , not significant

# Four of six variable-pairs are significantly correlated. Does this confirm coherence?

- Perhaps; certainly better than chance. But...
- Why aren't they all correlated?
- Might some of the correlations be due to structural or grammatical relationships between the variables?
- Are all these variables truly independent?

# Possible structural motivations

- Feeding relationships:
  - S deletion would increase surface absence of nominal agreement
  - Denasalization would increase surface absence of verbal agreement
- Parametric coherence: might an abstract “AGREE” parameter drive both nominal and verbal agreement?

## Another possibility: other social dimensions may affect usage

- Gender: female mean slightly more standard than male mean on all variables
- Denasalization: marked gender difference

weight	FEMALES	MALES
Above .50	2	10
Below .507	1	

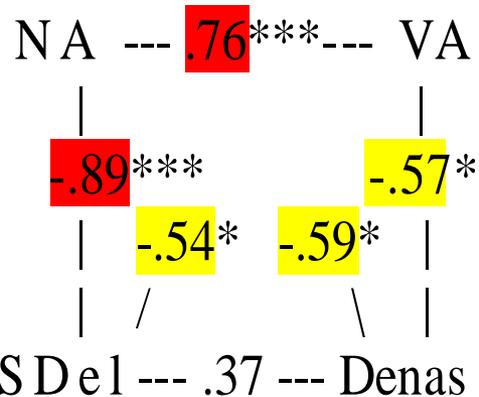
Do high rates of denasalization index male identity?  
Could such intervening variables obscure correlations?

# Within-gender correlation: female speakers

Variables

NomAgr/SDel      VerbAgr/Denas

Syntax (agreement)



Five out of six pairs show solid correlations!

# Beyond pairwise comparisons

- Divide individual results for each variable into thirds (high, mid, low rates of use of the prestige variant)
- Map the ranking group position of each speaker for all four variants
- Thus, each speaker will have a classification like hmhm, hhml, etc.

# Lectal clustering of variants

	speakers	%
All four variables same hhhh, mmmm, etc.	4	20%
Three variables same hhhx, llx, etc.	8	40%
Two same, others adjacent hhmm, mmhl, etc.	4	20%
Two same, others dispersed hhll, hhml, llmh, etc.	4	20%

# Clustering results show much better than random coherence

- 20% of speakers have all four variables agreeing; a random distribution would be 3.7%
- 40% have three variables agreeing, vs. random distribution of 11%
- Still, 20% of speakers show no meaningful clustering for these 4 variables

# Oushiro & Guy 2013

## 102 São Paulo speakers

- Same two syntactic variables (nominal and verbal number agreement)
- Two different phonological variables that are typical of SP Portuguese, and do NOT interact in any structural way with number agreement -- retroflex r and diphthongal nasal eN

# General results

- Table 1: General results for social factors

App.value	retroflex (-r)	diphthongal (eN)	(NP-0)	(VP-0)
Sex/gender	men (12)	women (24)	men (13)	men (6)
Age	younger (11)	younger (16)	stable (19)	stable (15)
Education	up to high school (16)	post-high school (8)	up to high school (33)	up to high school (26)
Area	periphery (26)	central (6)	central (7)	-

- Predictions:

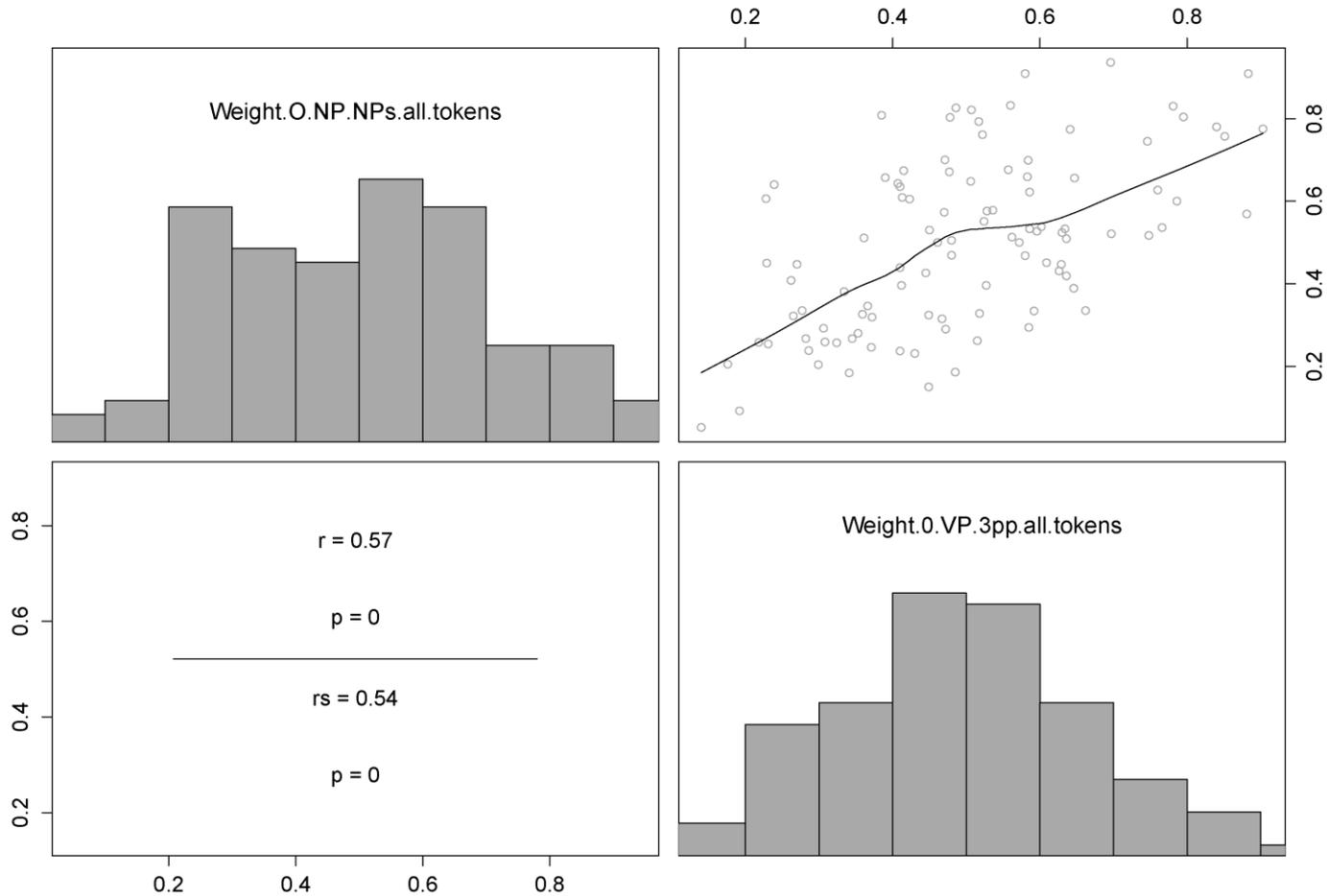
$$(NP)-(VP) > (-r)-(VP) / (-r)-(NP) > (eN)-(-r) / (eN)-(NP) / (eN)-(VP)$$

