Social meanings across the vowel chart
Variable co-occurrence in California English

Dan Villarreal
d.vill@pitt.edu

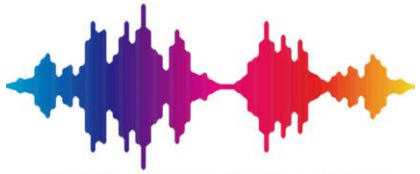
James Grama
james.grama@uni-due.de

American Dialect Society
January 7, 2024

University of Pittsburgh

UNIVERSITÄT DUISBURG ESSEN

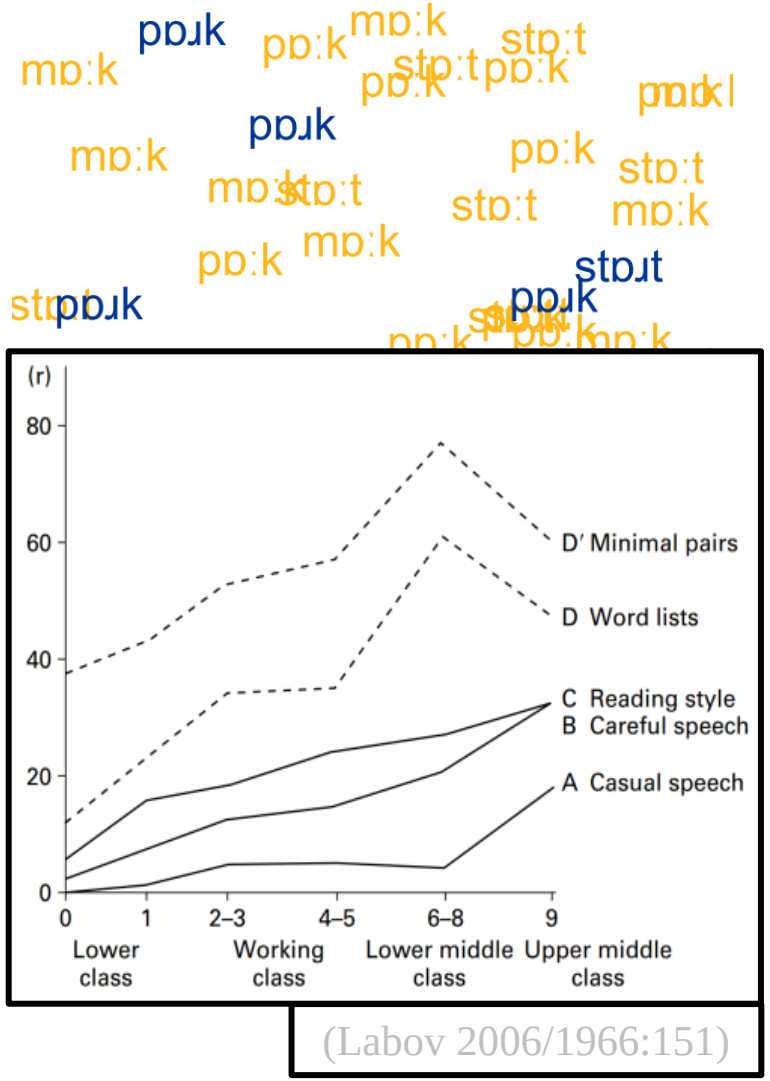
Offen im Denken



SociolinguisticsLab

The power of individual variables

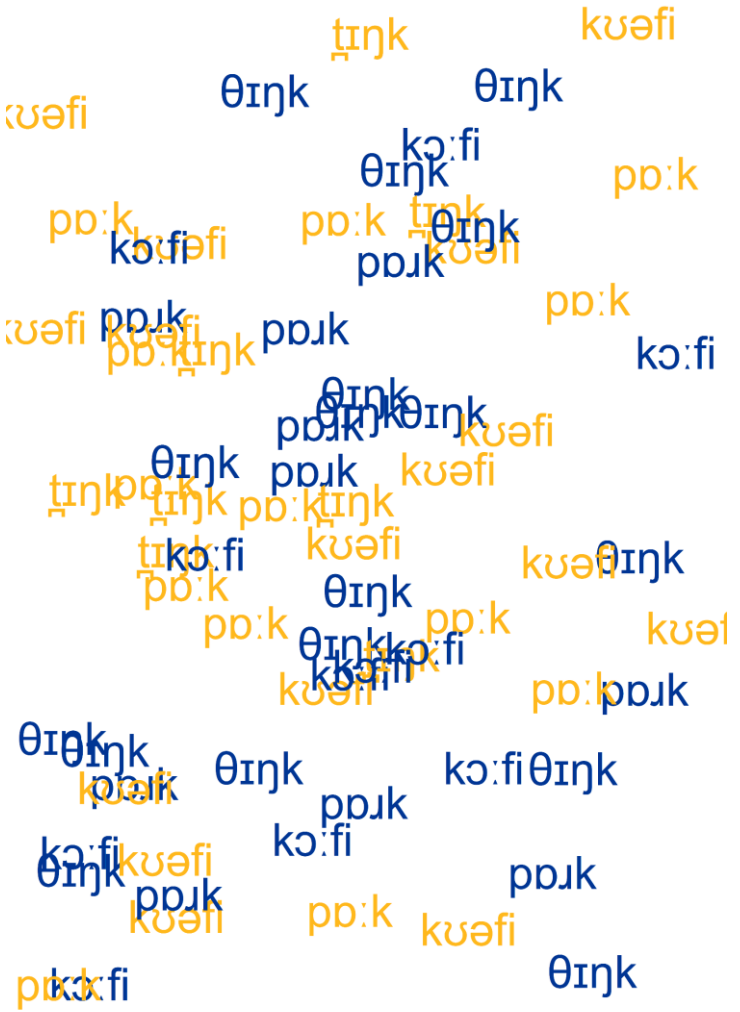
- Social processes are reflected and reproduced in speakers' production and listeners' perception of individual linguistic variables
  - Ex. Hidden in the apparent "free variation" of (r)…
  - …are socioeconomic stratification and class anxiety
