

Evaluating Mechanisms Underlying the Geographic Localization of Lexical Innovation



Aparna Ananthasubramaniam
aparna-ananth.github.io
[@AparnaAnanth729](https://twitter.com/AparnaAnanth729)
akananth@umich.edu



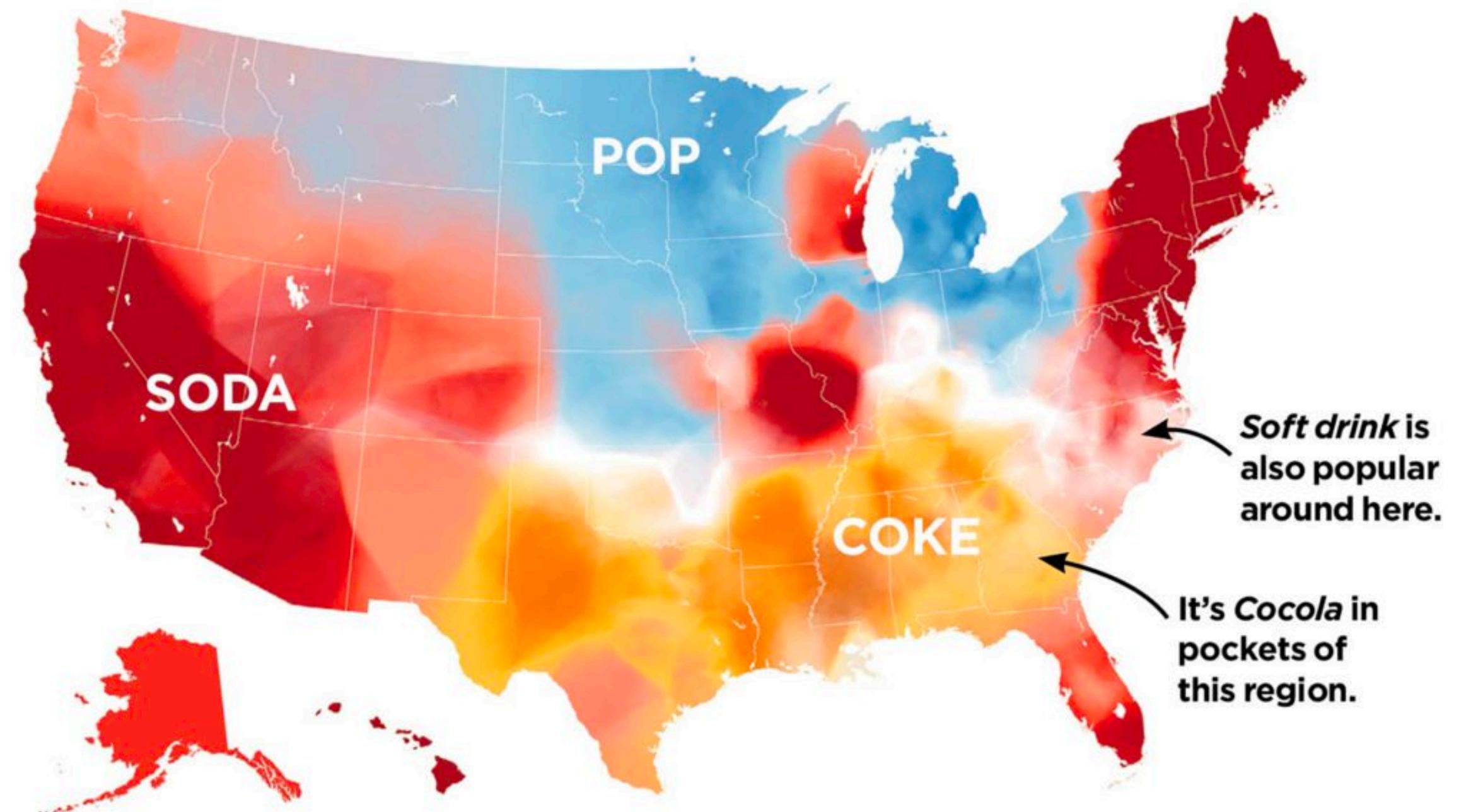
David Jurgens
jurgens.people.si.umich.edu
[@DAVID__JURGENS](https://twitter.com/DAVID__JURGENS)
jurgens@umich.edu



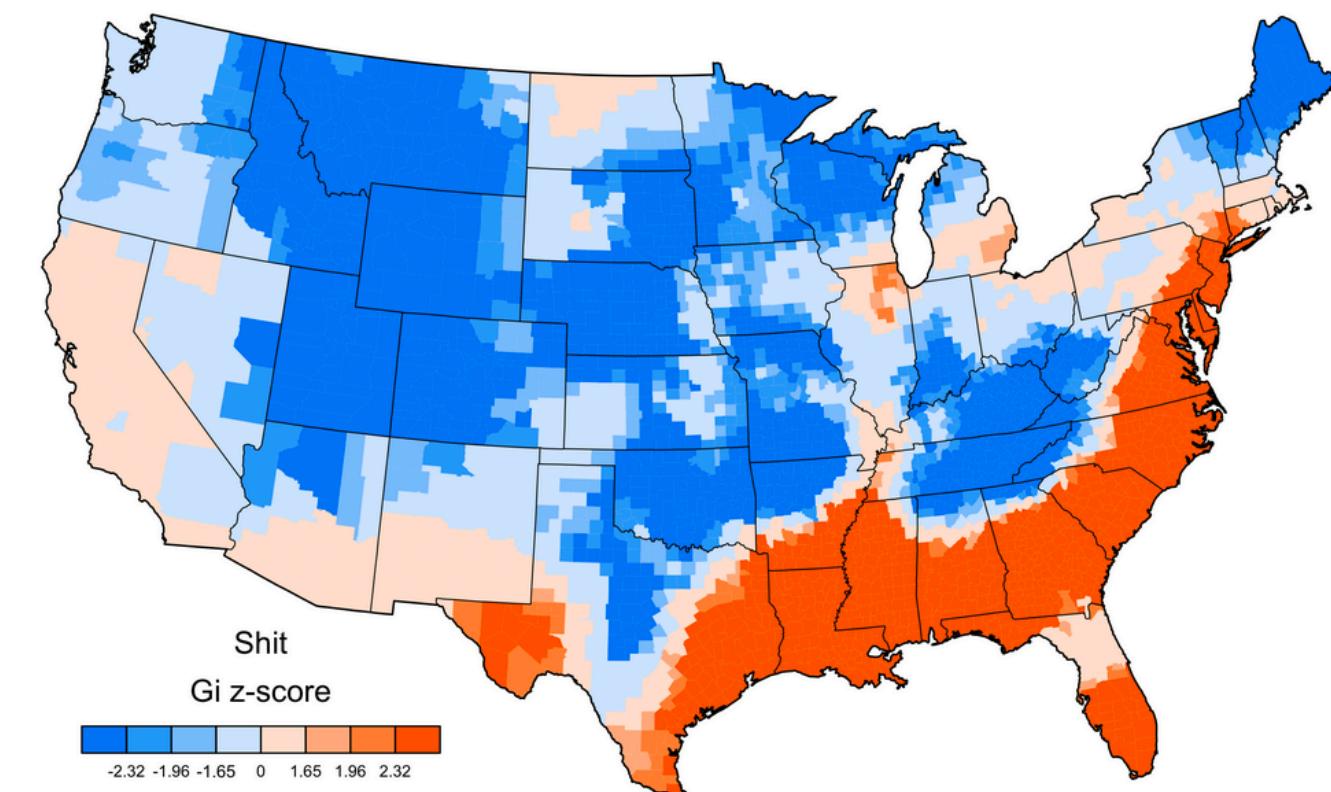
Daniel Romero
dromero.org
[@DanielMRomero](https://twitter.com/DanielMRomero)
drom@umich.edu

Why and how does linguistic “localization” happen?

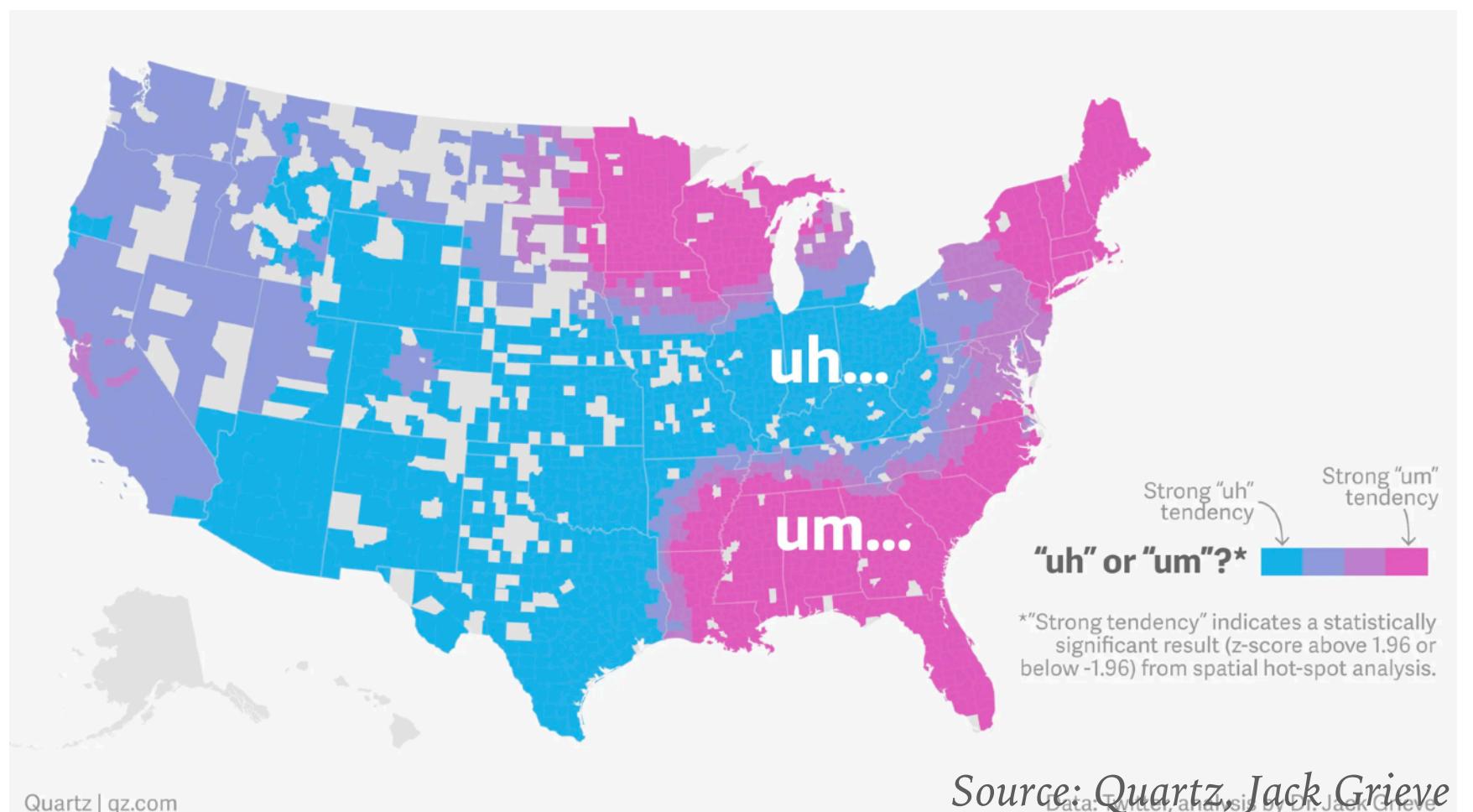
What is your generic term for a sweetened carbonated beverage?