# São Paulo study: Correlations (102 speakers)

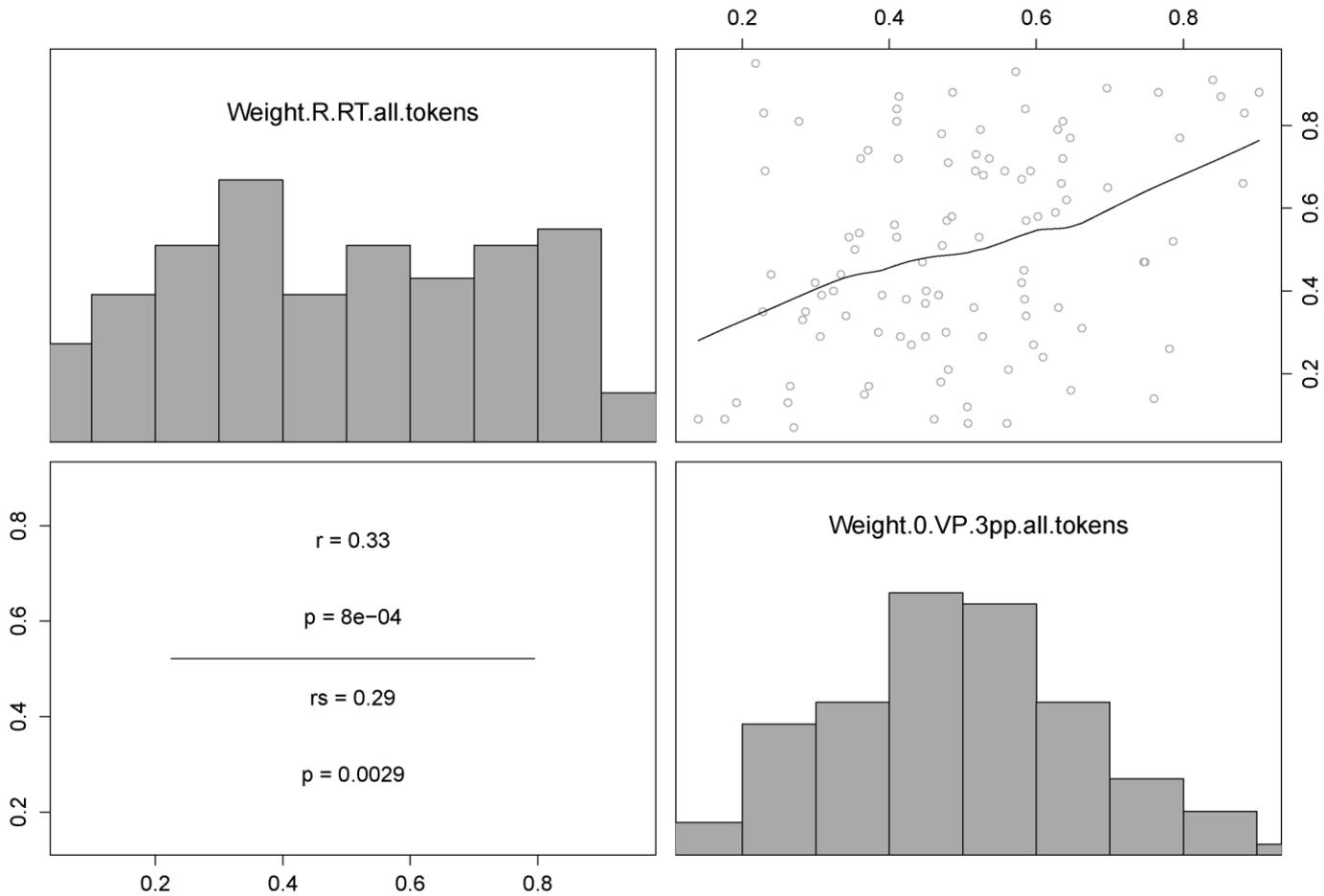
## Variables

	NomAgr		VerbAgr
Syntax (agreement)	NA	--- .57***	--- VA
	.2*		-.06
		.33**	-.17
Phonology (retroflex r, Nasal diphthong)	/		\
	(r)	--- -.14 ---	(eN)