- In other words, there's considerable descriptive power in these momentary instances of speech, even if isolated from their context



(Labov 2006/1966:151)

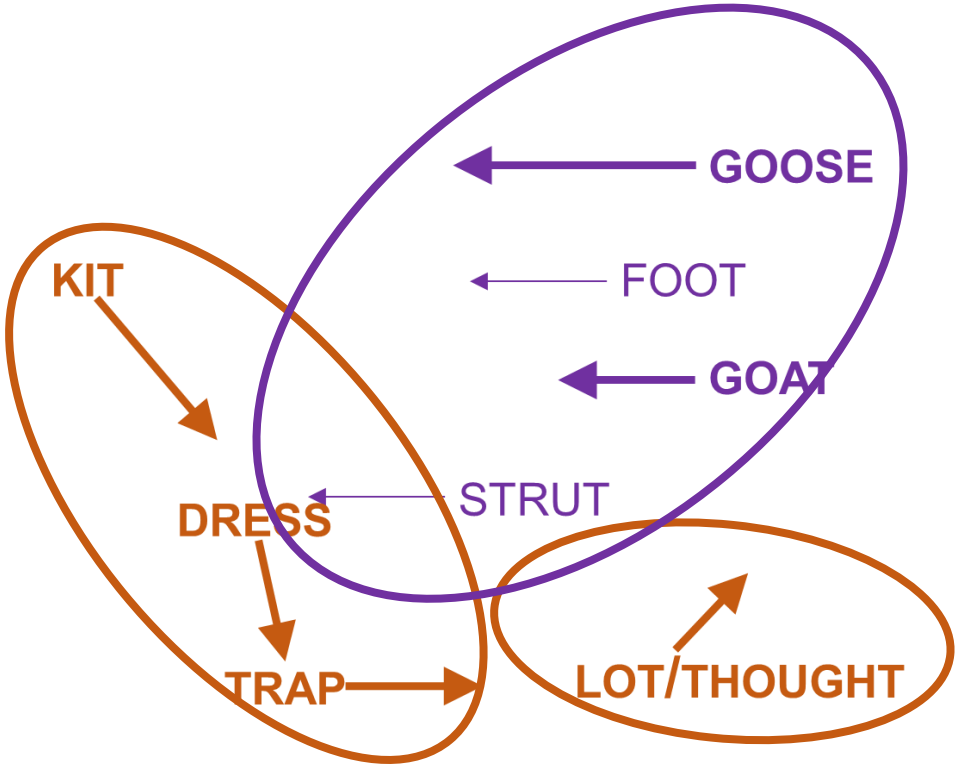
But individual variables aren't all

- Speakers and listeners don't experience language variation as individual variables connected by disparate instances, but as a multiplicity of variables in even short stretches of running speech
- How do we deal with this problem of **variable co-occurrence**?
- Look at a whole bunch of variables?
  - But then we drown in a sea of data
- Appeal to supra-variable structures (style, register)?
  - But we still need to understand individual variables to understand how they form styles/registers