Source: Reader's Digest, Josh Katz



Source: Stan Carey, Jack Grieve



Source: Quartz, Jack Grieve

Spatial Distribution of Linguistic Variables is Well-Studied

- Dialectologists, sociolinguists, and historical linguists catalogue linguistic regions and their evolution over time

Source: Labov 1963

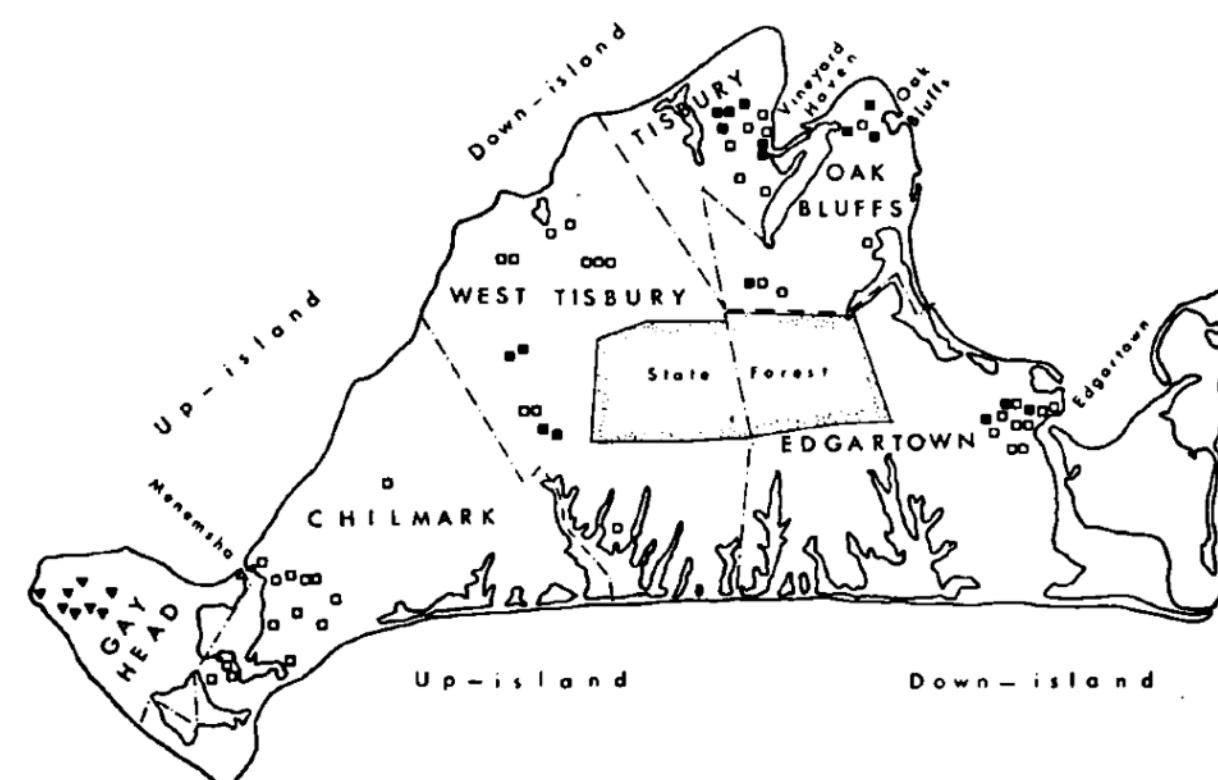
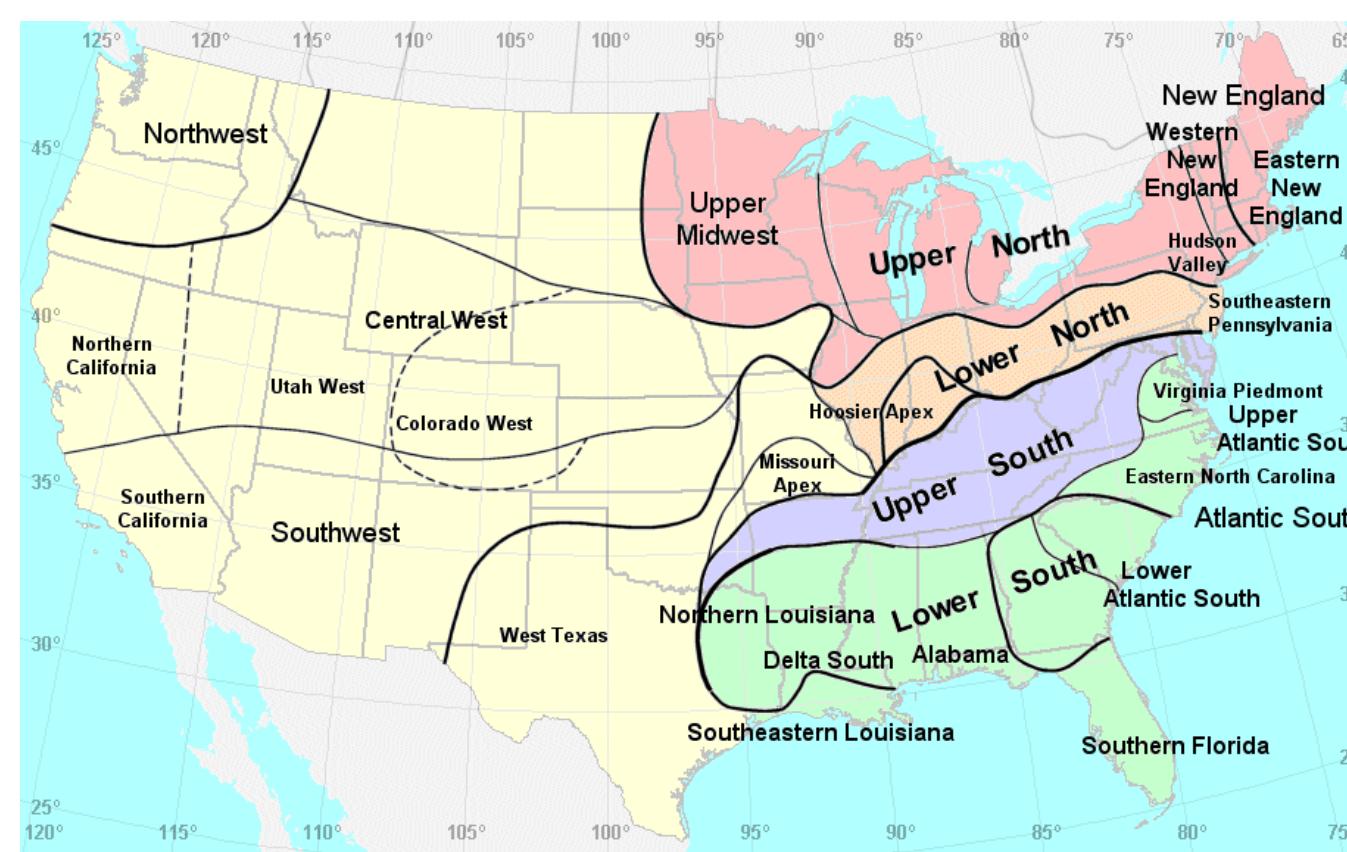
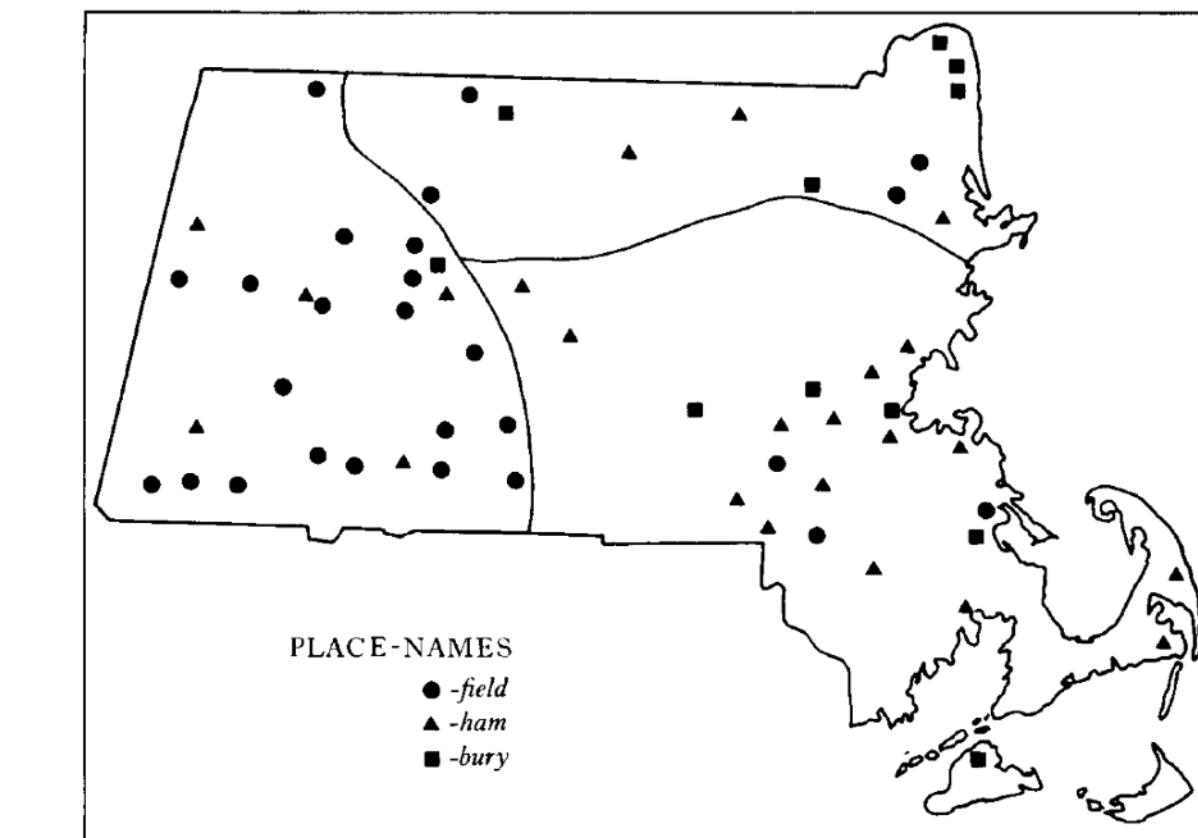


FIGURE 1. Location of the 69 informants on Martha's Vineyard. Ethnic origin of the informants indicated by the following symbols: □ English, ■ Portuguese, ▼ Indian. Symbols placed side by side indicate members of the same family.

Source: Labov, Ash, Boberg 2006



Source: Chambers and Trudgill 1980

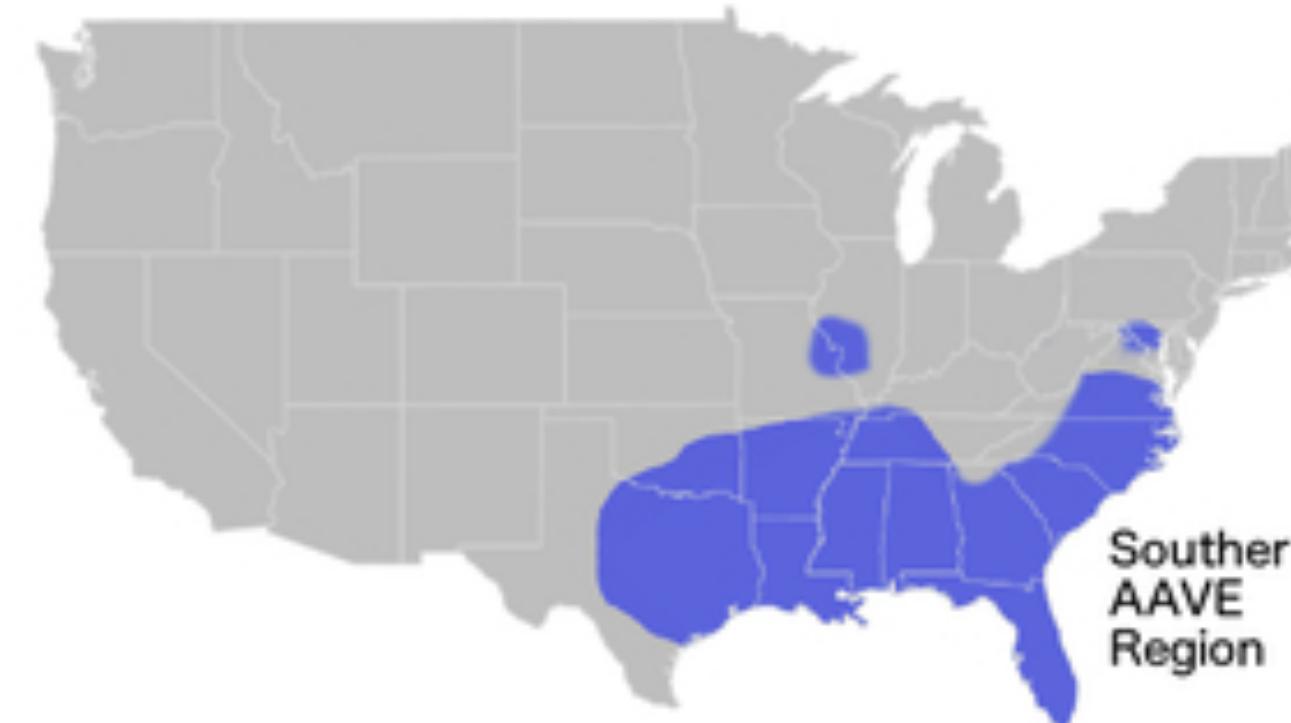


Map 7-7. The state of Massachusetts, showing the major isogloss bundles which divide it into three regions, and the occurrence of certain place-names prior to the American Revolution (after Green and Green 1971). Each of the regions shows a preference for a different type of place-name

