

Snapshots in time:

Coding social factors in changing communities

Devyani Sharma and Nathan Young

1 Introduction

How should researchers code and document social factors about their communities and participants when those factors are themselves in flux? In this chapter, we examine communities that are undergoing major social change and ways to operationalize social factors and document metadata to maximize transparency and replicability (Yaeger-Dror & Cieri, 2014). We also note along the way the importance of recognizing situations where this may not be achievable.

The social factors discussed here overlap with those discussed in many other chapters in the present volume. Our focus here is the challenge of handling such factors amidst major change. This includes identifying which factors are relevant for analysis, conceptualizing those factors, eliciting or obtaining the relevant data, coding and analyzing it, and documenting relevant background. We review a range of independent variables—age, generation, social class, social network, bilingualism, and cultural practices—and comment only briefly on dependent variables. For reasons of space, we limit our main discussion to late-modern London and Stockholm but make reference to other relevant studies.

We advocate triangulating complementary types of data to mitigate blind spots, “approach[ing] a single problem with different methods, with complementary sources of error” (Labov, 1972a, p. 118). We therefore include sociological and ethnographic perspectives alongside our main variationist focus.

2 The social context

Before entering a community (Cieri, this volume, Di Persio, this volume, D’Arcy, this volume) and while conducting fieldwork, researchers familiarize themselves with the history, media, and general social tenor of the community at hand. Most communities are undergoing some degree of social change, but not all will be undergoing rapid and transformative change. Higher-level awareness of the sociohistorical context is necessary for establishing whether one is witnessing gradual or transformative change.

2.1 Migration and change in Western cities

We examine linguistic change amid substantial inter-generational and transformative social change, focusing on late modern Europe. Late modernity, an often under-defined term, is characterized by Wacquant (2008) as “post-Fordist”, in which both manufacturing and the welfare state have weakened. Late modernity has been tied to linguistic change through postwar low-income migration to European cities and resulting patterns of segregation and polarization. The most striking of these outcomes has been the birth of new European multiethnolects (working class varieties used by multiple ethnicities—particularly, but not limited to, non-white groups) whose emergence coincides directly with this epochal shift. European multiethnolects have been analyzed using general language contact models (e.g. Cheshire et al., 2011), but also, more recently, in terms of the intersection of class, ethnicity, and urban geography. Though rarely described in these terms, these varieties are examples of what has been described as “catastrophic” (abrupt) change (Labov, 1994, pp. 42–44; Lightfoot, 1991, 1997; Poplack & Malvar, 2007). In the present chapter, we reflect on instances of sharp thresholds of change and how best to track and document these.

Migration and change can affect language in very different ways globally (Smakman & Heinrich, 2015; Stanford, 2016). Urban environments in OECD (Organisation for Economic

Co-operation and Development, sometimes termed developed, post-industrial, or Global North) nations often bear more similarities to the above European cases (e.g. migration-linked effects in Sydney English, Grama et al., 2020) than change in urban contexts in culturally and socio-economically different locales (e.g., Abd-El-Jawad, 1987; Satyanath, 2015). We also see linguistic change patterning differently in non-Western or more rural contexts, or in situations with very different sociohistorical conditions, e.g. in Australian Aboriginal language situations or rural China (Chirkova et al., 2018; Meakins, 2008; both discussed later).

In this chapter, our primary examples come from two urban Western communities where the authors have conducted research. Sharma's research in the Punjabi community in West London examined first-generation migrants and two age groups of second-generation British Asians (Sharma, 2011; Sharma & Sankaran, 2011). Young's (2019) research examined a comprehensive sample of social class and ethnicity across the social spectrum in Stockholm.