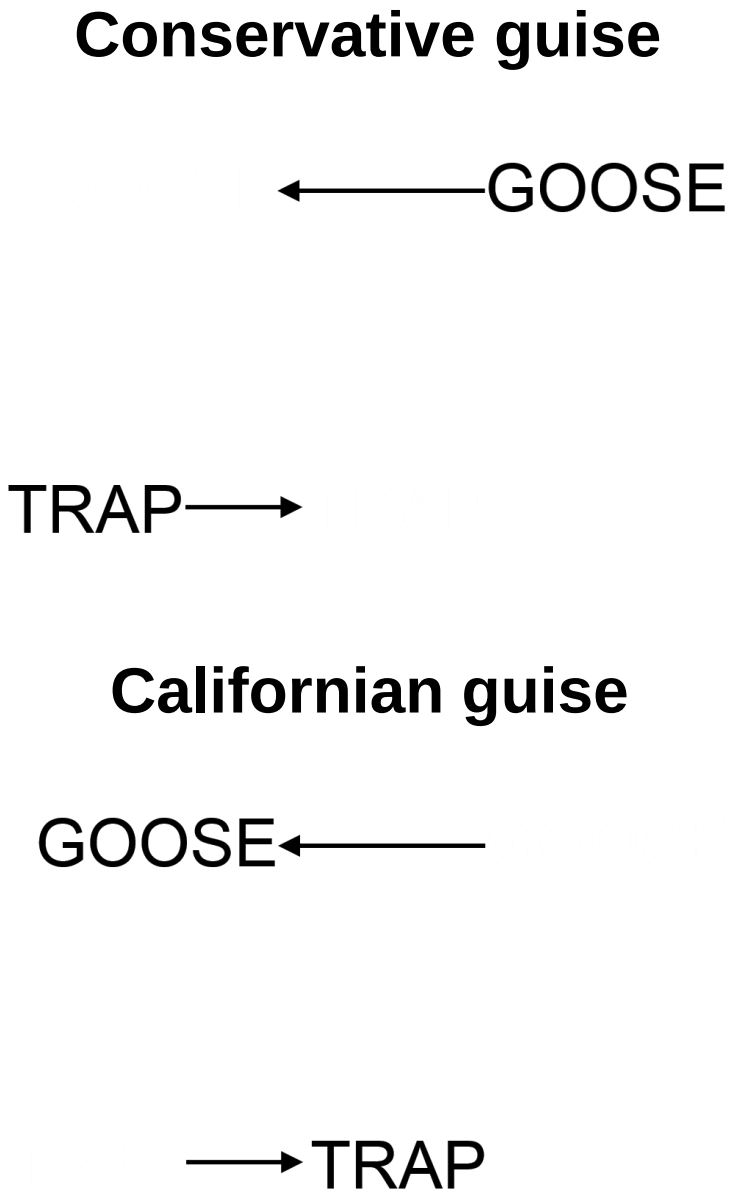
California English vowels

- ***Low-Back-Merger Shift*** (Becker 2019)
  - Merging of LOT, THOUGHT
  - Lowering/retraction of TRAP, DRESS, KIT (in that order, diachronically)
- ***Back Vowel Fronting***
  - Fronting of GOOSE, GOAT
  - Less attention paid to FOOT, STRUT
- These variables not only co-occur in speech, they're structurally related



Villarreal (2018)

- Today, we **re-analyze Villarreal's (2018)** matched-guise technique study of California English social meanings
- Californian listeners (*n*=97) rated stimuli on 12 attribute scales
- TRAP/GOOSE manipulated into two guises
- Carrier phrases: 10-12s clips from spontaneous-speech cartoon retells
  - Across stimuli: Same topic, diff. linguistic content
  - 12 speakers