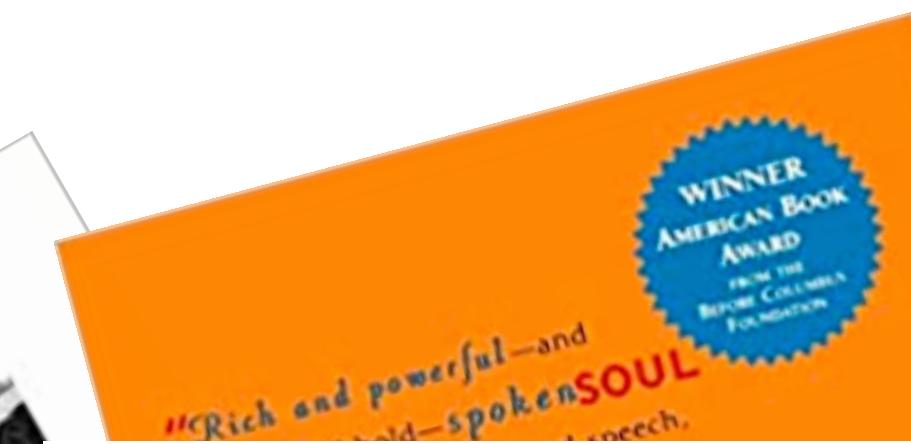
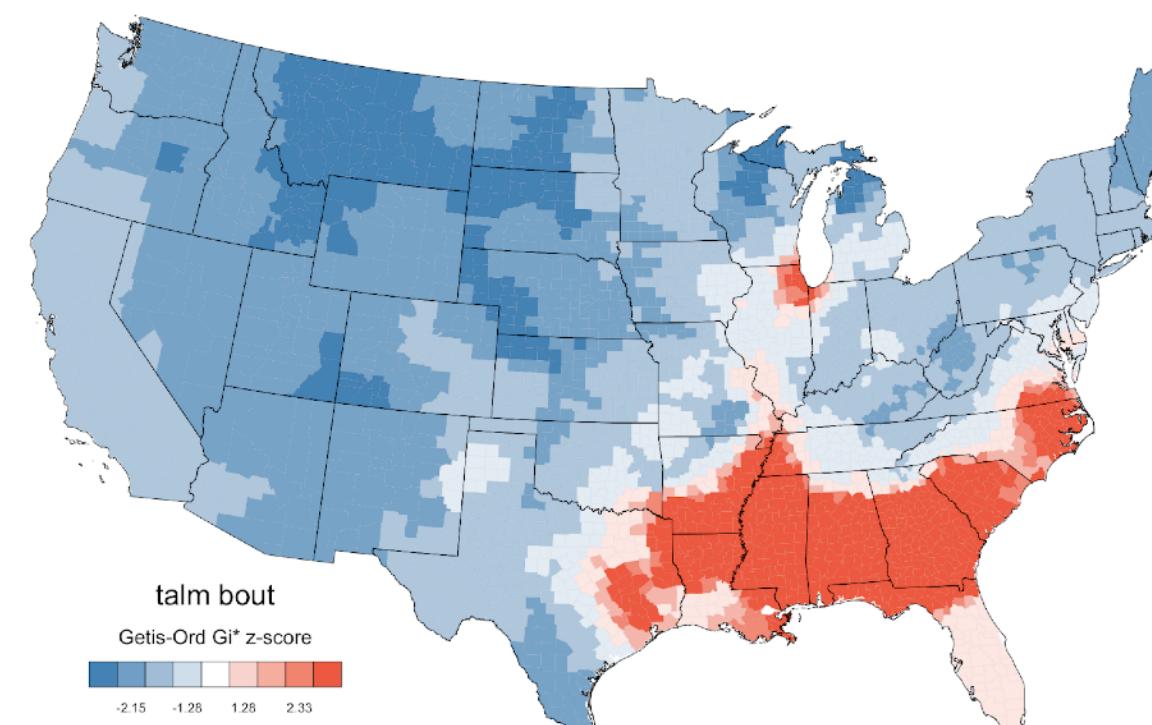
Studying Regional Variation in Language has Broad Applicability

- Linguistic regions are culturally, socially, and historically significant

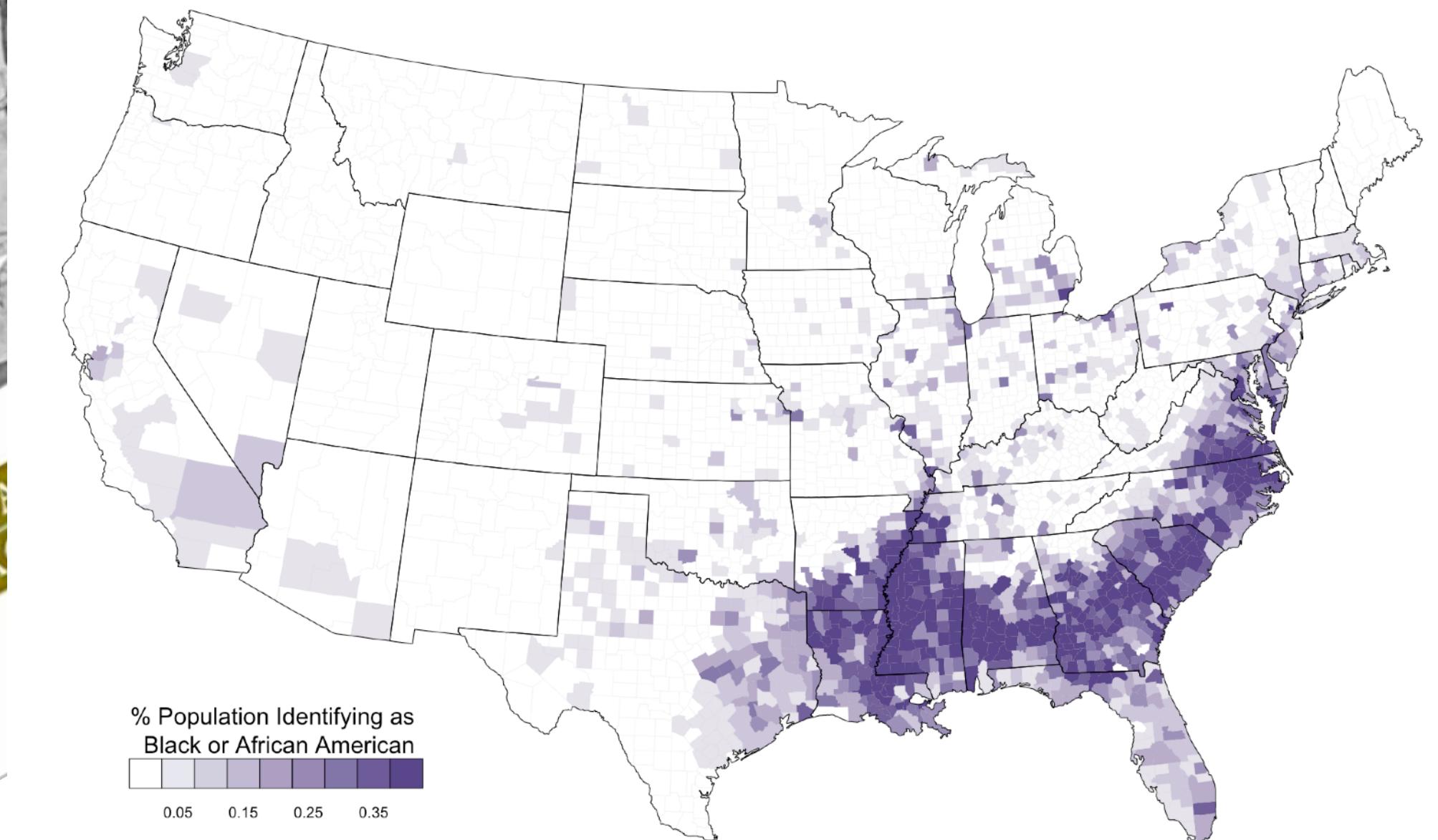
African American English (Source: Jones 2015)



Use of “talm bout” on Twitter



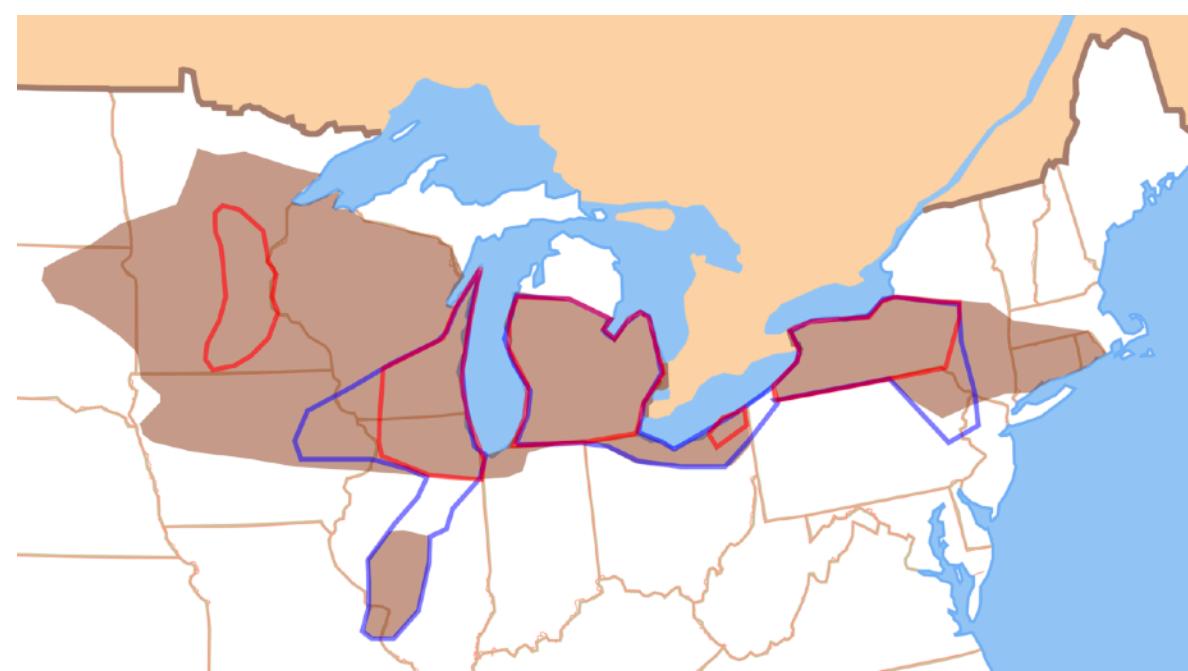
US Black or African American Population (Source: Census Bureau)