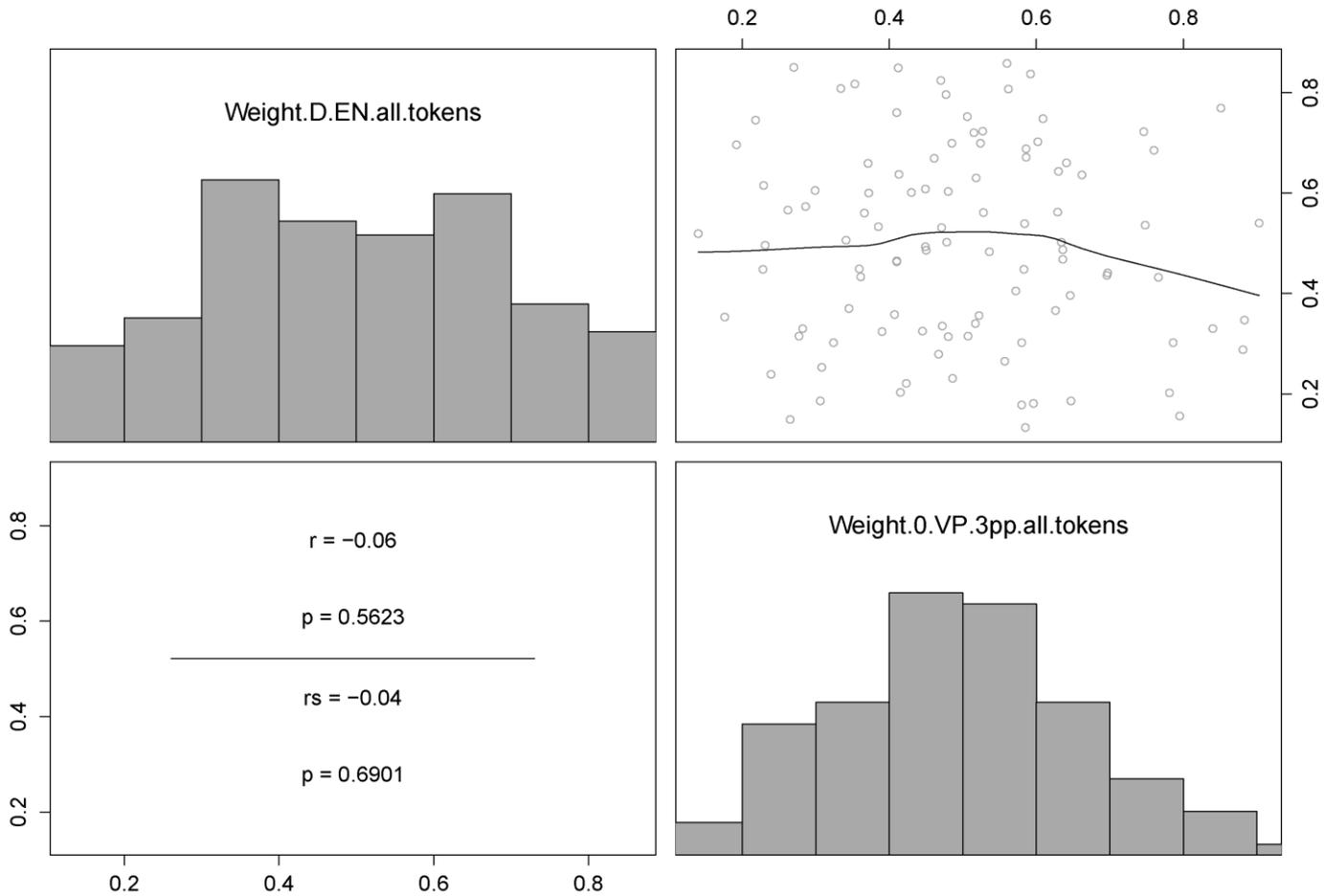
# Correlations: NP and VP



# Correlations: VP and (-r)



# No correlation: VP and (eN)



# Discussion

- strong correlation between morpho-syntactic variables (cf. Guy 2013)
- correlation between (-r) and morpho-syntactic variables (NP) and (VP), which are structurally unrelated
- (eN), undergoing change from below (Oushiro 2012), seems to be less available than the socially marked variables for composing coherent sociolects

# Spanish in NYC: Emergent dialect coherence?

- Erker 2012 looks at two measures of coda /s/ lenition, and at filled subject personal pronouns
- Contrasting treatment of these variables by speakers from Caribbean (e.g. PR, DR) and Latin American mainland (e.g., interior Mexico, Colombia)
- Compares newcomers with long-term residents of NYC