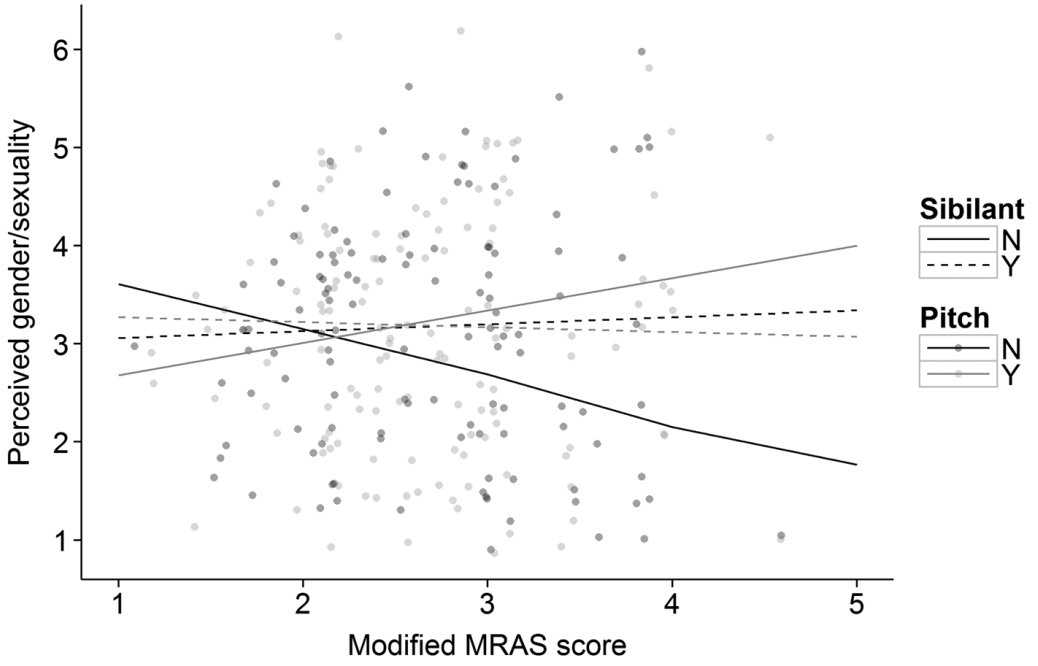
Villarreal (2018)

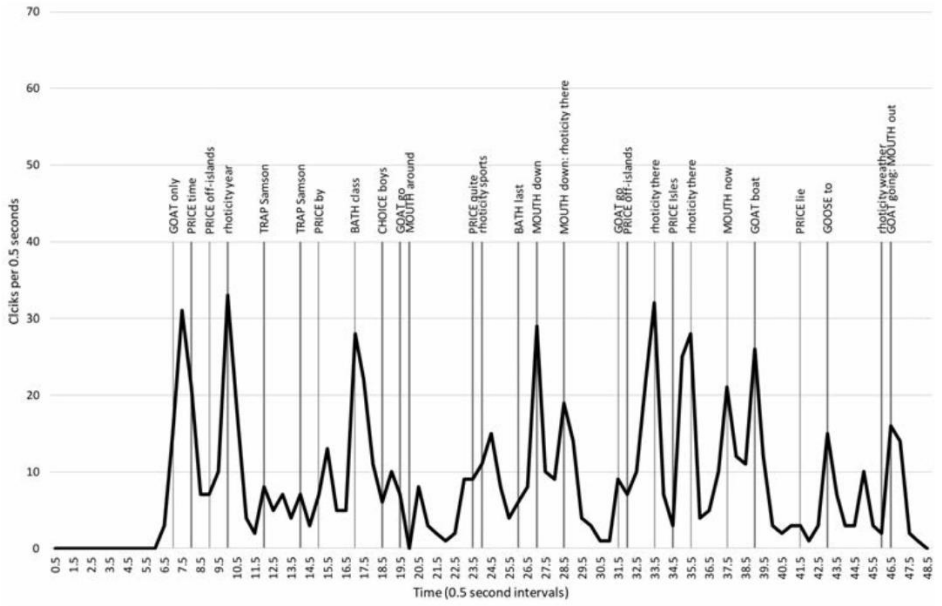
- Cali. guise rated sig. higher for *Californian, sounds like a Valley girl, confident*
- But! There was substantial variance in ratings that was captured by predictors other than guise
- Suggests social meanings were sensitive to both TRAP/GOOSE **and** other cues in the signal

Scale	Guise	Region	Gender	Ethnicity
suitable			✓	✓
feminine		✓	✓	✓
Californian	✓		✓	✓
fast		✓	✓	✓
young		✓		✓
confident	✓	✓	✓	✓
relaxed			✓	✓
familiar				✓
rich		✓		✓
friendly		✓	✓	✓
speaks like me				✓
Valley girl	✓	✓	✓	
# sig. effects	**3**	**8**	**8**	**11**

Researcher choice

- Our hypothesis: the smaller guise effect is due to **researcher choice**: hypothesizing that a particular variable is socially meaningful
- Sociolinguists are hardly immune to overlooking variation meaningful to communities (e.g., Wolfram 2007)
- Other approaches: factorial MGT design (e.g., Levon 2014), real-time reactions (e.g., Montgomery and Moore 2018)





The present study

- Given the limited effect of the guise manipulation compared to other social predictors in Villarreal 2018, this data is ripe for re-analysis via a "bottom-up" approach that mitigates researcher choice
- Research questions:
  1. Which variables (vowels) most impacted listeners' perceptions? Are these the same that researchers pay the most attention to?
  2. Within these variables, which *features* most impacted perceptions?