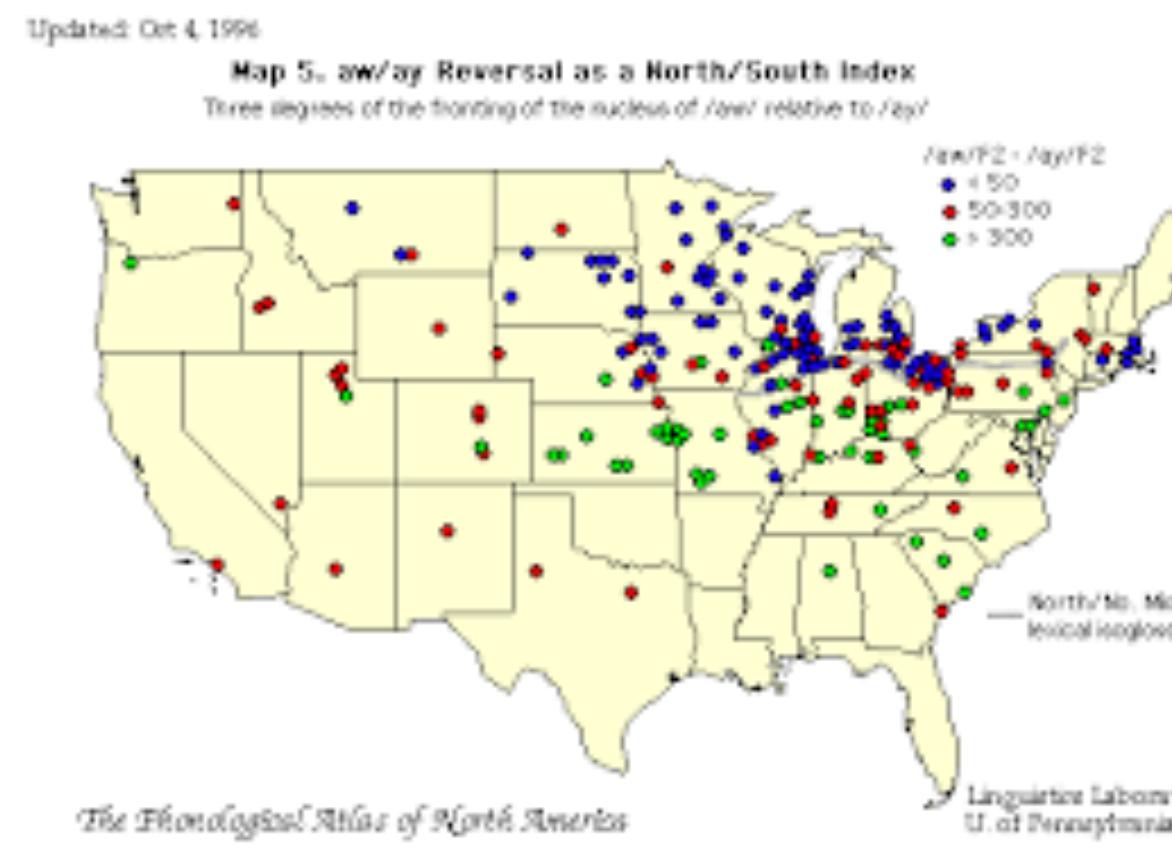
Studying Regional Variation in Language has Broad Applicability

- Linguistic regions are culturally, socially, and historically significant

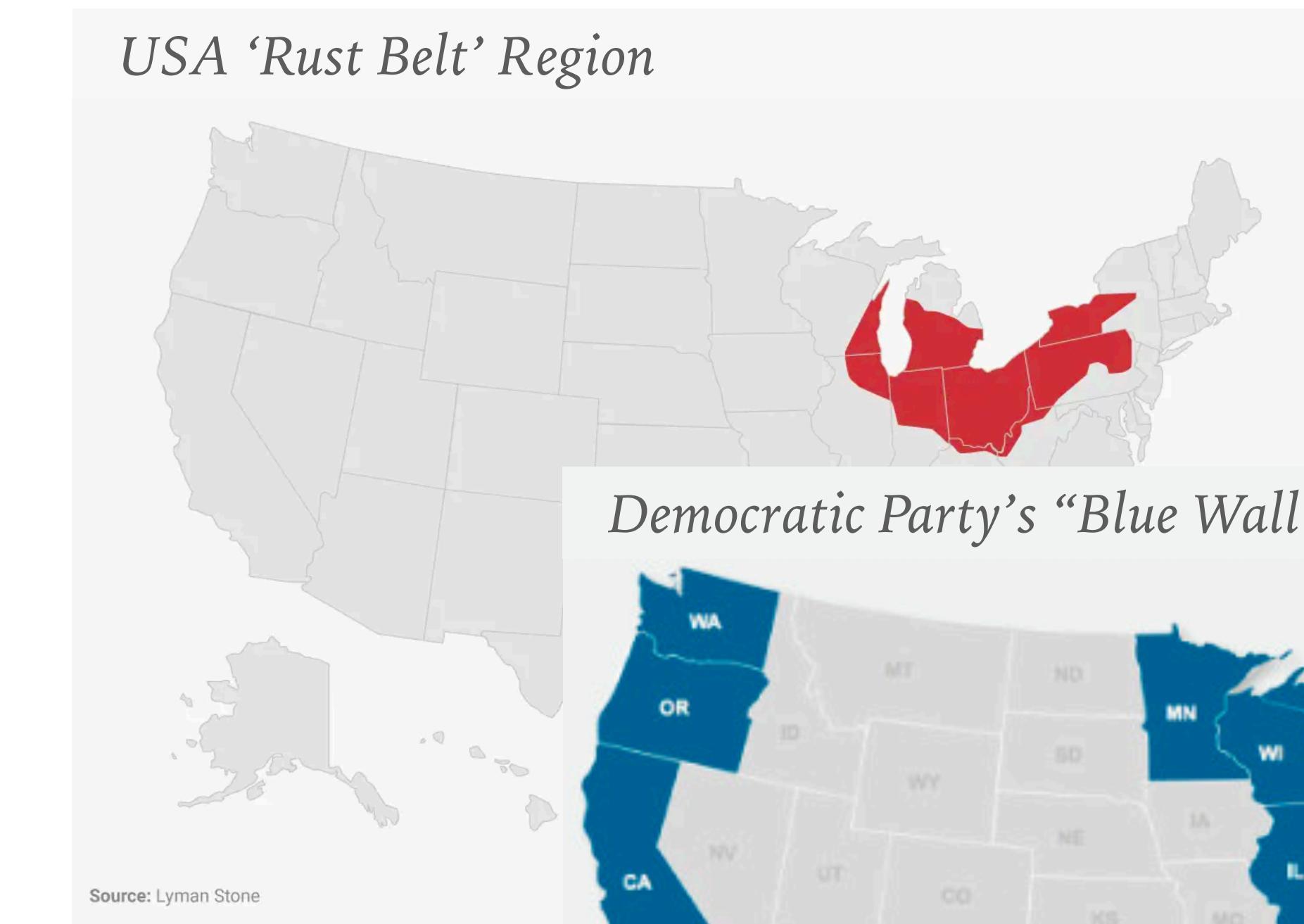
Isoglosses for Northern Cities Vowel Shift
(Source: Labov, Ash, Boberg 2006)



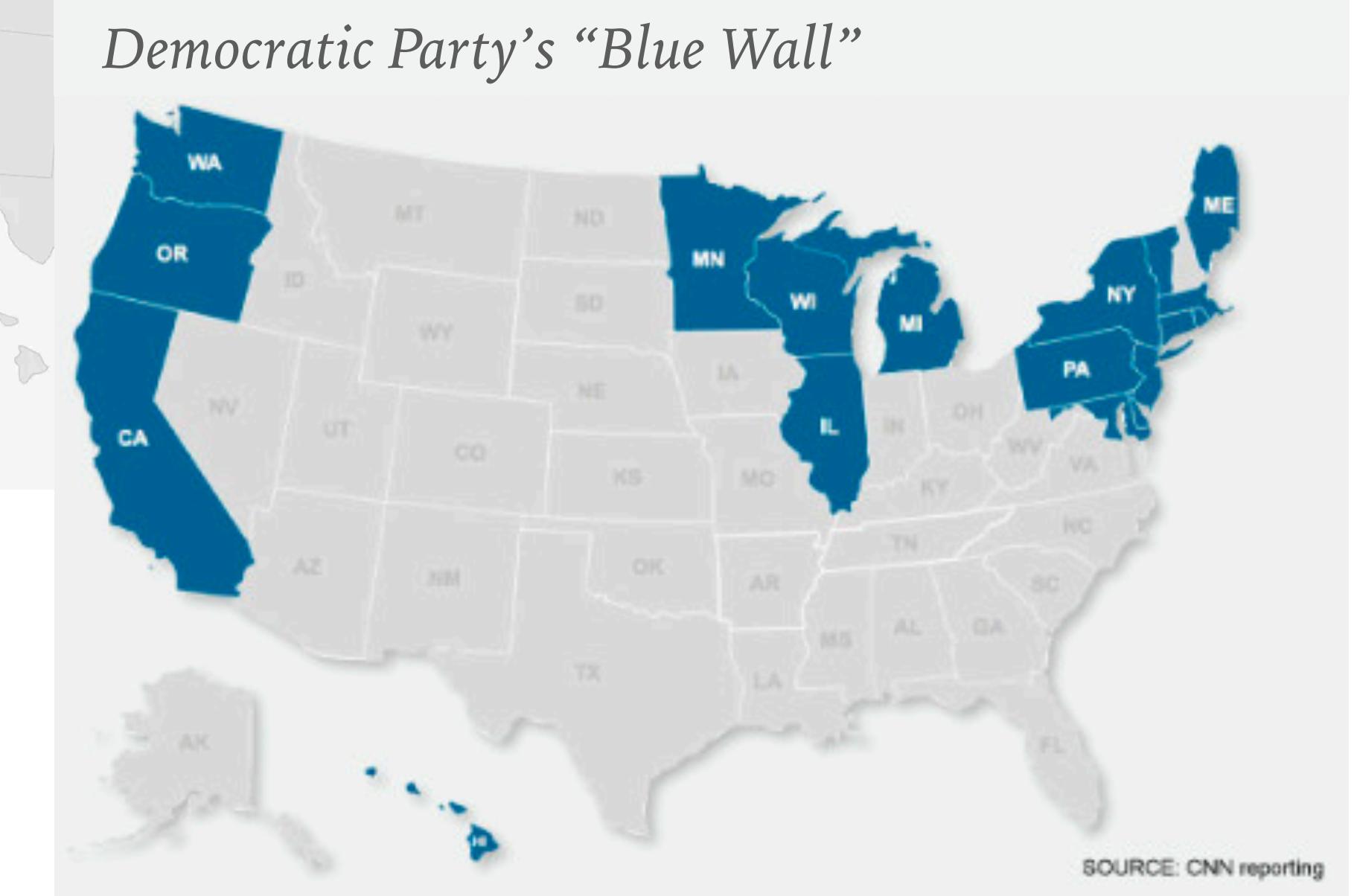
Phonological Atlas of North America



USA ‘Rust Belt’ Region



Democratic Party’s “Blue Wall”