# Pronoun Rates

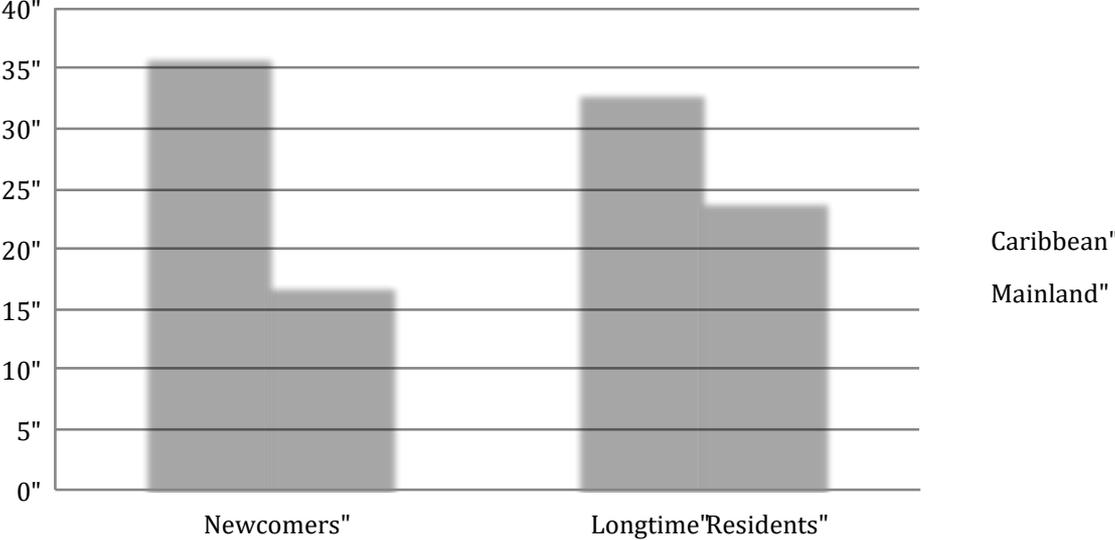


Figure 3. Deletion rates by region and exposure group

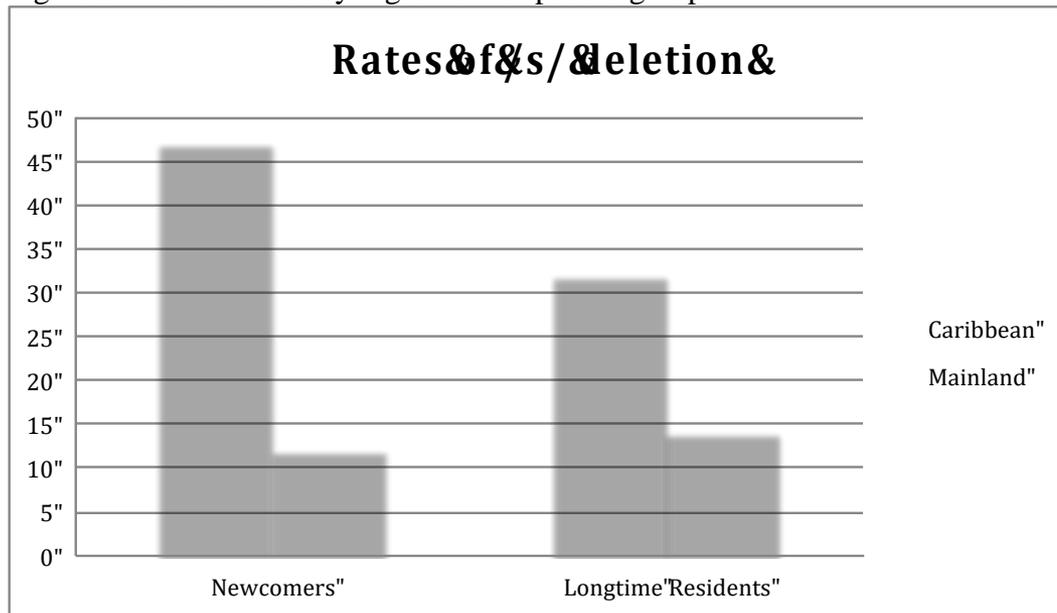
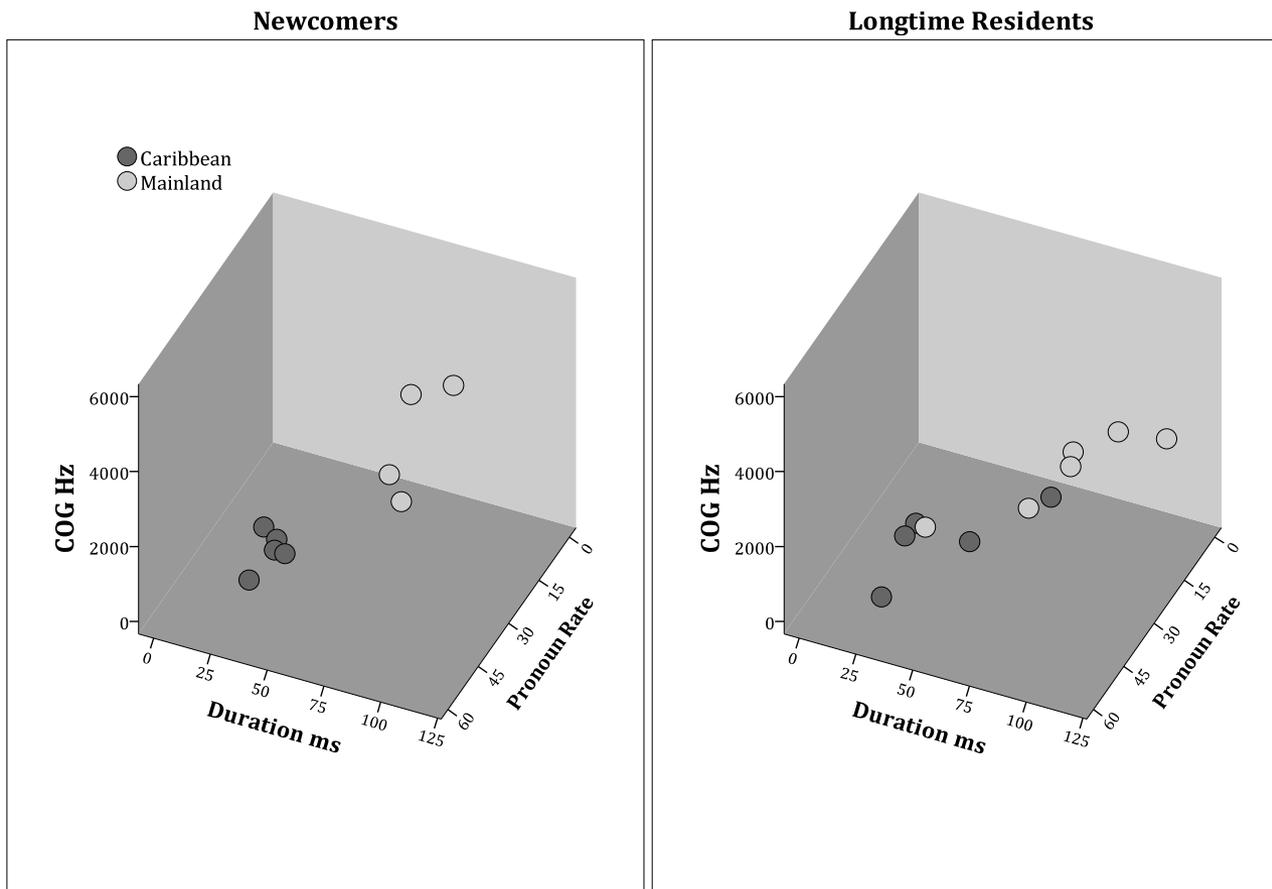


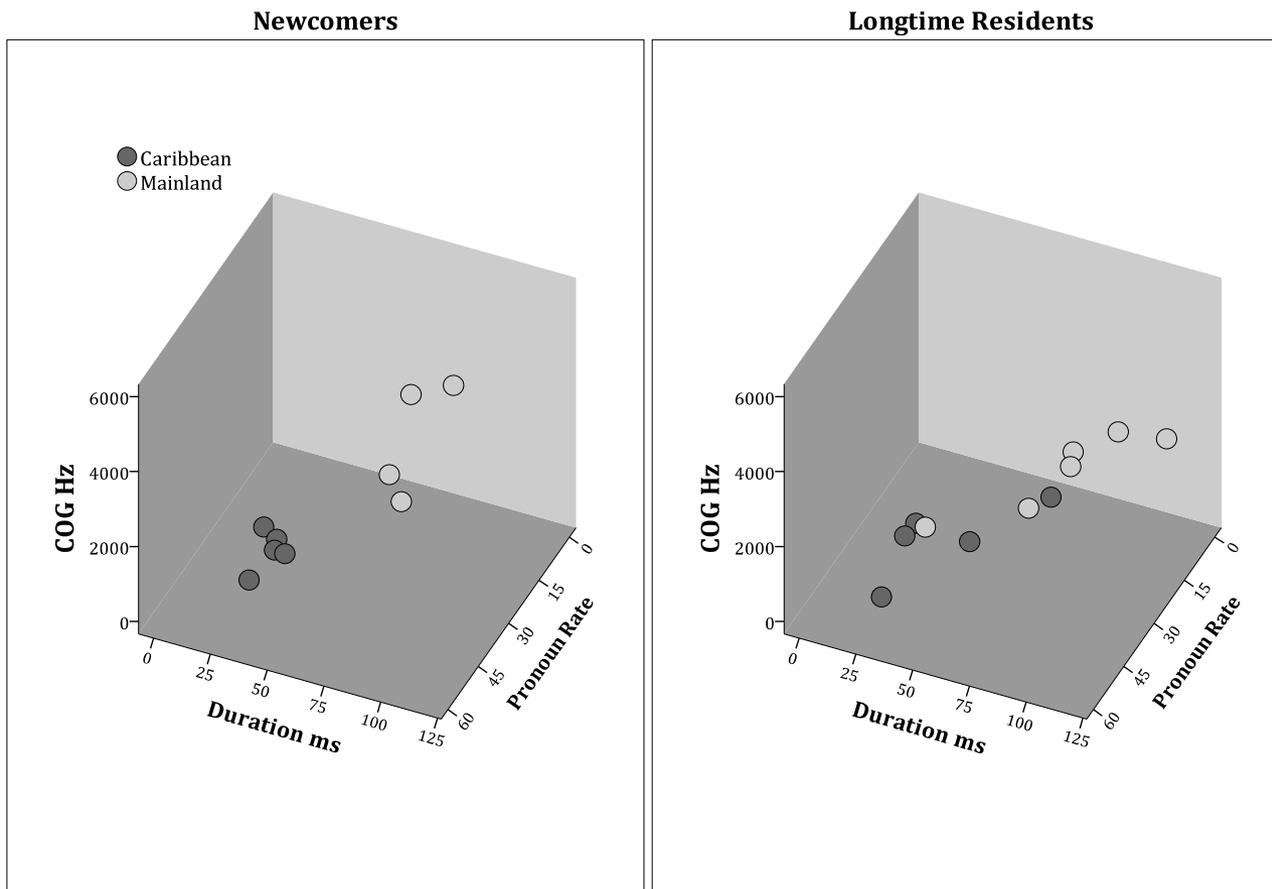
Figure 7. Mean duration and COG of /s/ by speaker, region, and exposure group.