Methods: Re-analysis

- **Problem:** Analyzing multiple cues means dealing with sparsity in features

(A) **Stu** has a **list** of things that he wants to pack and so he walks offscreen to **go** and and find those things and then Matt the cat is **meanwhile putting** all of his cat toys inside of the **suitcase**

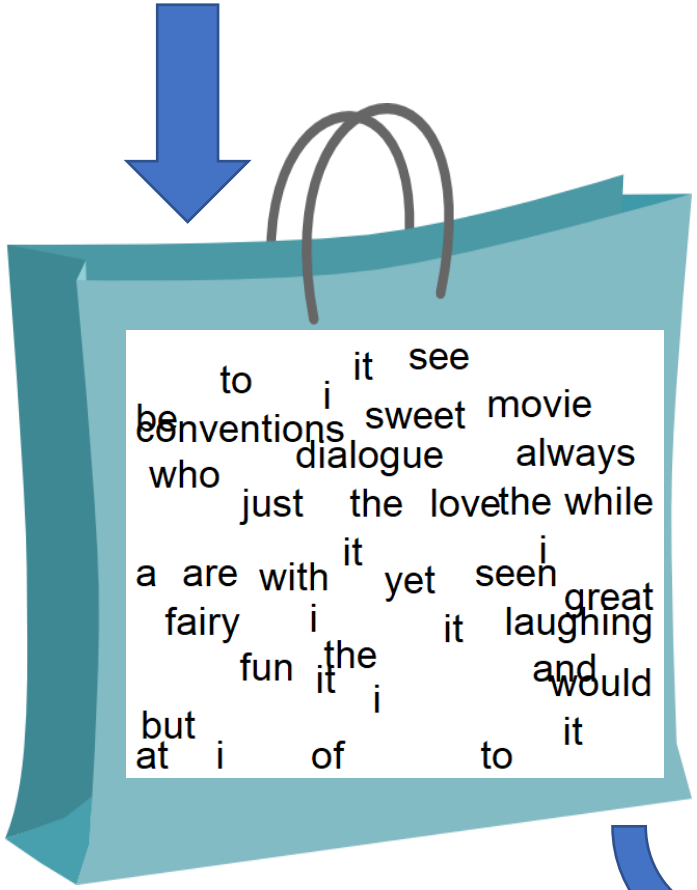
(B) **Stu's** just **sittin'** on the couch **readin'** a **book** and Matt the cat's just **sittin'** on the couch like on the **edge** and uh **Stu sees** the spider crawlin' around he starts freakin' out and he starts pointin' at it and **he's** tryin' to tell like Matt the cat to **get** it

Variable	(A)	(B)
FLEECE	1	3
KIT	1	2
FACE	1	0
DRESS	0	2
TRAP	4	5
PRICE	1	1
LOT	0	1
STRUT	0	0
GOAT	1	0
FOOT	1	1
GOOSE	2	2

Methods: Re-analysis

- **Problem:** Analyzing multiple cues means dealing with sparsity in features
  - *Bag of words* (from comp ling): representation of texts as counts of lexical items, ignoring word order & grammatical relations
  - **Solution!** "Bag of features"

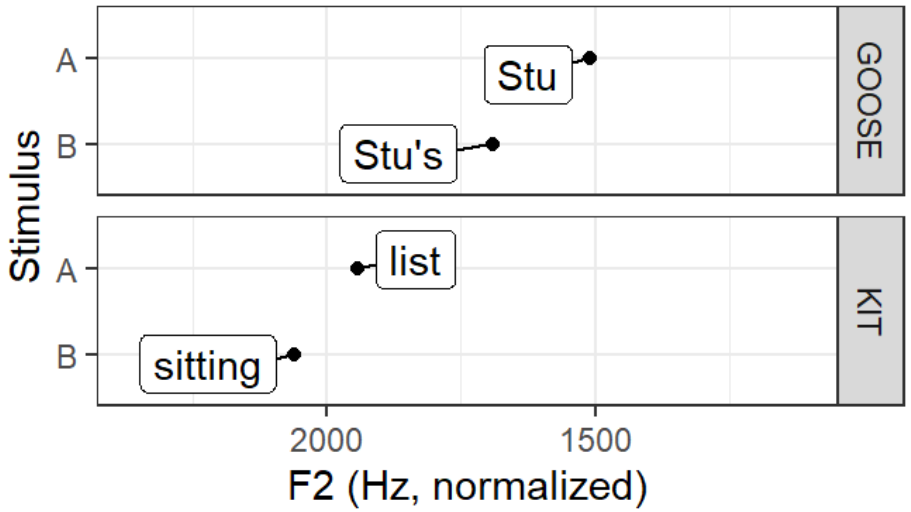
I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre.