2.2 Understanding sociohistorical conditions in a given community

Large-scale sociohistorical conditions may include changes in demographics, ethnic relations, social recognition or marginalization of groups, and institutional policies and practices. Although such information cannot always be directly encoded as a variable, it is crucial to document it for interpreting and comparing variation. It is often presented in published pieces but separated from the data itself. As the research community moves towards more shared corpora and data, we strongly recommend that researchers include metadata on sociohistorical background. Without this, a researcher who uses the corpus decades later may misinterpret the status or social motivations of speakers, and may, for example, erroneously treat their data as more comparable to an independent dataset than it is.

What sorts of information would give the researchers themselves as well as later scholars a reasonable understanding of current and past sociohistorical context? Some of the best information about a community—a subset of which could be summarized in metadata—comes

from sociological statistics wholly independent of the community. Gal's (1978) use of historical marriage records in Oberwart, Austria, to show that Hungarian-speaking peasant women (more than men) were marrying out into the German-speaking community, stands as a brilliant early example of furnishing a simple sociological statistic to strengthen her interpretation of a gendered and socioeconomic dimension to language shift. We include such examples not necessarily to suggest that future studies mimic them, but to give examples of the creativity that can be brought to bear on the question of how to understand and document the dynamics of a community. Researchers often gather information of this kind—relating to class, ethnicity, occupation, schooling, or cultural practices—directly from participants, but the use of independent community data can dramatically increase the validity of such reported information and the accuracy of any profile of a community.

With such information in hand, an analyst can also step back and explore whether trends in the speech data correspond to particularly discrete breaks (as opposed to incremental trends) in the public record data. This is particularly necessary when data span a large generational or age range. The suggestion is to exploratively harvest publicly available data on factors that we know might track 'sharp' thresholds between social or political eras, for example school markets and segregation, the emergence of racialized working-class enclaves, income inequality, and so on.

Young's (2019) research examined the intersection of race and social class in Stockholm and their manifestations in changing speech rhythm. Certainly, segregation by class and, to some extent, ethnicity is not new to Stockholm. But over the last two decades, migration, income polarization, and segregation have shot up simultaneously and abruptly. Between 1978 and 2005, Stockholm saw an average of 8,600 non-Western migrants per year with very little variance. This shifted abruptly to an average of 15,500 per year between 2006 and 2018 (p. 13). Around this same time, in 2000, Sweden witnessed the highest level of

income polarization since 1936 (p. 260). School segregation shows an even starker shift. After a series of school reforms, Swedish education has rapidly moved from one of the most egalitarian institutions in the world to one driven by a publicly funded voucher system, free school choice and the right to run schools as commercial enterprises (Forsberg, 2018, p. 1). Holmlund et al. (2014) found that between 1988 and 2010, a measure of diversity (“variance decomposition”) among Swedish schools for students with foreign-born parents rose from 0.09 to 0.23. As a comparison, public schools in the Southeastern United States in the 1990s maintained a multiracial variance between 0.25 and 0.29, (Stroub & Richards, 2013, p. 514). Growing income stratification and non-Western migration have worked in tandem with decades-long school segregation to racialize the social-class hierarchy, something that is now said to be a signature feature of the European strain of late modernity (Lentin, 2008; Neergaard, 2017). In many European cities, a racialized working-class subgroup rapidly developed its own linguistic variety, as a result of such patterns of school segregation, social exclusion, and relegated (sub)urban enclosure.

In Stockholm Swedish, speech rhythm stratifies in ways that parallel these sociological developments. Stockholm has traditionally had a tidy set of iconic sociolinguistic variables that moved from backed articulations (working-class) to front (upper-class) as one climbed the socioeconomic hierarchy. For example, [o:], [ɑ:], and [a:] are the traditional working-, middle-, and upper-class variants for /ɑ:/ in LAT (‘lazy’), respectively. Other similar “variant clines” are [u:]–[o:]–[ɔ:] for /o:/ in LÅS (‘lock’) and [x]–[fj]–[ʂ] for /fj/ in SJU (‘seven’). But the racialization of its class hierarchy has disrupted this system, and speech rhythm – itself a salient sociolinguistic variable – no longer stratifies in such a top-to-bottom fashion. Rather, the white “Swedish” working class has some of the highest rhythmic alternation in the city while the non-white working class – native born but racially designated as “Immigrant” – has the lowest rhythmic alternation in the city, popularly characterized as “staccato”. Within an intermediate

range lies the speech rhythm of the middle and upper-middle classes. This working-class-internal ethnic opposition emerged in the years that followed the aforementioned changes in school segregation (Young, 2019, pp. 254–264).