Hypotheses about Underlying Mechanisms have not been Directly Tested

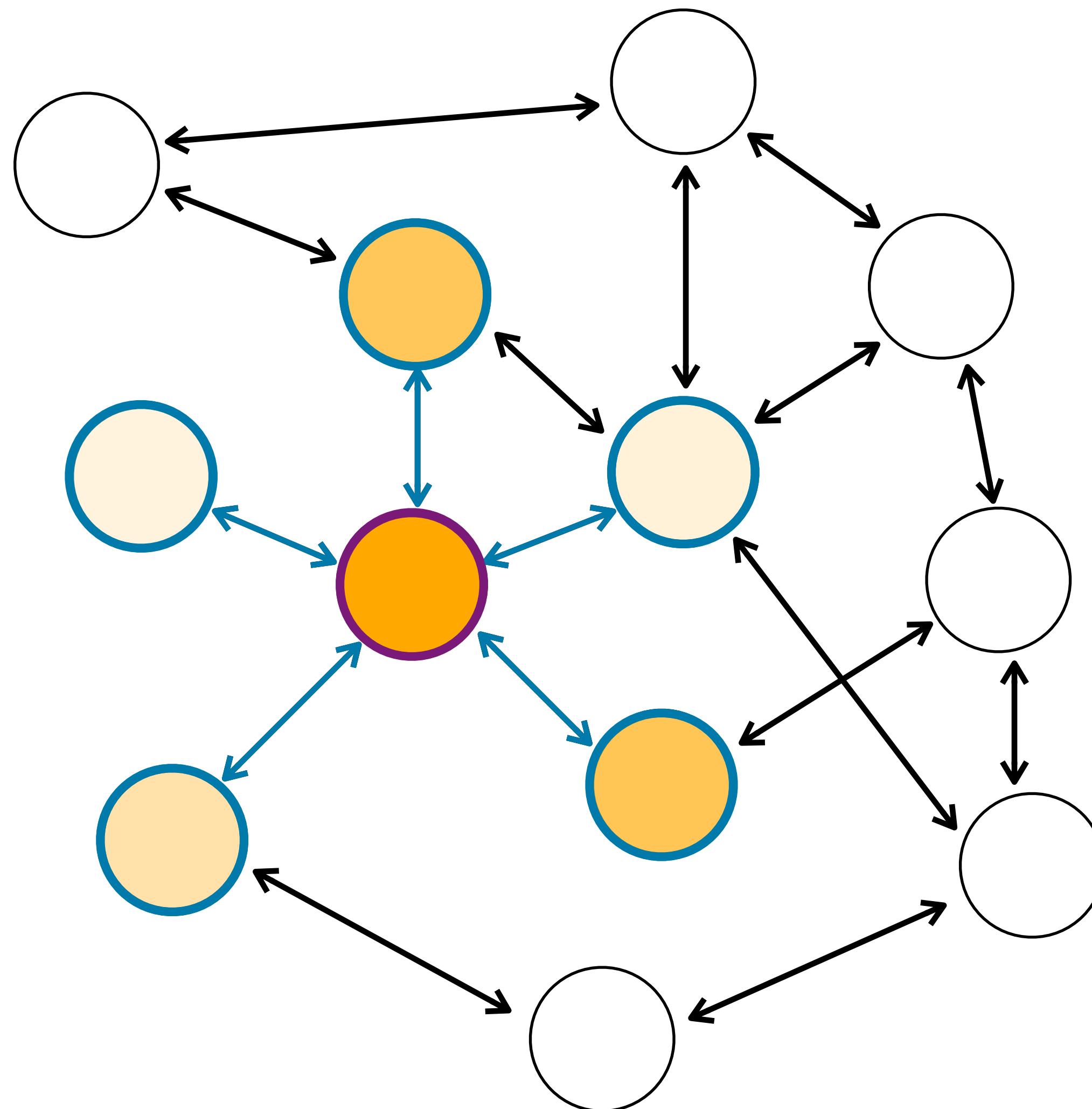
- Linguistic variables (e.g., phonemes, words) and regions are well-defined, measurable
- Mechanisms driving localization are difficult to observe
 - Hypothesized variables are abstract and complex
 - Internal states, like social and cognitive processes, are not measurable
 - Inference from individual speech patterns is intractable



Our Work in Progress

- Main Idea: Linguistic localization can be used to **evaluate hypotheses** about the adoption of language variation and lexical innovation (Cangelosi 2012)
- Approach: Use **agent-based models** to test counterfactuals
- Research Questions:
 - What social and cognitive **variables** generate the important language regions?
 - What are the **mechanisms** underlying localization of linguistic variables?

Agent-Based Model Formulation



After an **agent** is exposed to the new word:

1. **Agent** decides whether to use the word based on their **mental representation**
2. If used, **their connections** are “exposed”
3. **They** adjust **their mental representations**:
 - Fraction of connections using the word
 - “Stickiness” of the word, cognitive attention
 - Identity (demographic): race, SES, geography

Model Parameters

Estimated from Twitter:

- Network (mutual mentions)
- Tie strength
- Agent characteristics:
 - Geolocation (disclosed)
 - Demographics (Census)

Model Parameters (hard to measure):

- Word characteristics
- Cognitive characteristics

Validation:

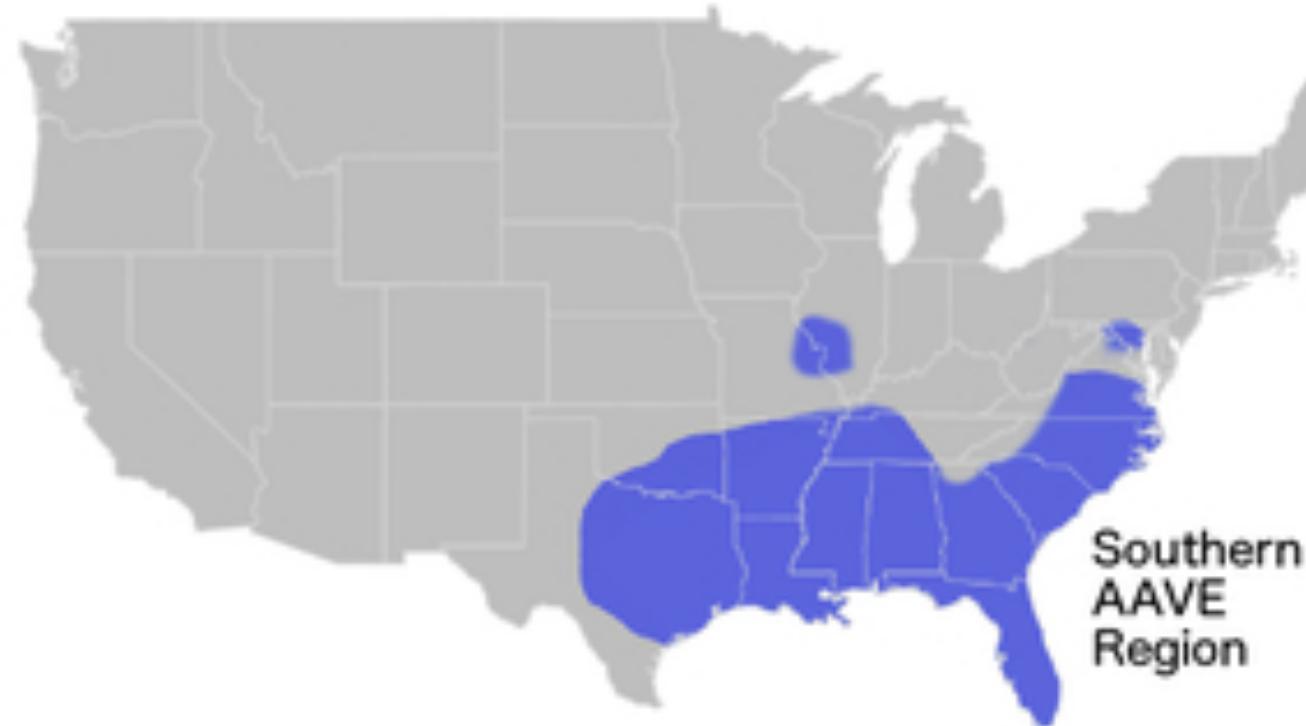
- Localization of lexical innovation

Early Results

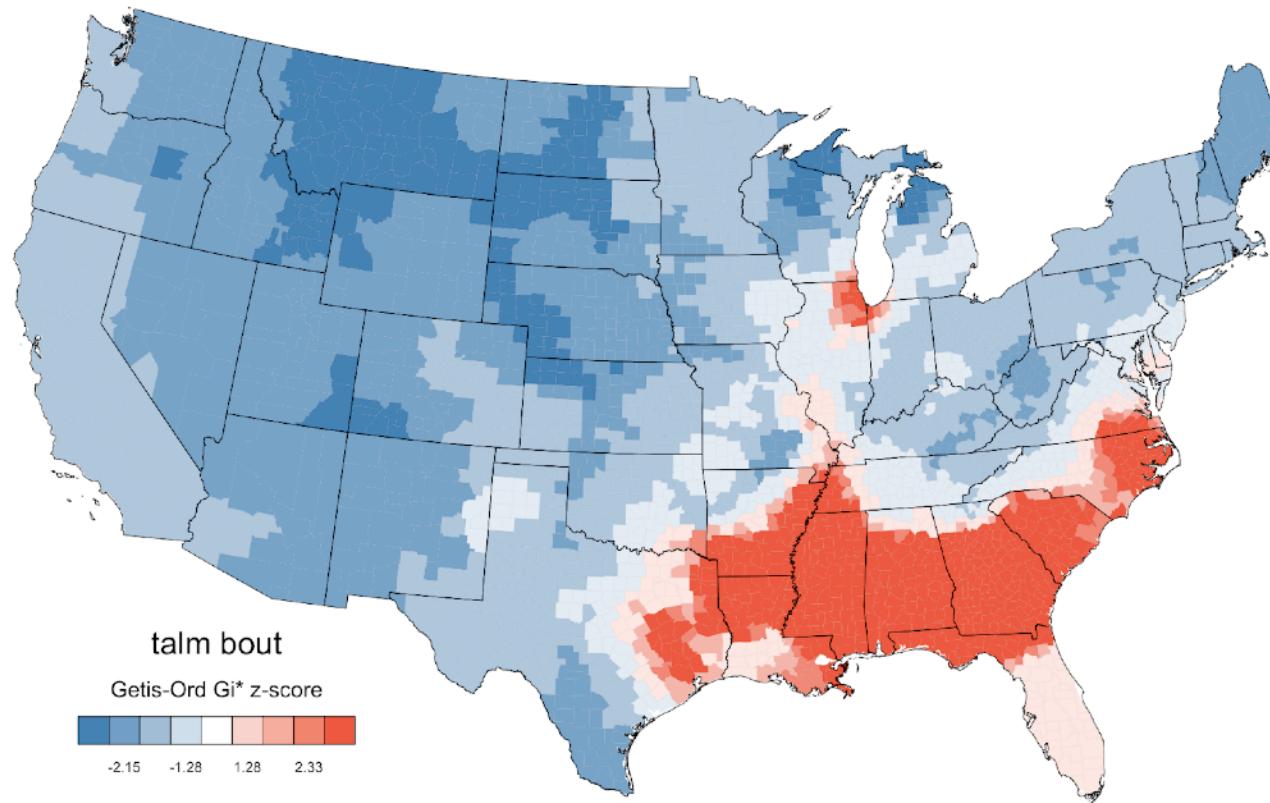
- Words in our model localize to important linguistic regions
- These regions are the product of the network topology and speaker sociodemographic identities
- Future work: testing hypotheses about mechanisms underlying diffusion of linguistic variation and innovation