Word	#
it	6
I	5
the	4
to	3
and	3
seen	2
yet	1
would	1
with	1
who	1
whimsical	1
…	

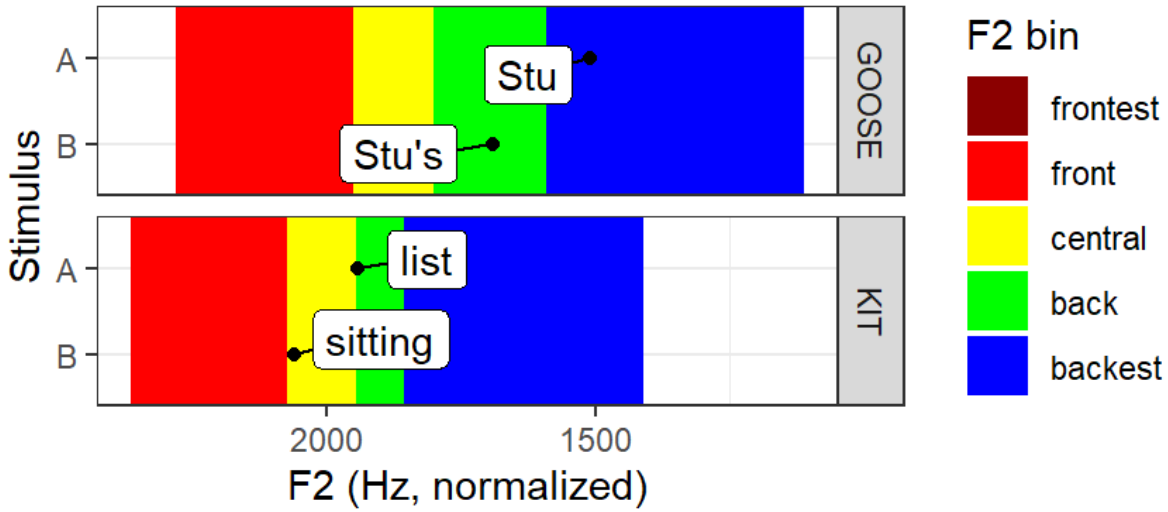
Methods: Re-analysis

- **Problem:** Analyzing multiple cues means dealing with sparsity in features
  - **Solution!** "Bag of features"
- **Problem:** Converting continuous variation into counts

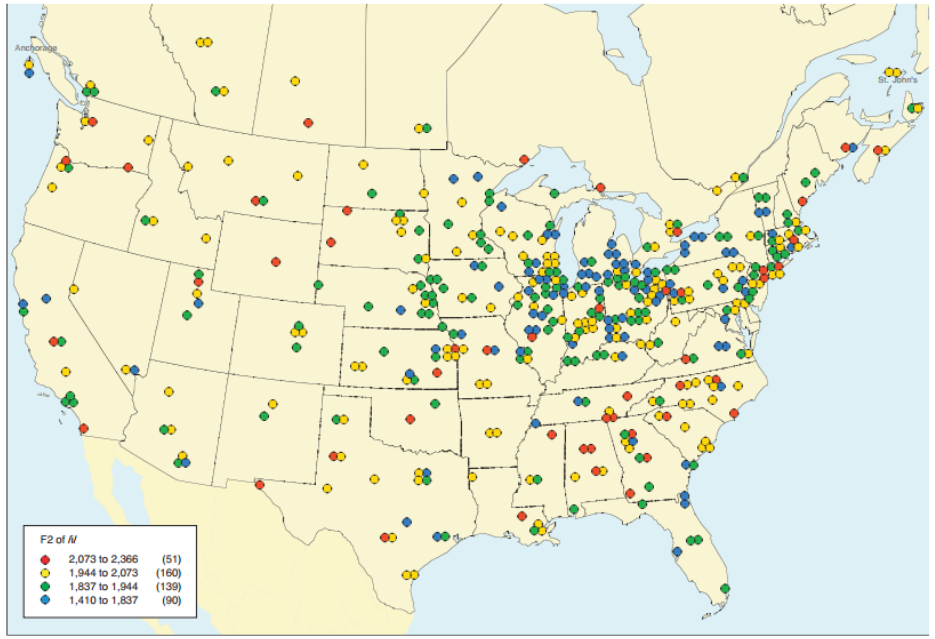


Methods: Re-analysis

- **Problem:** Analyzing multiple cues means dealing with sparsity in features
  - **Solution!** "Bag of features"
- **Problem:** Converting continuous variation into counts
  - **Solution!** ANAE benchmarks