Many studies of populations involving in- or out-migration have pointed to these sorts of transformative change points in social or generational history—early examples include Dubois and Horvath (1998) and Kerswill (1994). This can, though need not, pose a challenge for variationist analysis. Factors should be nuanced enough so as not to mask or miss pivotal junctures of change. Taking a very high-level view, it may even be possible to argue that many *first wave* (Eckert, 2012) studies in sociolinguistics were rooted more in the **modern** epoch, and that patterns of migration in the **late-modern** epoch, particularly in Europe, have created markedly different conditions of contact, input (e.g. group second language acquisition in London; Cheshire et al., 2011), and generational change. This is not brand new—The Industrial Revolution similarly actuated intense migration and contact, and new working-class koinés rapidly emerged (Kotsinas, 1988: 144; Kerswill, 2018). So societies can either be in situations of stability or of epochal change, and a sense of this ‘top level’ context is necessary to contextualize both published work and shared corpus data.

3 Identifying and coding key social factors

A natural starting point for identification and coding of factors is the balance of etic and emic perspectives (Pike, 1967). Etic factors are those treated as comparable across groups and so imposed by the researcher with little adaptation to specific communities. The still common tendency to code sex as male/female (or gender as man/woman) with little adaptation across studies is one example (Eckert, this volume), as are simple classifications of social class or race (Fix et al., this volume, Mesthrie, this volume). By contrast, emic factors are devised with sensitivity to the cultural context, often using salient distinctions, concepts, beliefs, or priorities stemming from the community. It is common to fine-tune etic factors based on emic or

ethnographic work: for example, while social network analysis is founded on very etic theorization of such constructs as the strength and distribution of social ties, the qualities that an analyst tracks within those constructs are almost always the result of emic understanding, such as ethnicity, religious practice, political beliefs, gang membership, and so on.

Factors coded for variationist data will always combine both types. Common etic factors are likely to be coded in any study—class, network, age, gender, ethnicity—but emic knowledge may inform exactly how these are implemented, and further community-specific factors may also be included. Whether superficially etic or emic, coding protocols must document each factor carefully in a shared corpus to avoid misinterpreted factors and erroneous comparisons to other data.

3.1 Age and generation

Age and generation are etic in the sense that they involve universal chronological properties. However, their subdivision into groups must be emically-grounded (Eckert, 1998) and more carefully handled in communities undergoing change. In particular, speaker age at the time of recording, speaker date of birth, and the date of recording may each correspond to distinct linguistic profiles (D’Arcy & Tagliamonte, 2018; Labov et al., 2013). Fruehwald (2017) describes time of interview in terms of ‘zeitgeist’, and D’Arcy and Tagliamonte (2018) recommend probing the data “in as many ways as are available leading to new insight into age vectors, social meaning, geographic differences and lifespan shifts” as well as exploring “multiple tests of time and geography, weighed by nature of data, type of linguistic variable, the social, economic and cultural circumstances of external situation”.

Generational experience can differ substantially in a short space of time and sometimes even come to be enregistered, or recognized as distinct—well-known recent American examples include baby boomers, millennials, and Gen X. Hall-Lew’s (2009) study of younger and older

Chinese San Franciscans documents with historical and ethnographic detail the very different experiences two successive generations have had.