African American English

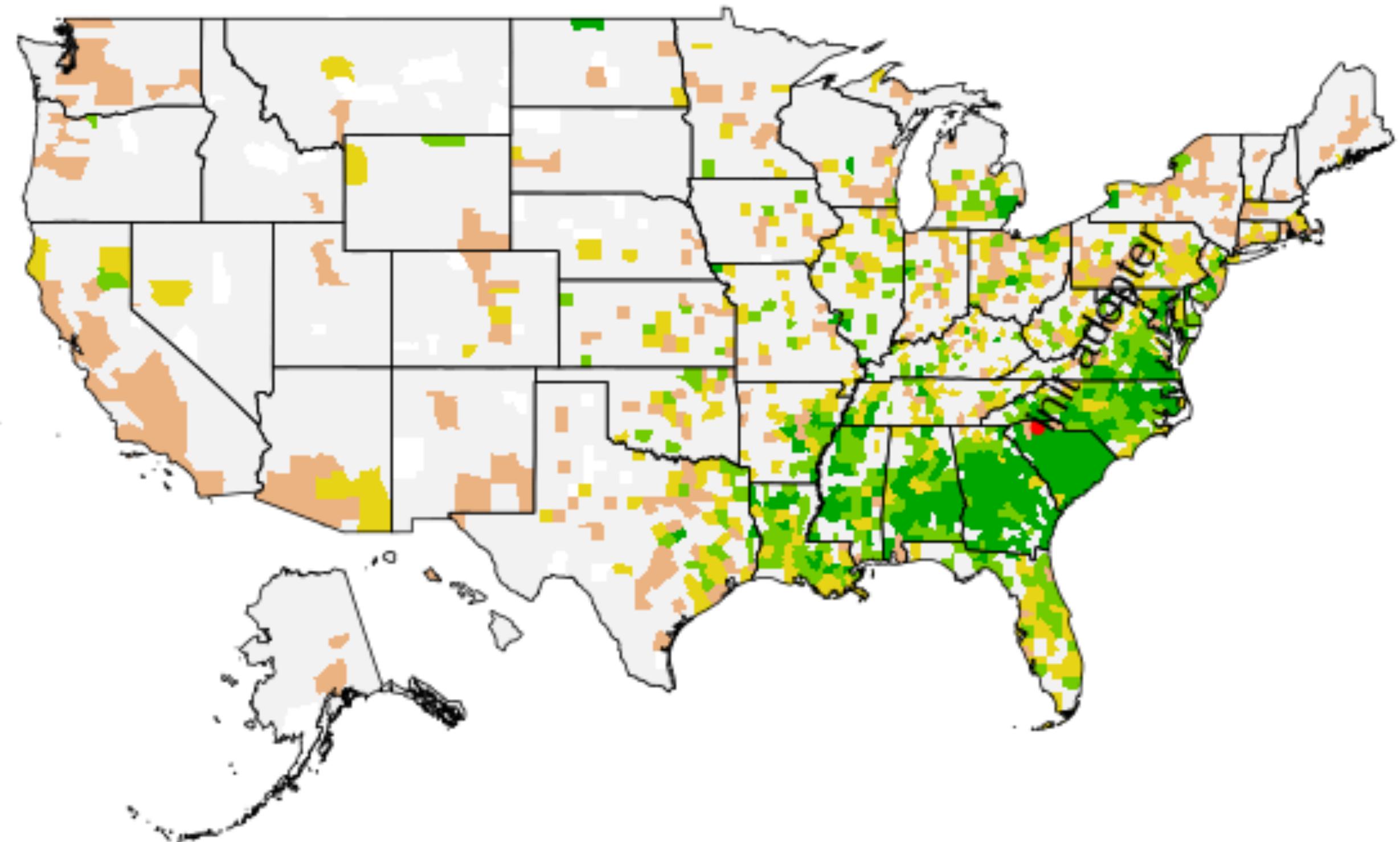
African American English (Source: Jones 2015)



Use of “talm bout” on Twitter

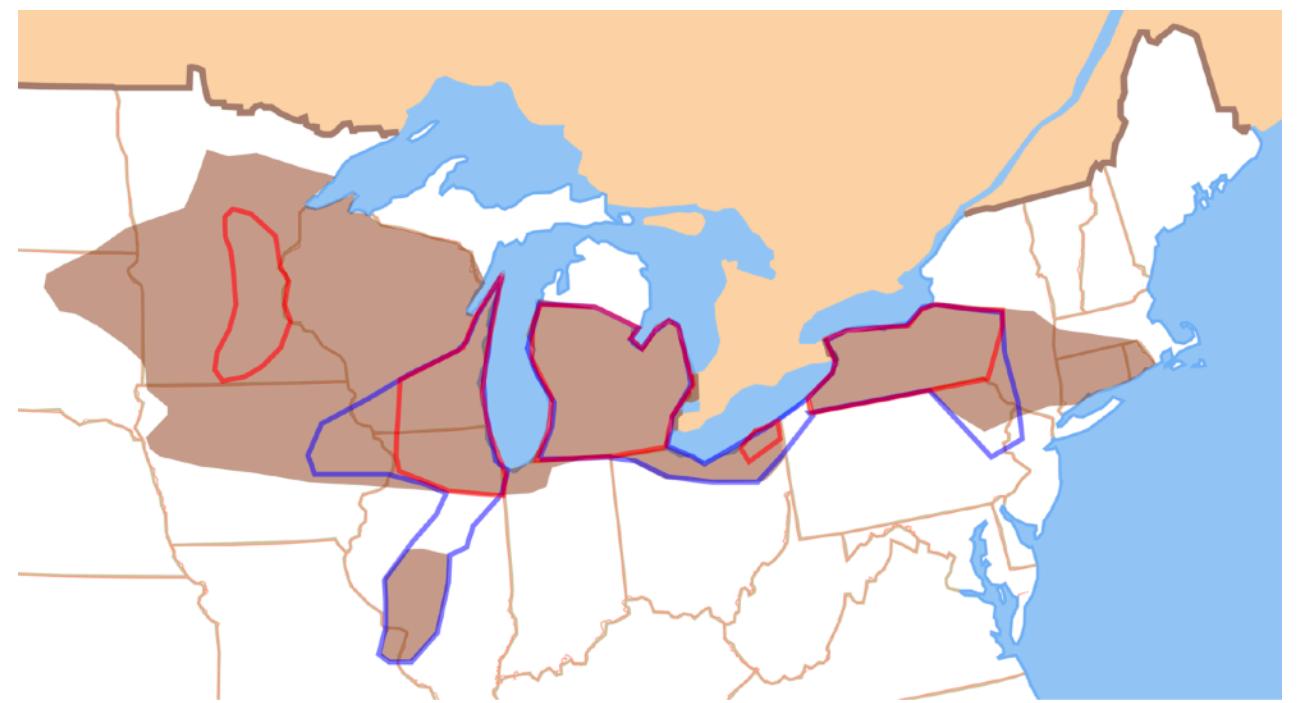


Our Model (identity = race, initial adopter in the South)

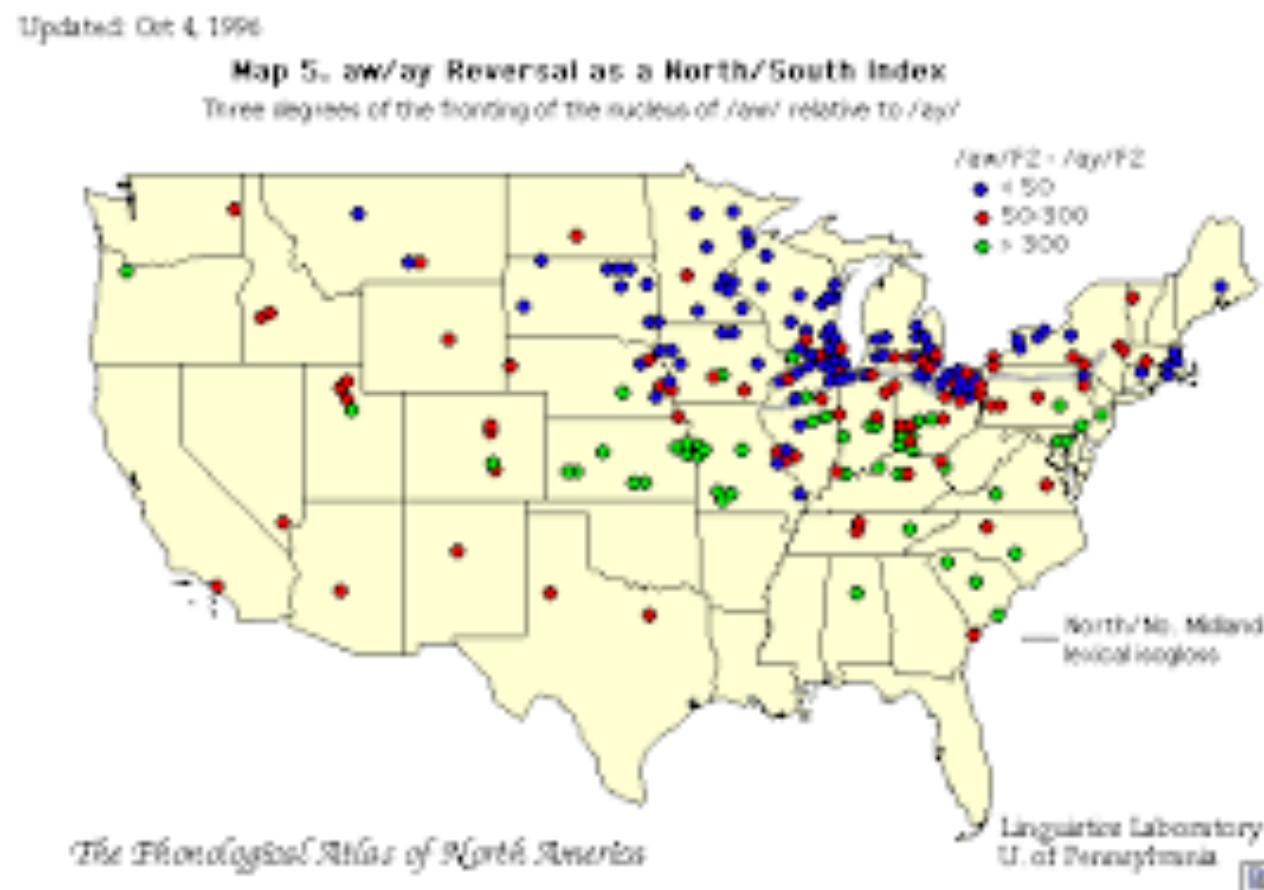


Northern Cities Vowel Shift

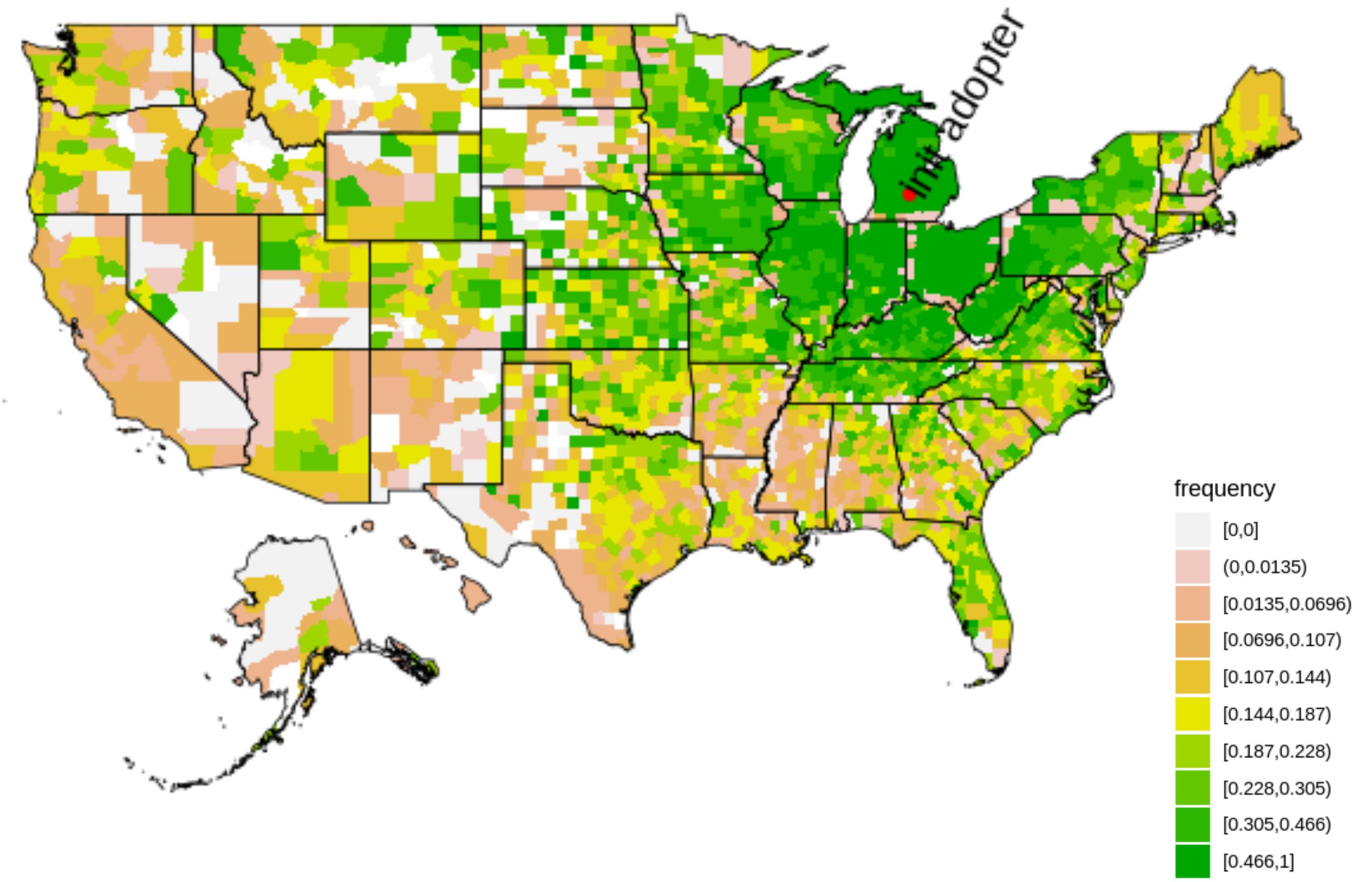
*Isoglosses for Northern Cities Vowel Shift
(Source: Labov, Ash, Boberg 2006)*