# Spanish: dialect coherence

- Mainland dialects and Caribbean dialects are both internally quite consistent
- Mainland: All speakers have low rates of SPP and of aspiration and deletion of coda /s/
- Caribbean: All speakers have high rates of SPP, aspiration and deletion of /-s/

Figure 7. Mean duration and COG of /s/ by speaker, region, and exposure group.



# New York City English (Becker 2014)

- Three traditional features of NYCE (per Labov 1964, and others)
  - Non-rhoticity (vocalization or deletion of coda /r/)
  - Raised nucleus of BOUGHT
  - Short-a split (tense BAD vs. lax BAT)
- All of these features are receding in contemporary NYCE

# Changes in progress in NYCE

- Rhoticity: coda r productions have increased steadily since c.1940s (cf. Becker, Mather, etc.)
- BOUGHT vowel is lowering in apparent time (cf. Becker)
- Short-a split (BAD vs. BAT) involves changing contexts. More on this later...

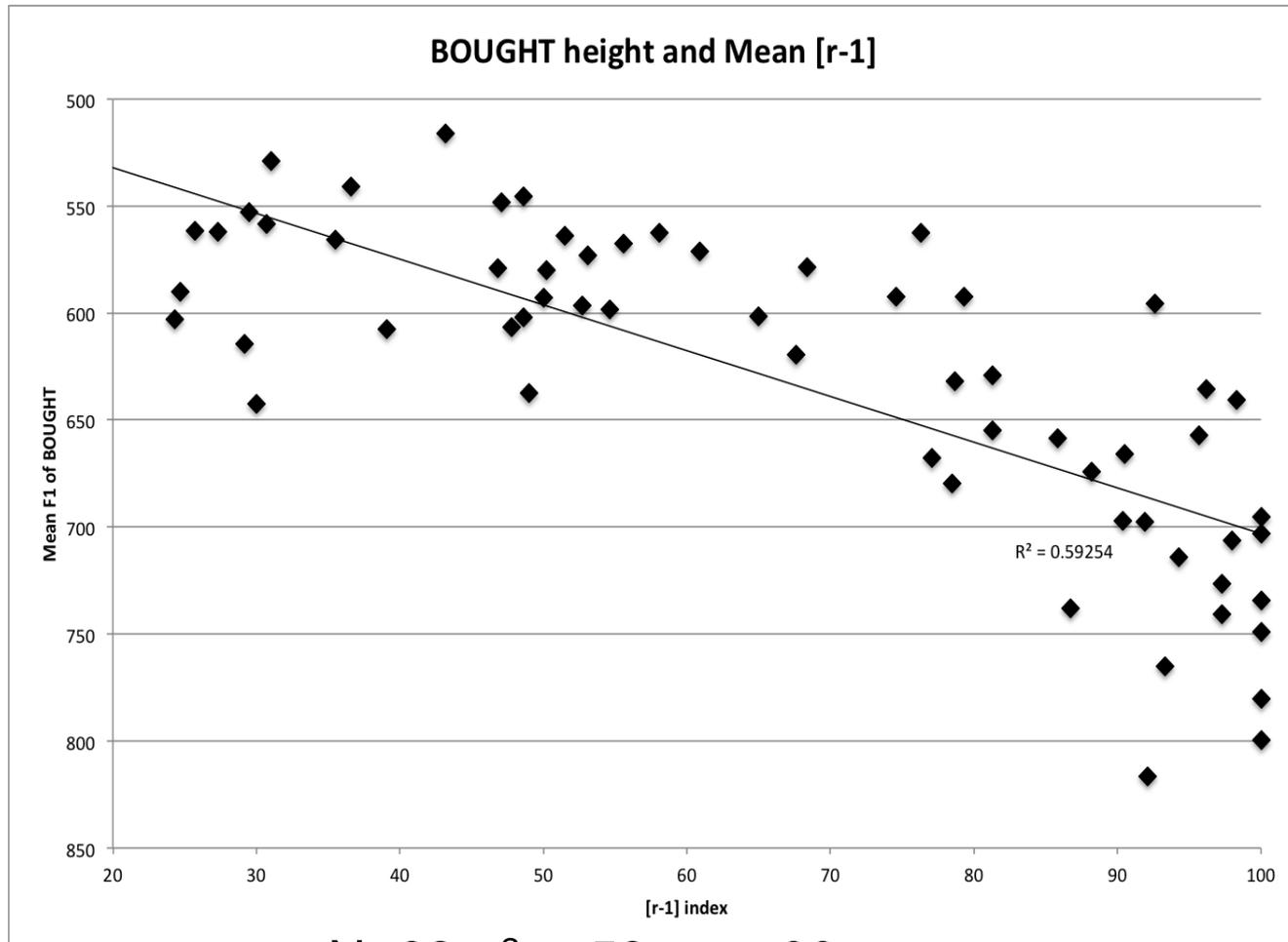
# Escape from New York

- All of these changes move New Yorkers away from traditional NYCE features, towards the phonology of the wider Midlands dialects of American English
- Hence, ‘changes from above’
- Likely motivation: the widespread stigmatization of NYCE in the American popular imagination

# The coherence question

- If all these variables index a  $\pm$ NYCE dimension, do they correlate? ...i.e.:
- If speakers lower BOUGHT, do they also use more coda /r/ ... and/or
- Do speakers who seek to construct an NYC-oriented identity simultaneously preserve non-rhoticity and raised BOUGHT?

# BOUGHT lowering and rhoticity



N=62,  $r^2 = .59$ ,  $p = .00$

# What varies?

- Speech communities show coherent patterns of effects of linguistic constraints on variables. This has been formulated in variationist theory in terms of constraint effects on variable rules.
- Differences in overall rates of use of variables is represented as an input probability ( $p_0$ ), independent of constraint effects.
- Does the indexical, agentive use of variables by speakers involve varying the input probability, or can they choose variants in ways that disregard linguistic constraints?

- **Shared Constraints Hypothesis:**  
Speech community members share common constraint effects on linguistic variables, but may differ as to overall rates of use.
- **Grammatical Difference Hypothesis:**  
Differences in constraint effects indicate different grammars.

