Preaching to the Choir: What all linguists need to know about defining and operationalizing ethnicity and race in research

Alicia Beckford Wassink, Robert Squizzero, Martin Horst, Alex Panicacci, Emily M. Bender
University of Washington

NWAV 51
Queens College, CUNY
October 2023

Presentation Roadmap

1. Equipping linguists in all fields with a critical lens for understanding (and using) race and ethnicity
    - Use of terms
    - Recognizing racializing practices
    - Collecting demographic data
2. Who are we?
3. Why this paper?
4. The study of textbook guidance
5. “Defining” race and ethnicity
6. ”How much demographic data do I need?”

Who we are:

Dialectology
Variationist Sociolinguistics
Sociophonetics
Syntax
Interactional Sociolinguistics/French and Italian studies
Computational linguistics
Social Psychology
Phonetics

Language submission title: Commentary “Collecting and using race and ethnicity data in linguistic studies” (Squizzero, Horst, Wassink, Panicacci, Jensen, Moroz, Conrod, and Bender, in prep.)

Why this study?

1. Linguistics has been criticized for undertheorized application of the notions of race and ethnicity (from outside and inside)
    - 2019 LSA Statement on Race

Charity Hudley & Mallinson (2011)
García Sanchez (2014)
Cheshire (2016, pc)
Charity-Hudley (2017)
Lanehart (2022)

2. Improved research ethics & respect for communities we represent

Cameron et al. (1992)
Rice (2010, 2012)
Eckert (2013)

3. Colleagues' requests for recommendations and templates

4. Better alignment with our sister fields (anthropology, archaeology, psychology, sociology)

López et al. (2017)
Fuentes et al. (2019)
García (2020)
Charity Hudley et al. (2020)

“All linguistic research has the potential to reproduce or challenge racial notions.”

- Linguistic Society of America Statement on Race (2019)

Our Study: Survey of Research Design Advice

62 textbooks found and reviewed

Publication years: 1951 and 2020

All subfields queried (including core subfields, applied linguistics, corpus linguistics, language documentation, anthropological linguistics)

New researchers are most likely to find guidance about **conceptualizing (speech/language) community** in critical sociolinguistics, applied sociolinguistics, language documentation, language variation and change.

2 actually "went there"

(Milroy & Gordon 2003; Heller et al. 2018)

New researchers are most likely to find general guidance about **designing demographic prompts** in methods texts about studying language variation and change.

Misconceptions

My study is "just about language." Sociolinguistic issues are not at the heart of the linguistic enterprise, so it's not my concern. (Kiesling 2011, Milroy 1987, Kibrik 1977)

My research is socially "neutral"/uncontaminated by investigation.

I can judge who my subjects should be/how to describe the community.

Race = Ethnicity

We act in the social world and [must] reflect upon ourselves and our actions as objects in that world." (Hammersley & Atkinson 2006)

Race & ethnicity defined

- Race and ethnicity are viewed as overlapping and used interchangeably

- In linguistic research, it is important to distinguish between:

1) Approaches that are static and essentializing (usually race-based)

2) Approaches that are practice-based (usually ethnicity-based)

Race

- *Race* refers to a group sharing physical features, especially skin color, facial features, eye shape, and hair texture (Bobo 2001; Spears 2020)
- Definitions are pointless outside of an acknowledgement of racism (Lanehart, 2023)
- *Race essentialism* is the tendency to view race as biologically based, immutable, and informative (Haslam, Rothschild & Ernst 2000; Prentice & Miller 2007)
    - Race essentialism has been linked to racial stereotyping and prejudice (Levy & Dweck 2003; Williams & Eberhardt 2008)
- Essentialist racial classification schemes based in biology or genetics are:
    - unreliable (Garcia 2020; Relethford 2009)
    - severely flawed (Keita et al. 2004)
    - completely arbitrary (Omi & Winant 2014)

Ethnicity

- *Ethnicity* refers to a grouping based upon:
  - Shared signs (in the semiotic sense),
  - Shared aspects of a common culture, or
  - Shared practice

- Material manifestations of shared aspects of culture may include:
  - Following patterns of dress
  - Adhering to diets or eating particular foods
  - Observing holidays
  - Practicing religions, and crucially
  - **Speaking languages and language varieties**

(Garcia 2020, writing on behalf of the American Anthropological Association and the Society for Anthropology in Community Colleges)

So should I ask about race or ethnicity?