Phonological Atlas of North America



Our Model (identity = race, SES, and geography; initial adopter in 'Rust Belt')



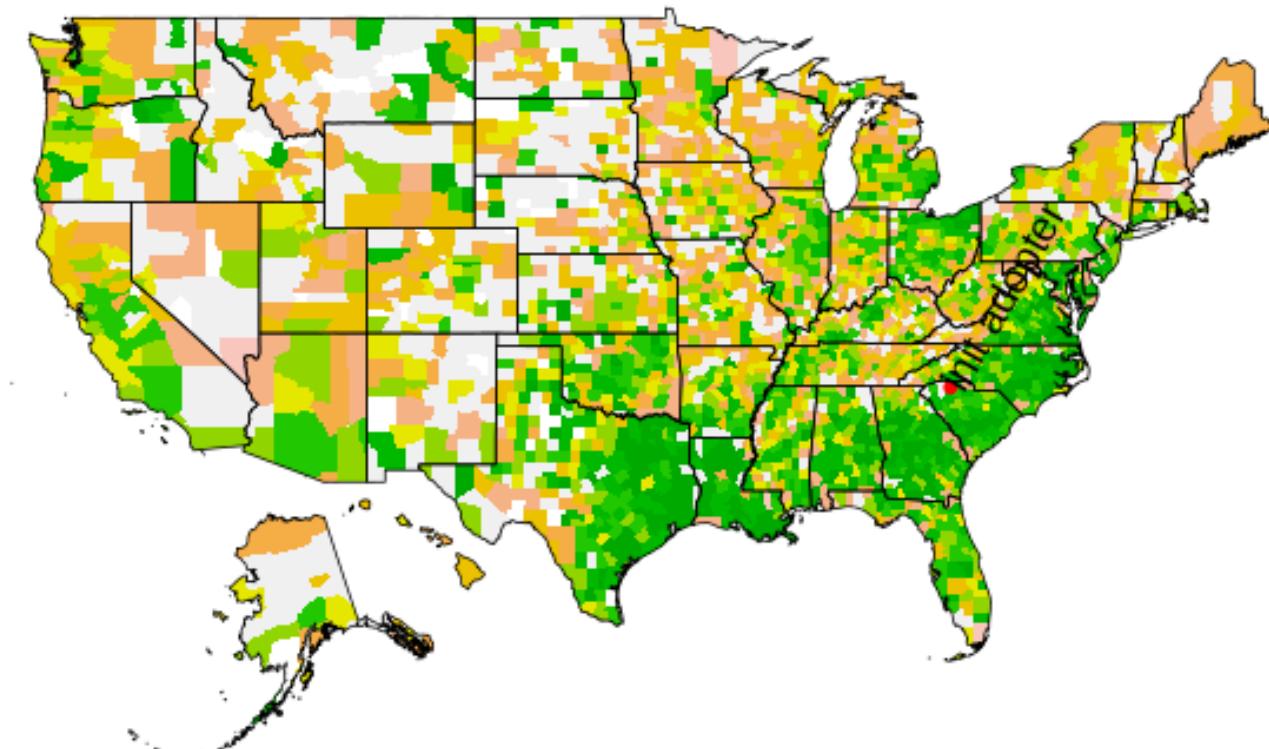
Early Results

- Words in our model localize to important linguistic regions
- These regions are the product of the network topology and speaker sociodemographic identities
- Future work: testing hypotheses about mechanisms underlying diffusion of linguistic variation and innovation

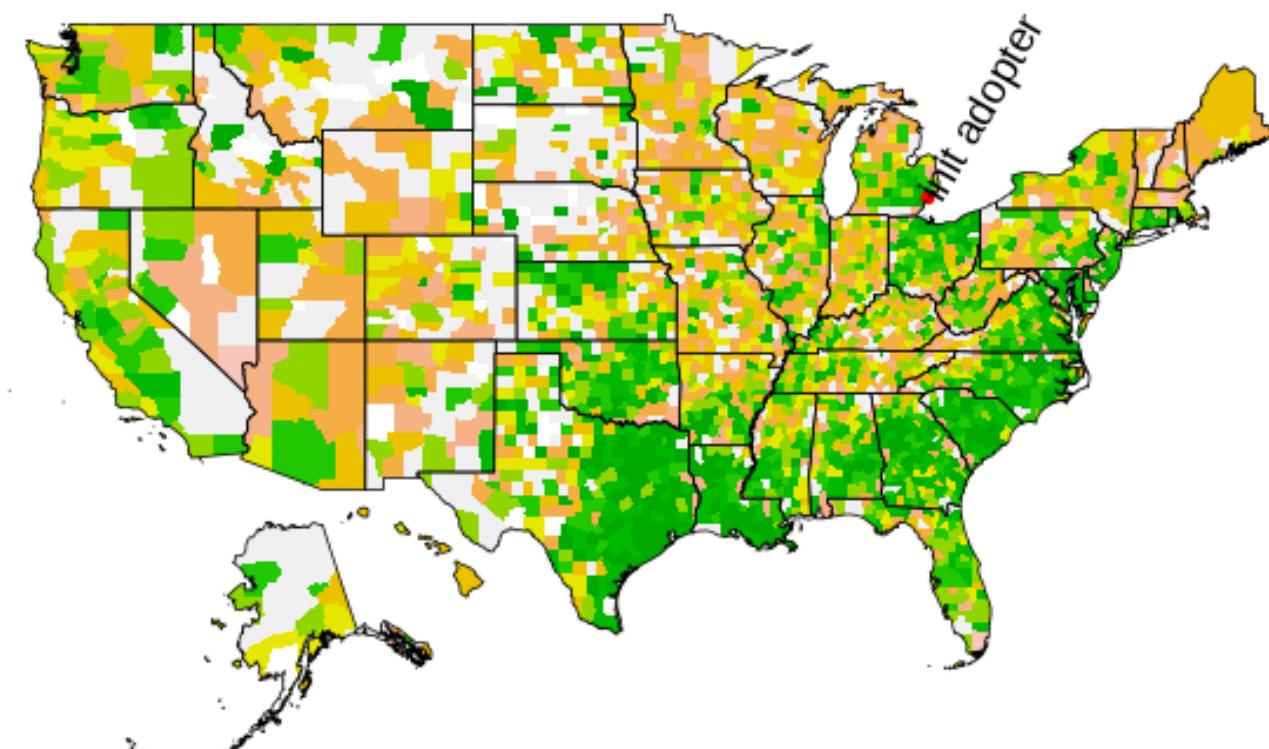
Network and Sociodemographic Identity are Both Necessary for Regions

African American
English

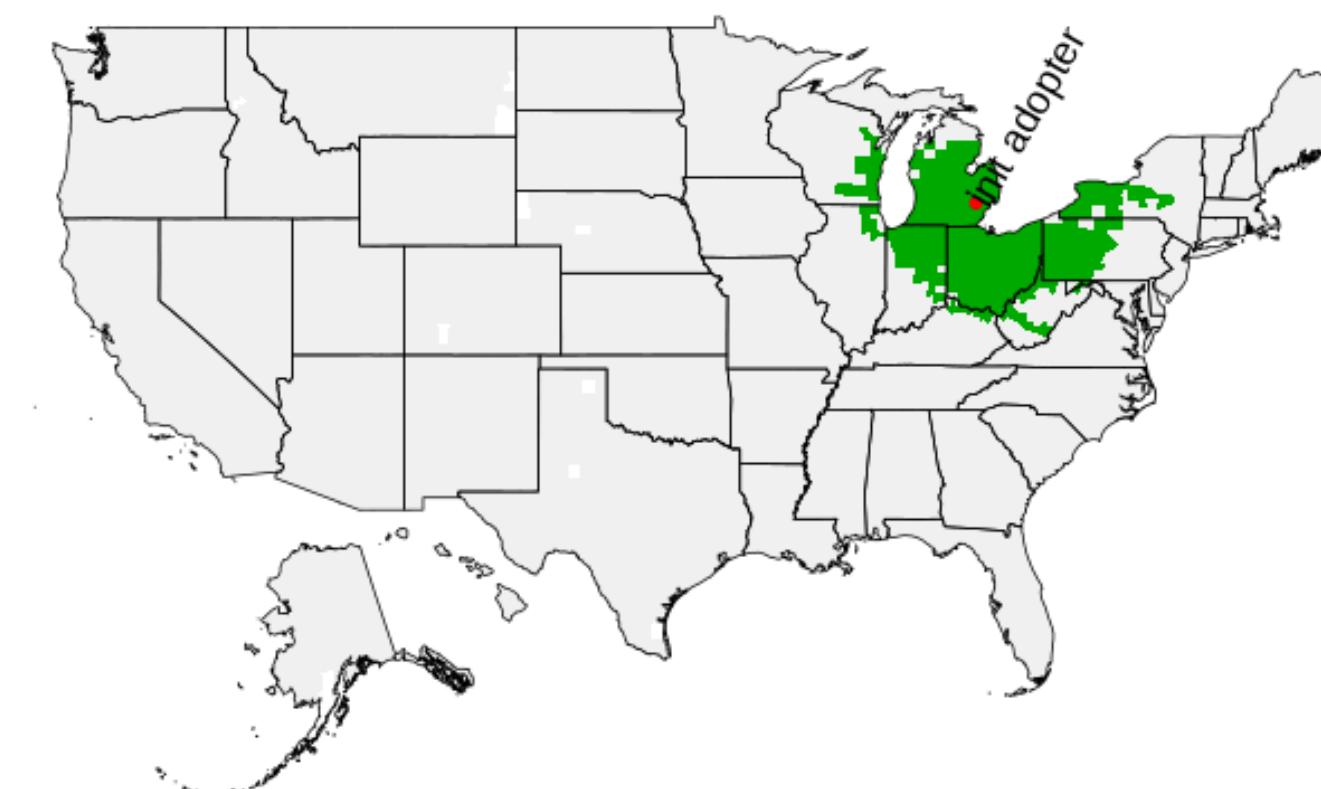
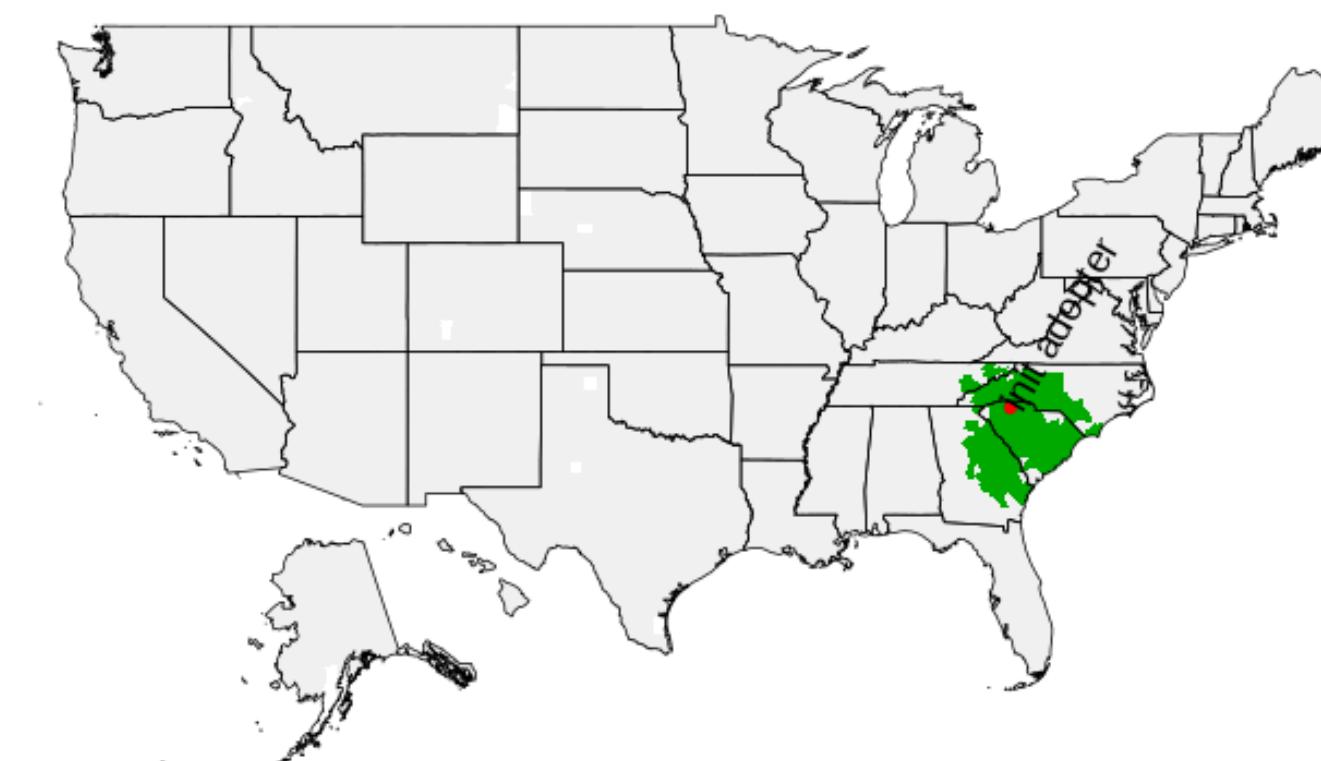
Network Only