Similar generational shifts were observed in Sharma's project in West London. The original focus of the project was the boundary between Gen 1 (adult migrants) and Gen 2, but fieldwork revealed a stark difference between older and younger Gen 2 British Asians that revealed a subtle one-generation lag in social organization. In the older Gen 2 group, rural Punjabi-style gender roles were maintained, such that women's social networks were smaller and more in-group, leading to their having correspondingly less diverse speech repertoires. By contrast, the younger Gen 2 group had a very different lived experience and developed gender roles that more closely resembled lower middle class British gendered networks (Milroy, 1987), leading to women rather than men having more diverse social networks and speech repertoires (Sharma, 2011). The emic recognition of this 'tipping point' led the researchers to avoid grouping all of Gen 2 together, allowing them to uncover markedly distinct generational linguistic and social factor effects (Sharma & Sankaran, 2011).

Hua et al. (2021) show with novel statistical methods that generation can be clearly independent of age in changing communities. They examined 185 variables across 3 generations of the Gurindji community in Northern Australia—a community undergoing language shift to the mixed language of Gurindji Kriol—and showed that some variables correlate with age in an incremental language shift pattern. Others co-vary with generation rather than age. They concluded that these three generations have a social reality in the historical events which led to the establishment of the Gurindji communities. Generation 1 established a new community, and so Gurindji identity was very salient in their land rights and labor union movements (Meakins, 2008). Generation 2 was the first to grow up in the new community and 'created' Gurindji Kriol; they were also the first with equitable access to English-based schooling. Generation 3 are schoolchildren and were the group to fully elaborate

a new Gurindji Kriol grammar. In qualitative interview content, each generation makes reference to stereotypes of the other generations' speech, tied to perceived differences in lived experience and social context. Thus, even if generation is not included as a factor, ethnographic knowledge of generational stereotypes or transformations should be documented.

In the case of Stockholm, changes in speech rhythm similarly did not emerge gradually. Young (2019, p. 264) refers to these varieties in discrete terms, as Rinkeby Swedish 1.0 and Rinkeby Swedish 2.0. As described above, non-white working-class speakers had lower intervocalic alternation in their prosody than speakers from other groups. However, speakers born before 1987 have much higher and more mainstream alternation than speakers born after 1987. Importantly, these two cohorts also achieve the staccato effect via strikingly different phonetic means, which adds evidence to the interpretation of a discrete or sharp shift. When participants' school attendance was examined in relation to annual data on the ethnic and socioeconomic makeup of Stockholm schools, a corresponding discrete break was found. All speakers' schools were relatively diverse until 2001, after which the variance decomposition of the speakers' schools doubled (p. 259). This meant that younger speakers born after 1987 attended predominantly ethnic-minority schools for most of their pre-teen and teen years, a time when linguistic innovation is particularly intense. (Cf. Dodsworth & Benton, 2017 for a robust new approach to the use of school data in modelling community change and linguistic variation.)

These insights from Stockholm have parallels in many other European multiethnolects. For example, Phrao and Maegaard's (2017) multiethnolectal Copenhagen data from the 2000s may have substantial phonetic differences from Quist's (2000) 1990s data despite a mere decade separating the two. The characterization of London Jamaican and Multicultural London English as separate varieties (Kerswill & Sebba, 2011) is similar: Although these speakers have

nearly identical social profiles, they are cohorts that grew up in dramatically different environments and so developed very different constellations of speech features.

3.2 Class-linked measures

Some exceptions to universal stratification by social class have been noted in the literature (Rickford, 1986; Chirkova et al., 2018). As with many other factors, social class or status can be considered a broadly etic factor that needs emic fine-tuning.

