

TOPIC: Demographic coding for sociolinguistic archival preparation

SUBTOPIC: Coding for Asian (American) ethnic identities

RATIONALE behind more rigorous coding for Asian (American) ethnicities in corpora:

- to more accurately reflect the social reality of speakers' identities
- to make U.S. English speech corpora more readily comparable across field sites
- to further incorporate those ethnic groups that have traditionally been overlooked in large-scale dialectological work to see, among numerous issues:
 - whether those groups vary in their use of regional features and why;
 - if there are linguistic features specific to a given ethnic community (substrate-influenced or not)

THE CHALLENGE: Coding for ethnicity requires going well beyond census categories

- The primary challenge in constructing metadata for a data archive is the observation that ethnicity, like many other aspects of identity, is continually negotiated and reproduced in discourse, and therefore fluid and changeable, at least to a certain extent.
 - But if we are to remain dedicated to an empirical approach to analyzing ethnicity and sociolinguistic variation, and to the creation and preservation of corpora, then it is necessary that we find a way to cope with the malleable nature of ethnic identity.
- Issues of classification specific to 'Asianness' (some of which are actually of relevance for all ethnic coding decisions)
 - What level of specificity? (e.g., is a speaker better represented as 'Chinese' vs. 'Cantonese'?) There are many different levels of intra-group differentiations among people who are often considered by "outsiders" to be members of a single group.
 - In Wong's field site (NYC), for instance, three of the main 'Chinese' groups are Cantonese, Fujianese, and Taiwanese. And Cantonese can be further divided between the earlier immigrants from rural Taishan and the more affluent urban immigrants from Hong Kong and Guangzhou. These different 'Chinese' groups are distinguished not only by different heritage languages, but also by different social statuses as well as stylistic capital.
 - Immigrant generation is also another common source of intra-ethnic differentiation, particularly between the immigrant generation vs. the American-born generations.
 - What level of generality? When is the category 'Asian American' appropriate? How community members conceive of inter-group differentiations varies between communities, individuals, or even within individuals in the same moment of discourse.
 - In Hall-Lew's field site (SF), for instance, the seemingly pan-ethnic term 'Asian' is the most common identification term. But 'Asian' is typically understood to refer to East Asians, specifically Chinese, Japanese, and Korean. This is evidenced by discursive constructions such as "Whites, Asians, and Filipinos." Some speakers use 'Asian' to mean 'Chinese' only.
 - The general issue of people who identify ethnically as Mixed:
 - Mixed-race individuals are statistically more prevalent among Americans of Asian heritage than other ethnic groups. Some

identify as ‘Hapa’. This group may include, depending on the individual:

- having part Asian and part non-Asian ancestry
 - having 2+ different Asian backgrounds
 - identifying as one ethnicity that was a minority in their heritage country (e.g., ethnic Chinese in Vietnam)
 - trans-racial adoptees (e.g., born in China, raised by white parents in the U.S.)
- In Hall-Lew’s corpus there are several Hapas. These speakers are more likely than non-Hapas to shift their orientation to different ethnicities within a given speech event. Hapas are not outliers, either; there is a higher percentage of multiethnic individuals in this neighborhood than non-mixed Japanese, Korean, Asian Indian, Vietnamese, or African Americans.

DISCUSSION: what we did, what we are working to do, what we would like to see:

- One way to get at ethnic identity/orientation is to use questionnaires to gather information based on some quantifiable scale. That information can later be used to look at both inter-ethnic differentiation (Walker & Hoffman 2010) as well as intra-ethnic differentiation (Wong 2007). [Give example of Wong 2007, and Walker & Hoffman (2010)’s EO score & social network]
- Potential problems with survey-type questionnaires:
 - they usually compel survey takers to select rigid ethnic/cultural identifications out of (spatial/temporal) contexts. The complexity and fluidity of identities may get washed out.
 - they encourage the use of essentializing, if not misleading, ethnic/cultural categories (“Chinese”, “American”, “white”), whether provided by the survey-designers or by the survey-takers. Again, other salient identities may get washed out.
- Another way to get at ethnic identity/orientation is to collect information about identity through ethnographic interviews (and observations if possible).
- Researchers can then consider **self-identification** as it appears in the discourse of the very data being archived. This includes bald statements of ethnic self-identification as well as more context-dependent statements about affiliation with ethnic groups, past and current, and participation in ethnic practice.
 - Close analysis of the statements in our data that led us to make decisions about (a) assigning relevant ethnic labels and (b) coding for other aspects of ethnicities (e.g. heritage language, ethnicity of networks, etc).
- For legacy corpora to be truly socially smart, each speaker must be tagged with multiple codes for ethnicity. One codes may then be labeled as more representative than the others, but only if the researcher has grounds to do so.
 - Adopting multiple codes of ethnicity allows for both macro-social categories (e.g. Asian American, Filipino) as well as fine-grained distinctions (e.g. ethnic orientation, ethnicity of social network).
 - Note other very closely related labels that we are also (we hope) coding for: Heritage Language (+ Competence & Frequency of Use); Generation in the US; etc.
 - Standardized sets of survey questions may be called for to ensure that researchers are collecting similar metadata (e.g. practice vs attitudes)

Selected References:

- Espiritu, Yen Le. 1992. *Asian American panethnicity: bridging institutions and identities*. Philadelphia, PA: Temple University Press.
- Hall-Lew, Lauren. 2009. *Ethnicity and Phonetic Variation in a San Francisco Neighborhood*. PhD Dissertation, Stanford University, Stanford, CA.
- Hoffman, Michol and James Walker. 2010. Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change* 22: 37-67.
- Reyes, Angela and Adrienne Lo. 2009. *Beyond Yellow English*. Oxford: Oxford University Press.
- Tsai, Jeanne .L., Yu-wen Ying and Peter A. Lee. 2000. The meaning of "being Chinese" and "being American": Variation among Chinese American young adults. *Journal of Cross-Cultural Psychology*. 31: 302-322. (see also <http://www-psych.stanford.edu/~tsailab/GEQ.htm>)
- Wong, Amy Wing-mei. 2007. Two vernacular features in the English of four American-born Chinese. *University of Pennsylvania Working Papers in Linguistics: Selected Papers from NWAV 35*, 13(2):217–230.