- *Probably ethnicity*. When describing language, useful to adopt framing, practice-based approaches
- less likely to be essentializing
- less exclusionary
  - Sometimes, language users do not possess the phenotypic characteristics stereotypically associated the speech community to which they belong
  - Practice-based approaches describe what people **do**, not what people look like
  - Main exception: if you are investigating racism, you might ask about race

1. Reviewer comments we've received on journal manuscript submissions (including when we've been criticized)
2. Conversations about research ethics (with our community partners, on Twitter, in the Media, online)
3. Two years of laboratory group discussions
4. Reviewer reactions to two earlier versions of this paper
5. Texts mentioned earlier that provided guidance (or examples where these would have been helpful)
6. Wisdom shared by some of you...

How did we come up with our recommendations?

How much demographic information might I need?

Large Corpus-Driven:
Computational Linguistics, Historical Linguistics, Corpus Linguistics

Formal:
Syntax
Semantics
Morphology
Phonology
Typology
Pragmatics

Experimental:
Phonetics
Phonology
L1 & L2 Acquisition
Bilingualism
Psycholinguistics
Neurolinguistics
Sociolinguistics

Qualitative:
Discourse Analysis, Sociolinguistics, Sociocultural Linguistics, Language Documentation, Raciolinguistics

Probably less → **Probably more**

Large Corpus-driven

Typical types and amounts of data:

- Digitized
- Internet-sourced
- Vast corpora (millions of observations), largely anonymous

- Typical style of research question involves **hypothesis testing**
- The need for race and ethnicity data can arise for specific tasks which implicate social identities
  - When working with “unlabeled” social media data (Twitter or Reddit)
  - Linking language use and social identity (Hate speech detection)
  - NLP: building broadly useful tech (e.g. speech to text)

(Abreu 2015)
Bender and Friedman (2018)
Gonen and Goldberg 2019

Large Corpus-driven (cont.)

Recommendations

1. Closed dataset? **Be transparent** when self-identity data are not known
2. Your dataset? let participants opt-in, give informed consent, **self-report demographics**
3. **Avoiding essentializing** linguistic features as THE markers of racialized language varieties
4. **Beware of linguistic appropriation** in datasets
5. Compare large datasets to relevant studies drawn from the same user population

Abreu (2015)
Charity Hudley (2017)

Formal

Typical types and amounts of data:

- Grammatical intuitions (unnamed consultants)
- Small number of consultants
- Data from preexisting studies

- Typical style of research question involves **deductive reasoning and explanation**
- The need for reporting ethnicity data arises because such methods may:
  - Misrepresent the provenance of the phenomenon
  - Misreporting the state of the grammar of interest

Formal

Our Recommendations:

1. Collect minimal demographic information to establish regional and social location (**region, ethnicity, gender identification** and **language background**)
2. Identify your **sources** (judgements and examples)
   How many **speakers?** From what **regions**? Of what **ethnicity?**
3. If your source is yourself, write a brief **positionality statement**
4. Ask your sources what **relevant social categories** would apply if they were to write a brief positionality statement or bio for themselves. Include this information in a footnote or an appendix

Legate et al. (2020)

Experimental:

Typical types and amounts of data:

- Large judgement or random sample (primary data)
- Inferential, time series and descriptive analysis
- Coding for age, gender, language exposure, interlocutor type, treatment, group, etc.

Typical style of research question is **descriptive** or **correlational**:

- Sociolinguistics: Distribution of some linguistic feature, e.g., “Is use of Avertive *liketa* disfavored in constructions displaying multiple-negation in Mississippi AAE?”
- Acquisition: “Does X feature (FL learning) occur differently in group Y (intervention children) than in group Z (CPC)?” (Ferjan-Ramirez & Kuhl, 2020)

Experimental:

Recommendations:

1. understand which ethnic labels might be **relevant** in/to the community of interest. Immediate **social network** useful.
2. allow participants/caregivers to **not answer questions**.
3. allow participants/caregivers to **choose multiple options**.
4. multiple choice with well-justified categories, free-response or interview-style (consider including an option where participants/caregivers can name one or more labels not already included which are relevant to them)
5. Report **analyst's positionality**.