ANAE (Labov et al. 2006) p. 79

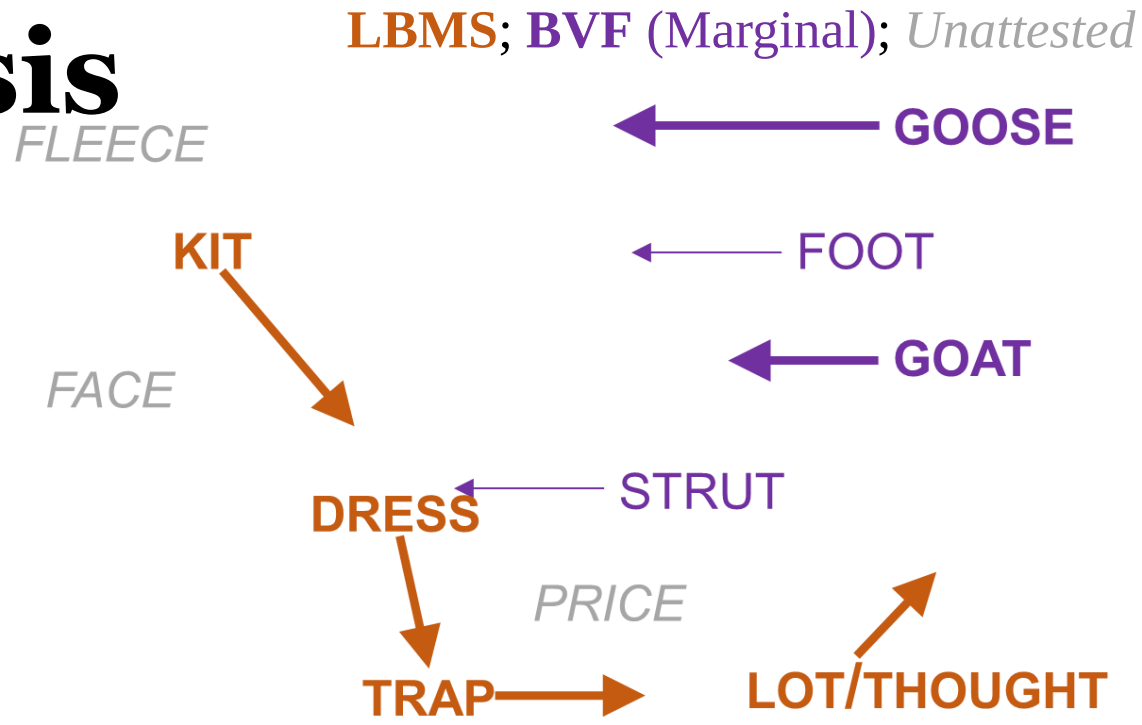


Map 10.2. The relative fronting and backing (F2) of */i/* in *bit*, *hid*, etc.

Feature	(A)	(B)
GOOSE_F2_backest	1	0
GOOSE_F2_back	0	1
KIT_F2_back	1	0
KIT_F2_central	0	1
…		

Methods: Re-analysis

- Features across 11 phonemes, incl. 3 with no attested evidence of change in California
  - 379 tokens across 24 stimuli
- Used Boruta algorithm (Kursa & Rudnicki 2010) to determine which feature(s) impacted dependent variable (cf. Dickson & Durantin 2019)
  - Separate runs for each attribute scale using `Boruta` R package



Scale	Feature	Result
Californian	FACE_F2_central	Rejected
Californian	FOOT_F1_highest	Rejected
rich	FACE_F2_central	Rejected
rich	FOOT_F1_highest	Confirmed
…		

Findings 1
Which variables most impacted social meanings?