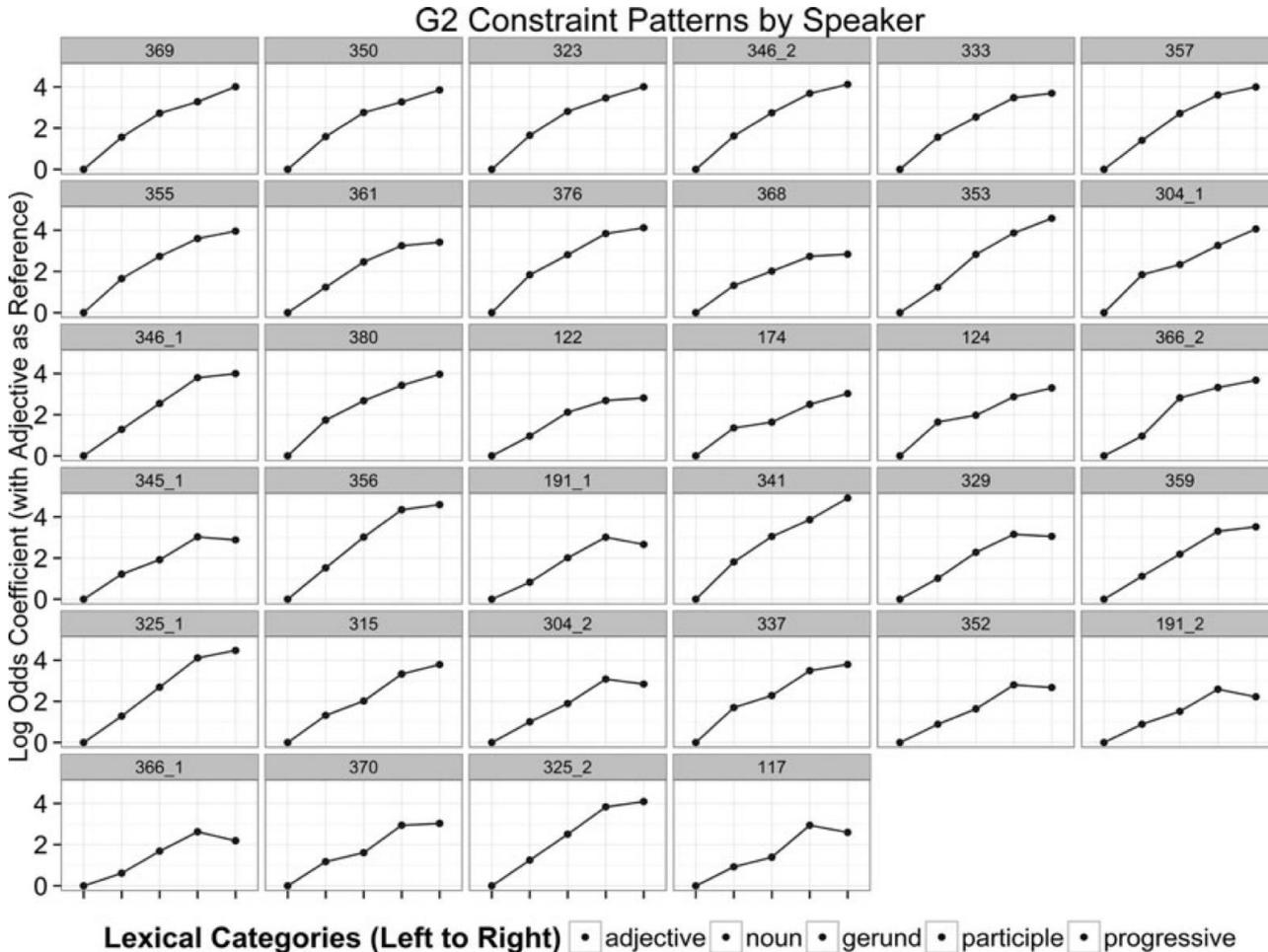
# Popular Brazilian Portuguese: constraints on vowel denasalization

DENASALIZATION	Factor Weights				
	Elvira	Lucia	Bira	Sidnei	23 Spkrs
Preceding Context					
Nasal Consonant	.16	.21	.20	.32	.25
Palatal Consonant	.81	.82	.89	.66	.76
Following Context					
Oral vowel	.84	.79	.76	.79	.76
Oral consonant	.78	.73	.61	.58	.64
Pause	.39	.39	.44	.44	.45
Nasal vowel	.38	.36	.30	.41	.38
Nasal consonant	.12	.22	.22	.26	.26