Qualitative

Typical types and amounts of data:

- Qualitative - Ethnographic observation of “lived routines of daily living”
- Period of observation spans years
- Large amounts of data recorded by the analyst

Research Questions vary. Characterization of some linguistic phenomenon within an individual language.

- Community participation
- Lapierre (p.c.): Age, gender, clan, familial and social roles (marriage), village of origin
- Analyst records key facts

Key: *Ethnographic approach*

`Ethnos > Gk. “belonging”`

Recommendations: See previous slide; include practices

Conclusion: What linguistics has to gain by critical treatment of race & ethnicity data

- ➔ Better design considerations
- ➔ Better transparency (e.g., [under]representation of speaker types in NLP modeling)
- ➔ Better representation of under-sampled groups
- ➔ Avoiding social harm to participants
- ➔ Appropriate level of generalization for theory building
- ➔ … and more!

Acknowledgements

Anne Charity Hudley,
Arthur Spears,
Sonja Lanehart,
Laada Bilaniuk,
Maya Angela Smith,
Sara B. Ng,
Laura Munger

References (and reading suggestions)

Abreu, A. M. (2015). Online Imagined Black English. *Arachne*. Available at: https://arachne.cc/issues/01/online-imagined_manuel-arturo-abreu.html

Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, *6*, 587-604.

Bobo, Lawrence. 2001. Racial attitudes and relations at the close of the twentieth century. In *America becoming: Racial Trends and Their Consequences,* vol. 1, ed. Neil J. Smelser, William Julius Wilson, and Faith Mitchell, 264-301. Washington, DC: National Academy Press.

Bucholtz, M. 2020. Race, Research, and Linguistic Activism. *The Routledge Companion to the Work of John R. Rickford.*

Cameron, Deborah, Elizabeth Fraser, Penelope Harvey, M. B. H. Rampton, and Kay Richardson. 1992. *Researching language: issues of power and method*. London, UK/New York, NY: Routledge.

Charity Hudley, A. H. 2017. Language and racialization. *The Oxford Handbook of Language and Society.* Oxford, UK: Oxford Handbooks.

Charity Hudley, A. H., Mallinson, C., & Bucholtz, M. 2020. Toward racial justice in linguistics: Interdisciplinary insights into theorizing race in the discipline and diversifying the profession. *Language* 96(4). 200–235.

Comaroff, J., & Comaroff, J. (2009). *Ethnicity, Inc.* Chicago: The University of Chicago.

Eckert, P. 2012. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual review of Anthropology,* 41, 87-100.

Fasold, R. 2019. Comment on: LSA Statement on Race

Fuentes, Agustín, Rebecca Rogers Ackermann, Sheela Athreya, Deborah Bolnick, Tina Lasisi, Sang-Hee Lee, Shay-Akil McLean, Robin Nelson (2019) AAPA Statement on Race and Racism, *American Journal of Biological Anthropology* 169*(3),* https://doi.org/10.1002/ajpa.23882

García, J. D. 2020. Race and Ethnicity. In N. Brown, T. Mcilwraith, & L. Tubelle De González (Eds.), P*erspectives: An Open Invitation to Cultural Anthropology*, 2nd edn. 444–455. American Anthropological Association. http://perspectives.americananthro.org/

Hall, Stuart. (1980) Teaching Race. In Gilroy, P. and Gilmore, R. eds. (2021), *Selected Writings on Race and Difference: Stuart Hall.* Durham: Duke UP

Hanulíková, Adriana. 2018. The effect of perceived ethnicity on spoken text comprehension under clear and adverse listening conditions. *Linguistics Vanguard* 4(1). 1–9. https://doi.org/10.1515/lingvan-2017-0029

Haslam, N., Rothschild, L., & Ernst, D. 2000. Essentialist beliefs about social categories. *British Journal of Social Psychology*, 39(1), 113–127.