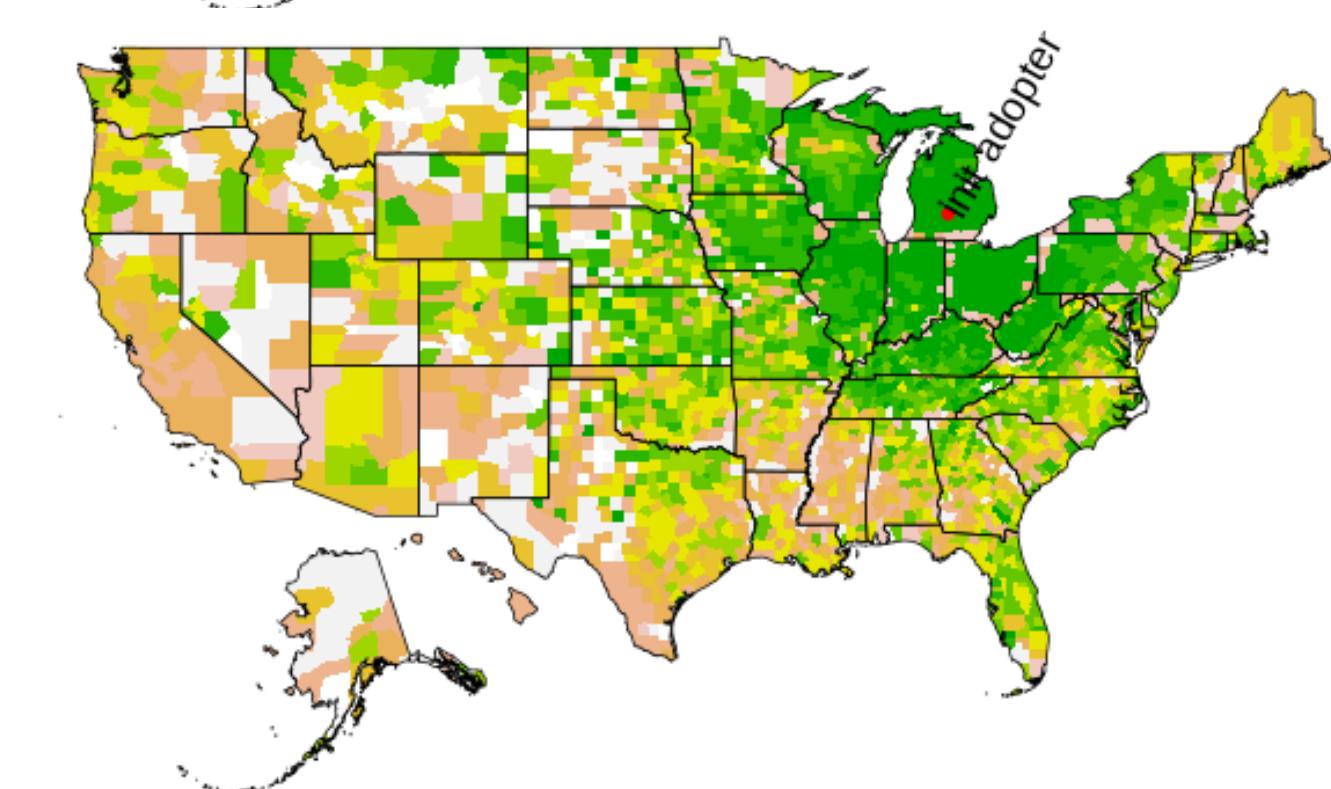
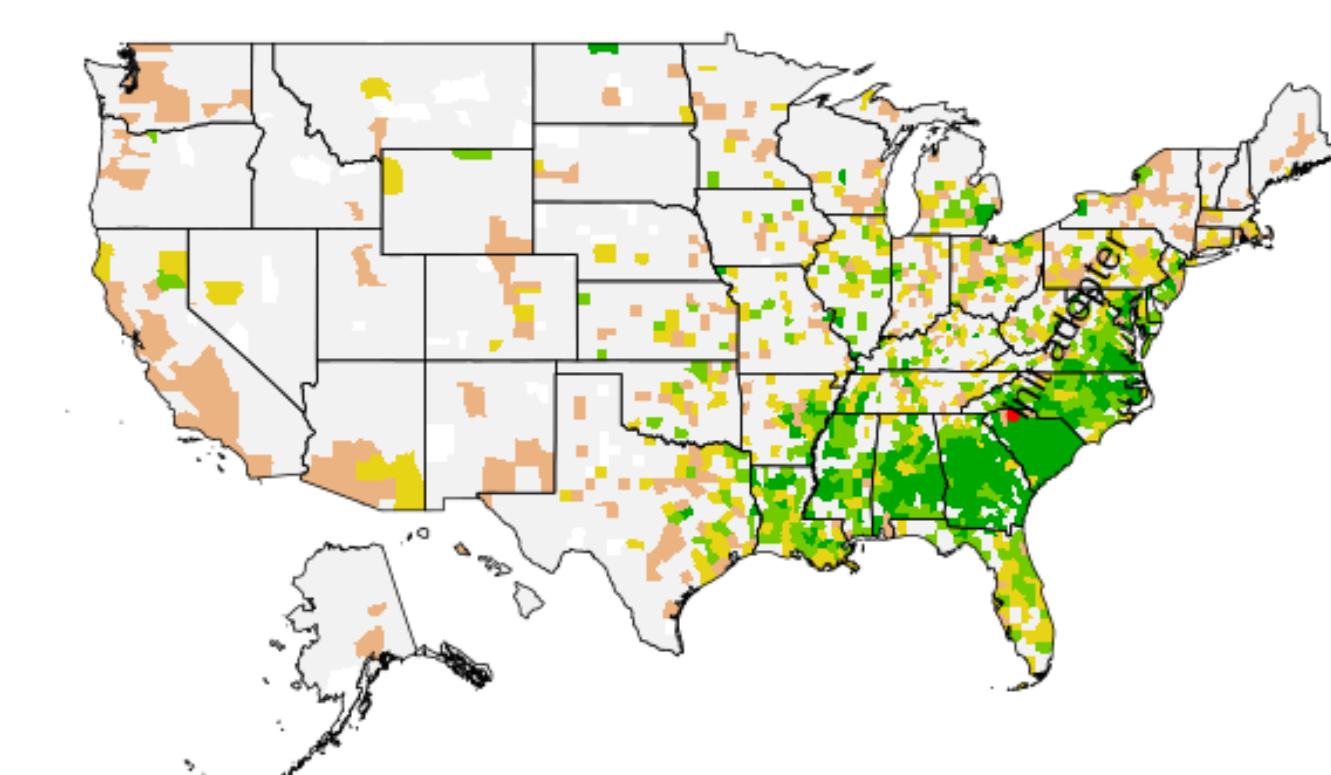
Northern Cities
Vowel Shift



Identity Only



Network + Identity



Localization of language is the product of segregation in both the network and the social context

Early Results

- Words in our model localize to important linguistic regions
- These regions are the product of the network topology and speaker sociodemographic identities
- Future work: testing hypotheses about mechanisms underlying diffusion of linguistic variation and innovation

Planned Work

- Test hypotheses about language change from sociolinguistics literature:
 - “Enregisterment” of new words (Agha 2003)
 - Cascade model of linguistic diffusion (Labov 2003)
 - Emergence of regionalisms (Eisenstein 2010)
- Examine model mechanisms:
 - Role of network vs. sociodemographic identity in localization
 - Emergence of robust linguistic regions

Takeaways

- Our model reproduces several socially significant language regions
- We were able to separate the effects of network topology from social context, finding that each alone is insufficient to describe linguistic diffusion
- We use our model to evaluate mechanisms underlying linguistic localization



Aparna Ananthasubramaniam



aparna-ananth.github.io

@AparnaAnanth729

akananth@umich.edu

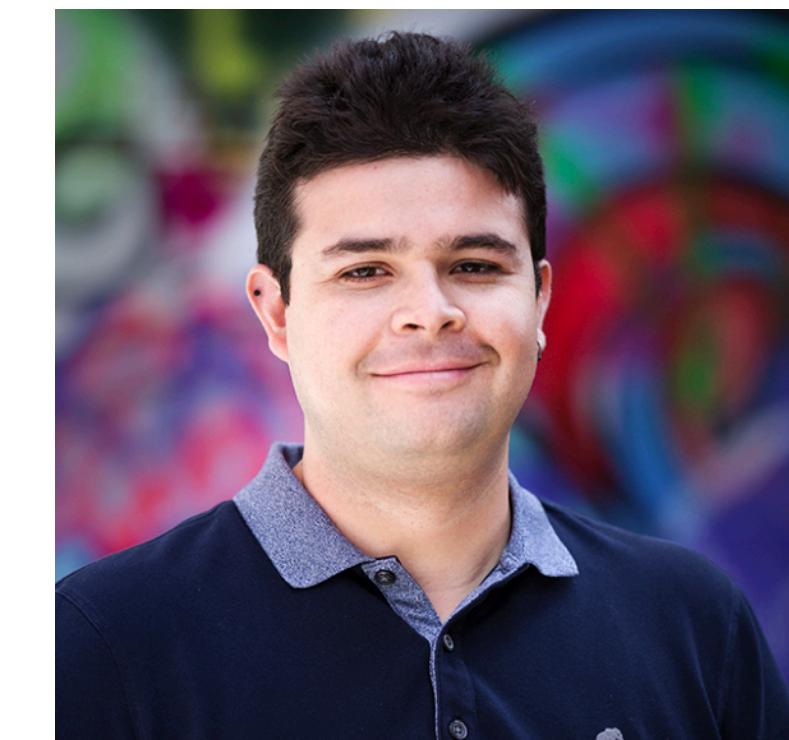


David Jurgens

jurgens.people.si.umich.edu

@David__Jurgens

jurgens@umich.edu



Daniel Romero

dromero.org

@DanielMRomero

drom@umich.edu

Appendix

Why Twitter?

Challenges:

- Biased sample of population
- Inaccurate network topology
- Performative, unrepresentative language used
- Trends may not generalize offline
- Ethical concerns around consent

Benefits:

- Scale: Innovation happens often and diffuses fast
- Measurability: Leaves a written trace, aggregate spatial timeseries
- Inherent Utility: Important source of language change in the present day
- Generalizability (in our context): Spatiotemporal patterns and demographic associations mirror offline results