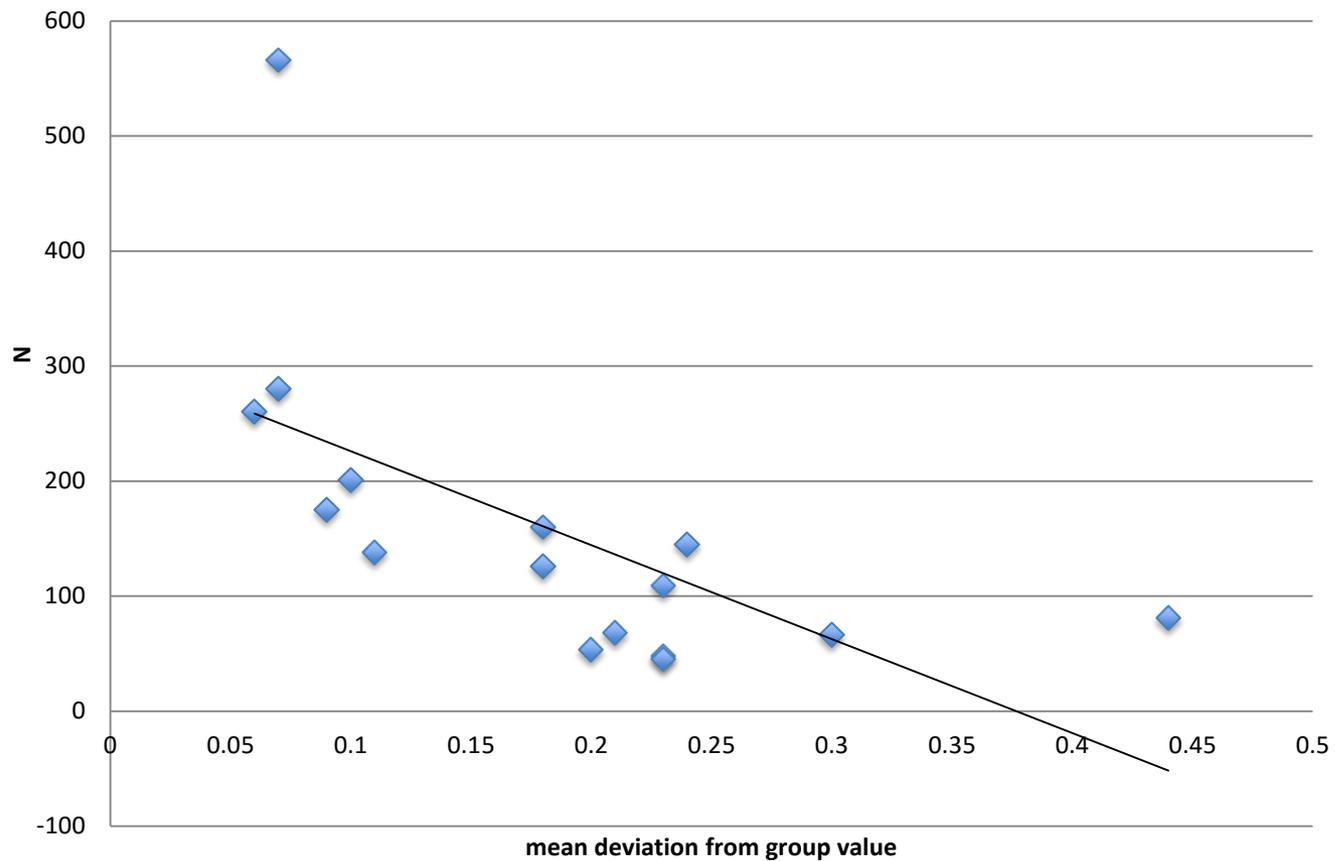
# PBP: constraints on agreement

SUBJECT-VERB AGREEMENT	Elvira	Lucia	Bira	Sidnei	23 spkrs
Morphological class					
1. come-comem	.13	.21	.14	.12	.24
2&3. fala-falam & faz-fazem	.41	.54	.40	.29	.43
4. está-estão	.51	.43	.65	.54	.52
5. sumiu-sumiram	.70	.60	.59	.86	.60
6. falou-falaram, fez-fizeram, é-são, etc.	.80	.74	.77	.73	.72
Subject position					
Immediately preceding	.77	.73	.79	.75	.67
Following	.22	.21	.09	.15	.31
Elsewhere	.51	.57	.73	.65	.52
Plural marking in subject					
Categorical	.72	.64	.61	.64	.65
Variable	.28	.36	.39	.36	.35