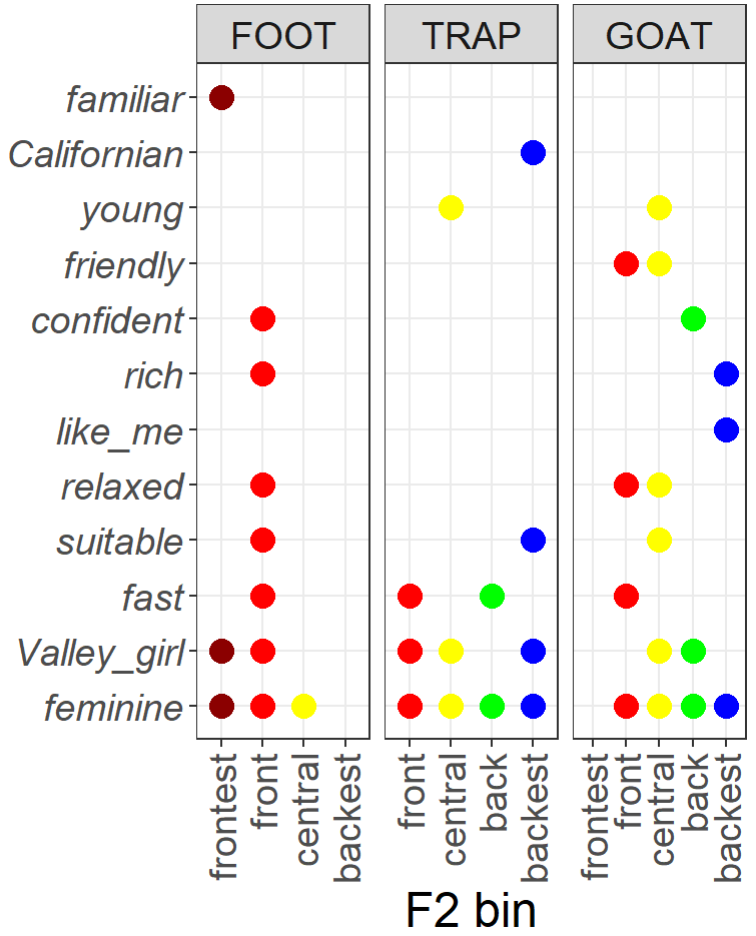
- For each variable, how many scales (out of 12) selected 1+ of that variable's features?
- Most-studied variables in California English not necessarily more impactful
  - FLEECE (almost completely unattested as changing) outranks several well-attested California English variables
- Selection frequencies do not reflect diachronic ordering (i.e., TRAP then DRESS then KIT)
  - Supports Eckert and Labov's (2017) claim that individual variables (not structural phenomena) are available for social meaning

Variable	F1	F2
FOOT	10	8
TRAP	8	6
GOAT	4	10
FLEECE	7	6
KIT	6	6
FACE	4	6
LOT	5	5
GOOSE	5	5
PRICE	5	4
STRUT	5	3
DRESS	3	4

LBMS; **BVF** (Marginal); *Unattested*

Findings 2
Which features most impacted social meanings?

- Within a variable, not all variants carry the same indexical oomph (e.g., Montgomery and Moore 2018)
- This holds for F2 features of FOOT, TRAP, GOAT
  - Central/backest FOOT appears unmarked, while frontest/front FOOT carries meaning
  - By contrast, for GOAT different bins affected different meanings: *friendly*/*young*/*relaxed* affected by front/central GOAT, but *confident*/*rich*/*speaks like me* affected by back/backest GOAT



Discussion 2
Methodological caveats

- This method reduces researcher choice, but isn't immune to it
  - Bag of features included 11 vowels, but left out consonants, prosody, etc.
  - Selection of scales in a given rating task ≠ range of possible social meanings
- Vowel variation needs to be discretized into features using benchmarks extrinsic to the data set (*does not* have to be ANAE)
  - ANAE-type benchmarks not widely available for other varieties
- Variable selection is not a Swiss army knife
  - Doesn't replace ethnographically informed accounts of social meaning construction in interactional moments; can't match experimental control of matched-guise technique
  - Next step for us: Drill down into directionality of effects

Variable co-occurrence and researcher choice

- As researchers, we choose to focus on particular variables for particular reasons—which may not lead us to the variables that are actually most salient to listeners
- Actually accounting for variable co-occurrence and mitigating researcher choice are challenging, but "bottom-up" approaches represent a step in that direction
- In modeling language variation, we need to pay more attention to variable co-occurrence (not just w/rt social meaning)
  - Speakers and listeners experience variables in context

Acknowledgments

- Dan's dissertation committee: Bob Bayley, Vai Ramanathan, Valerie Fridland, Santiago Barreda
  - Unofficial committee member: Christina Calvillo
- Funding: UC Davis Department of Linguistics
- Feedback: Nicole Holliday, Carina Ahrens, and Anne-Marie Mölders

Thank you!

Contact:
d.vill@pitt.edu
james.grama@uni-due.de

University of Pittsburgh

UNIVERSITÄT DUISBURG ESSEN
Offen im Denken

SociolinguisticsLab