Heng, Geraldine. (2011). The Invention of Race in the European Middle Ages I: Race Studies, Modernity, and the Middle Ages. *Literature Compass*, 8(5), 315-331. https://doi.org/10.1111/j.1741-4113.2011.00790.x,

Kang, Okim, Donald L. Rubin & Stephanie Lindemann. 2015. Mitigating U.S. Undergraduates' Attitudes Toward International Teaching Assistants. *TESOL Quarterly* 49(4). 681–706. https://doi.org/10.1002/tesq.192

References (and reading suggestions) cont.

Keita, S. O. Y., Kittles, R. A., Royal, C. D., Bonney, G. E., Furbert-Harris, P., Dunston, G. M., & Rotimi, C. N. (2004). Conceptualizing human variation. *Nature genetics*, *36*(Suppl 11), S17-S20.

Lanehart, S. 2023. *Language in African American Communities.* (Routledge guides to Linguistics). London: Routledge.

Levy, S. R., & Dweck, C. S. (1999). The impact of children's static versus dynamic conceptions of people on stereotype formation. *Child Development*, 70(5), 1163-1180.

Linguistic Society of America. 2019. Statement on Race. https://www.linguisticsociety.org/content/lsa-statement-rac

López, N., Vargas, D. E., Juarez, M. Cacari-Stone, L., & Bettez, S. 2017. What's Your "Street Race"? Leveraging Multidimensional Measures of Race and Intersectionality for Examining Physical and Mental Health Status among Latinxs, *Sociology of Race and Ethnicity*. 4(1):49-66.

Manzini, T., Lim, Y. C., Tsvetkov, Y., & Black, A. W. (2019). Black is to criminal as caucasian is to police: Detecting and removing multiclass bias in word embeddings. *arXiv preprint arXiv:1904.04047*.

Omi, Michael & Howard Winant. 2014. *Racial Formation in the United States* 3rd ed. Routledge.

Pauker, K., Apfelbaum, E. P., & Spitzer, B. 2015. When societal norms and social identity collide: The race talk dilemma for racial minority children. *Social psychological and personality science*, 6(8), 887-895.

Pauker K., Meyers C. K., Sanchez D. T., Gaither SE, Young DM. 2018. A review of multiracial malleability: Identity, categorization, and shifting racial attitudes. *Social and personality psychology compass.*

Prentice, D. A., & Miller, D. T. 2007. Psychological essentialism of human categories. *Current Directions in Psychological Science*, 16(4), 202-206.

Relethford, J. H. (2009). Race and global patterns of phenotypic variation. *American journal of physical anthropology*, *139*(1), 16-22.

Rice, K. (2010) Grenoble, Lenore A.; and N. Louanna Furbee, eds.. 2010. Language documentation: Practice and values. John Benjamins Publishing.

Smith, Maya A. 2019. *Senegal Abroad: linguistic borders, racial formations, and diasporic imaginaries.* Madison: University of Wisconsin Press.

Spears, A. 2020. Racism, Colorism, and Language within their macro contexts. In *The Oxford Handbook of Language and Race* (H. S. Alim, A. Reyes, & P. Kroskrity, Eds.). New York: Oxford.

Squizzero, Robert. 2020. Attitudes toward L2 Mandarin Speakers of Chinese and non-Chinese Ethnicity. In Kaidi Chen (ed.), *Proceedings of the 32nd Meeting of the North American Conference on Chinese Linguistics*, 521–538. Storrs, CT.

Squizzero et al. (2021) "Collecting and using race and ethnicity information in linguistic studies" *U Washington Working Papers in Linguistics.*

Williams, B. F. 1989. A Class act: anthropology and the race to nation across ethnic terrain. *Annual Review of Anthropology* 18, 401-444.

Williams, M.J., & Eberhardt J.L. 2008. Biological conceptions of race and the motivation to cross racial boundaries. *Journal of Personality and Social Psychology,* 94, 1033–1047. doi: 10.1037/0022-3514.94.6.1033