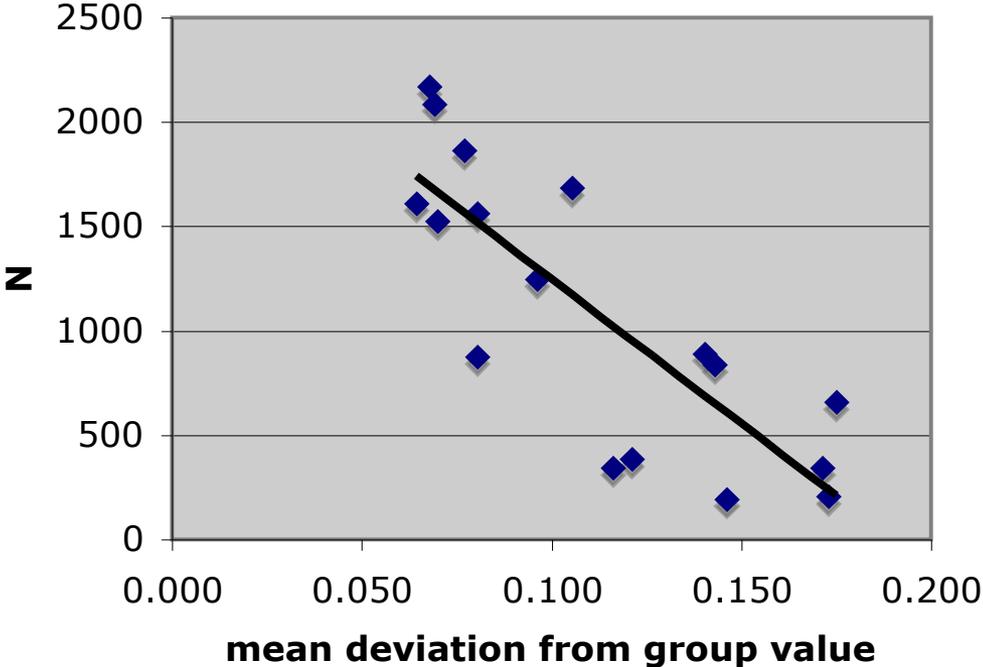
# Constraints on -ing; Forrest 2015



## Mean deviation from group value by data quantity – -t,d deletion; Philadelphia data



### Mean deviation by data quantity

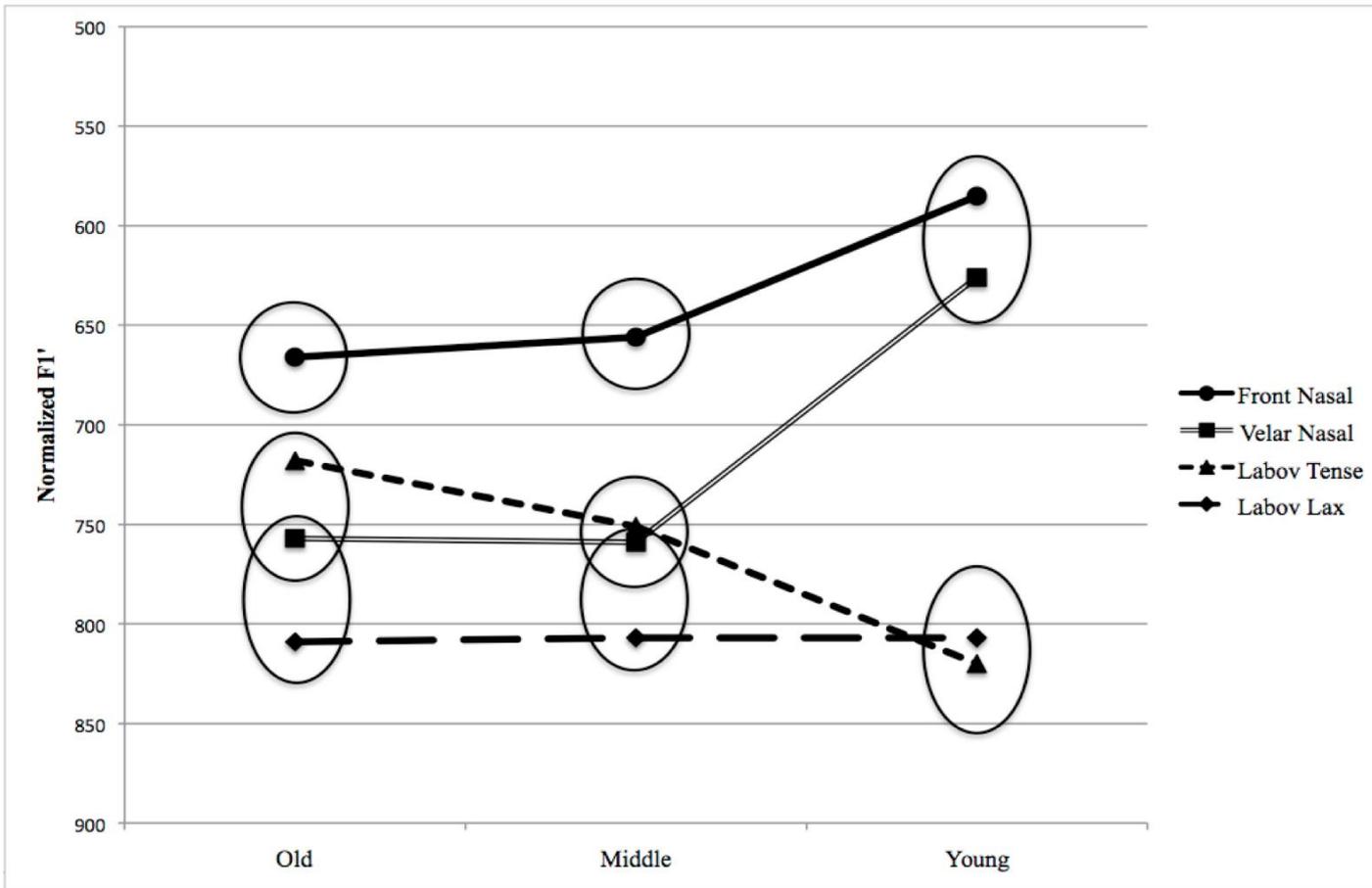


**-t,d deletion; ONZE Data**

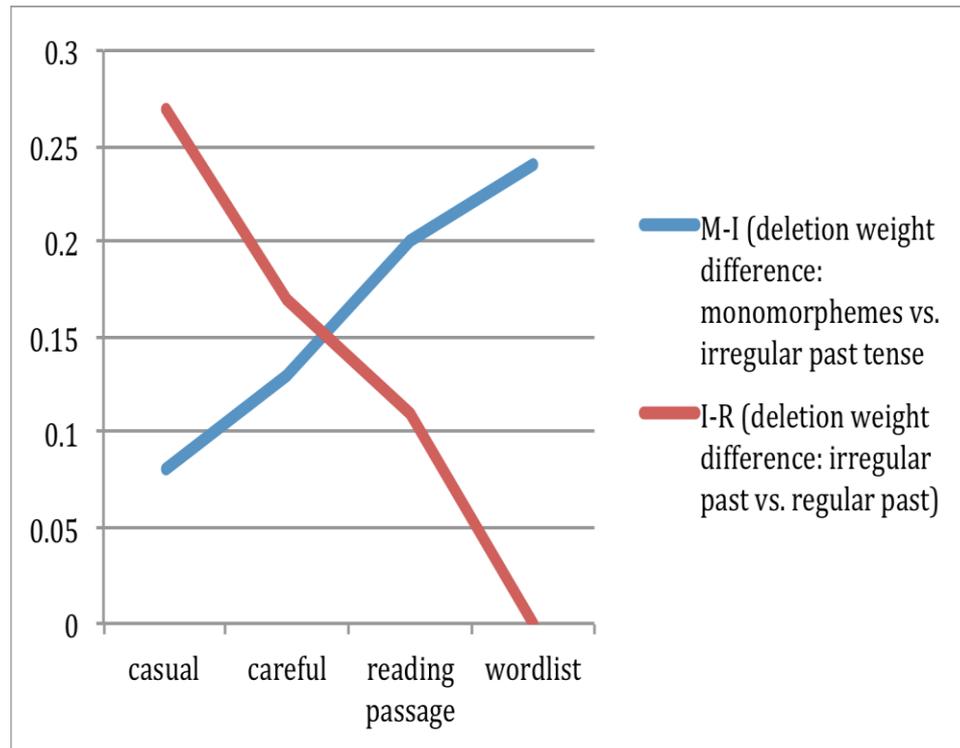
# Contexts for /æ/-tensing in NYCE

- The short-a (BAD/BAT) is shifting from the traditional NYC system (tensing before, *inter alia*, voiceless fricatives, voiced stops, and front nasals) to a nasal system, as found in other AmEng dialects (cf. Becker, Newlin-Lukowicz, etc.)

# Contexts for æ-tensing in NYCE (from Becker 2010)



# Singapore English: constraints on –t,d deletion vary with style



from Lim 2010, Guy & Lim to appear

# Summary and Conclusions

# Summary: Lectal Coherence

- Correlations occur well above the level of chance, but non-correlations also occur
- Some social groups seem more coherent (dialect groups, women, central urban areas)
- Some correlations linguistically driven
- Contextual constraints are stable within a community/ lect/ grammar
- Stable vs. dynamic variables behave differently

# Summary: Stable Variables

- Stable socially stratified variables correlate fairly well (cf. BP agreement, S-deletion, r-retroflexion)
- Dialectal variables correlate well (cf. Spanish pro-drop, –s lenition and deletion)

# Summary: Dynamic Variables

- Changes from below: new indexicalities, uncorrelated with older variables (cf. SP diphthongal eN)
- Changes from above: broadly correlated, but may move at different rates (cf. NYCE rhotacism, /æ/-tensing, *bought*-lowering)

# Summary: Linguistic Constraints

- Constant within a community / grammar
- Linguistic structures or processes may constrain correlations and coherence:
  - motivating correlations (feeding relations, parametric drivers)
  - inhibiting correlations (differences in acquisition or perception)
- No obvious differences between syntactic and phonological variables

# Conclusions, 1

- Social cohesion among variables may be weak, cannot be assumed
- Social variation is polydimensional; therefore patterns of correlation among variables may be complex or obscure
- Variables differ in identity associations
- Variables with common indexicalities show best correlations

# Conclusions, 2

- The data do not support an extreme version of either model:
  - too much clustering for completely free *bricolage*
  - too little for neatly bounded coherent lects
- The co-occurrence of variables is granular: some clusters of features are persistently found, but other features don't correlate

# Drivers of coherence

- Density of communication – shared experience
- The accommodation imperative ‘be understood’
- Common indexicalities among variables

# Drivers of differentiation

- Differences in experience
- The autonomy imperative ('be yourself')
- Innovation (especially 'change from below')
- Styling, stance-taking

# Coherence and *bricolage*

- The lects we name are indeed idealizations
- But community coherence is evident even in identity construction, styling, stance taking
- Bricolage is only communicatively effective against a background of shared community evaluations of the indexicality of variables.

# Towards a coherent theory of social meaning

- The speech community supplies the ‘grammar’
  - High density of communication and mutual accommodation drive linguistic similarity
  - Shared community understandings provide the indexical values of linguistic features
- The individual composes the ‘utterance’
  - Selections from the feature pool assemble indexical references into identities, stances

Danke

Grazie

Dziękuję

Arigato

Obrigado

Merci

Gracias

Thank you!

Comments and requests for copies of this powerpoint:  
[gregory.guy@nyu.edu](mailto:gregory.guy@nyu.edu)