

# Conceptualizing and coding social class in North American sociolinguistics

Suzanne Evans Wagner

Robin Dodsworth

## What are:

- the most relevant indicators of social status
  - for the widest variety of potential linguistic studies?
- the most efficient ways of gathering those indicators?
- the most generalizable ways of coding those indicators?
  - cross-sectionally
  - longitudinally

We'll review three approaches:

- **etic**
- **emic**
- **linguistic market**

then...

we'll look at **social network**

and finally...

we'll make some recommendations.

# Etic

- macro-level, mostly researcher-imposed categories
- indices of education, occupation, income etc
  - can be combined, weighted
- **example**
  - Labov 2001; Labov, Rosenfelder & Fruehwald 2013

Good approach for consistency and interpretability  
but may ...

- exclude measures of **cultural capital**
- be hard to apply to linguistics-size **datasets**
- still need tailoring to the specific **speech community**
- fail to capture categories that predict **linguistic** behavior

“...the leap from occupation to vernacular is obscure...”

Why *should* having a blue-collar job make someone talk a certain way?

Etic approaches give us the **what** but not the **why**.

# Emic

- micro-level
- categories emerge from participant-observation
- **example**
  - Labov 1972 *Martha's Vineyard*
  - Eckert 1989, 2000 *Jocks and Burnouts*

Good approach for e.g. younger speakers, the unemployed, new economic sectors, other hard to categorize groups.

but ...

- requires long-term or intimate **familiarity** with community
- categories may not be **generalizable**



Is there a way to unite emic and etic approaches?

# The linguistic market

- Bourdieu & Boltanski 1975
- Sankoff & Laberge 1978
- centrality of an occupation to the **linguistic market**
  - How much does the occupation require competence in the standard or elite variety/language?
- **example**
  - *avoir/être* in Montreal (Sankoff & Laberge 1978; Sankoff, Thibault & Wagner 2004)

This approach foregrounds the status conferred by language use.

But...

- relationship between language use and social status may be **circular**
- assessment of linguistic market centrality is **subjective**
- **time-consuming** method
- potential for high degree of **colinearity** with other indicators

# Social networks

- like linguistic market, tries to more directly capture the everyday experiences that shape language use
- network indices, e.g.
  - number of friends in the neighborhood
  - number of friends of a different ethnic group
- **example**
  - bipartite networks in Raleigh

An attractive approach because:

- more directly represents **talk in interaction** than emic/etic categories
- **everyone** has a network
- excellent **quantitative** methods – good for **replicability**
- effects of socioeconomic status may in fact be due (at least in part) to social network

But...

- enormous **quantity** of data needed
- data collection is **lengthy** and prone to **gaps**
- self-report data is **unreliable**
- a bit **mechanistic**: speaker agency/identity is backgrounded
- hard to **code** for sharing, generalizability

So which approach is best for metadata coding?

On balance, some kind of etic approach.

Sociology can inform, but doesn't always suit us.

Traditional indicators, especially occupation and education, are suitable for most (socio)linguistic studies.

No reason to think these indicators **don't** predict linguistic behavior

And it's more practical, methodologically, than the other approaches.



# Recommendations

## If you can record a lot of SES information:

- get multiple indicators, even if potentially colinear
- get most detailed information you can:
  - name or title of job
  - position in chain of authority
  - skill level
  - who the speaker interacts with regularly
- include qualitative notes
  - **example:** IHELP study of Lansing, MI

# Recommendations

If you can't record a lot of SES information:

- choose **occupation**
  - consistently a better predictor than education, income etc individually or combined
  - a window into speaker's everyday interactions
  - captures more about a speaker's identity/persona
  - a single education level = lots of occupations *but*
  - a single occupation usually  $\neq$  lots of education levels

# Recommendations

## When you're imposing categories:

- use existing categories (from linguistic studies!) where possible, for comparability
- impose the largest number of groupings you can, without losing statistical power
  - three is better than two
- but even two groups is helpful
  - **example:** legacy data from Lansing Auto Town corpus

# Recommendations

- If you're really unsure about categorizing, consult a class schema such as Wright, Goldthorpe etc (this presentation, Fabricius presentation
  - useful even if not tailored to your population
  - could promote consistency