Sharma found that standardized class indexes (Goldthorpe, 2000; Hollingshead, 1975), e.g. a 3:5 weighted index of educational attainment and occupational prestige, showed a poor correspondence to the observable socioeconomic status of many individuals. One reason for this is that systematic change within individual lifetimes is particularly common in migrant settings (Platt, 2005). Migrants arriving in the UK almost always experience “status loss” upon migration: “on the one hand, they tend to be positively selected on resources from the origin country; on the other, they often occupy the lower rungs of the status ladder in receiving countries” (Engzell & Ichou, 2020, p. 471). This can lead to mixed indicators of class (e.g., high education but low occupational category) as well as mixed self-perception in terms of class. Their children frequently experience the inverse: rapid social mobility, with radically different occupations to their parents’, despite sometimes living within the same household. Other ethnographic details for this community included women suffering greater drops in status than men after divorce.

Housing is a particularly important detail. Fox & Sharma (2017) show that two completely different dialects, differing at all linguistic levels, have developed in Asian neighborhoods in East London (Multicultural London English) and West London (British Asian English). They trace the difference not to demographics—both studies were conducted in Asian majority neighbourhoods—but to a difference in working class and lower-middle class housing. Though these are adjacent social class categories, the former is strongly linked to multiethnic public

housing estates (and correspondingly schools as well) while the latter involves just enough income to rent homes on streets where same-ethnicity families and friends live. Here, social class exerts a profound influence on new dialect formation via housing and schooling.

Social class measures can also incorporate patterns of socialization, habitus, and even taste. Early modernist sociological work did this in various ways, recognizing the value of subjective measures of class alongside objective measures (Alford, 1962). Hollingshead and Redlich's (1958) social class index was built on interviewers' subjective placement of New Haven families in a seven-point social hierarchy. This was later used as the response variable on a regression calculation that had income, neighborhood, and educational level as predictors, and the equation was cross-validated against media consumption information (e.g., *New York News* vs. *New York Times*). The resulting formula was used by Wolfram (1969, p. 32–39) in his sociolinguistic investigation of Black speakers in Detroit to demonstrate that “AAVE variants” were far from monolithic and very much class-stratified. These include word-final consonant clusters, morpheme-medial and final /θ/, syllable-final /d/, post-vocalic /r/, copula deletion, suffixal /z/, and multiple negation (p. 49–54).

In stable communities with relatively low social mobility, simpler class metrics can work because parental occupation, taste, educational attainment, and occupation may align in regular ways. By contrast, contemporary communities in Europe and other changing societies may involve crisscrossing mobility that confounds conventional models, as observed above among London Asians. Late-modern Stockholm is similarly characterized as an “escalator region” because ethnic Swedish migrants from other parts of the country typically climb the class hierarchy while foreign migrants fall in the hierarchy (Andersson, 1996).

To account for this complexity, Young (2019) devised class measures from numerous dimensions of social information about his Stockholm participants, including income, current occupation, educational level, parental occupation, parental education, and taste. Education and

occupation were coded in the typical manner. Taste, however, was coded as *lowbrow* or *highbrow* in accordance with whether the participant expressed interest in 60 different activities mapped by Experian Ltd and InsightOne Nordic AB (2013) as meaningful for the market segmentation of Stockholm (pp. 148–155). A Principal Components Analysis was then conducted on the six aforementioned metrics, and the resulting index correlated with the stratification of rhythm and vowels in a meaningful way.

3.3 Ethnicity and race

One challenge in comparing race and ethnicity effects across studies is the use of ostensibly similar terms that are non-identical in reference (e.g., in the United States vs. in South Africa). There is no simple template to resolve this, save to recognize ethnicity should not simply be seen as a set of “objectively definable categories but as sets of cultural practices” (Hall-Lew & Wong, 2014, p. 572). In this brief section, we note a few ways in which terms can be contextualized, and we point readers to relevant further sources.

The most frequently coded elements are the participant’s race or ethnicity (see Mestrieh, this volume), or that of their network ties. For the coding of participant ethnicity, we refer the reader to Hall-Lew and Wong’s (2014) detailed discussion of shared conventions for recommended coding conventions for ethnicity where open-source data sharing is involved. As they observe, “like many other aspects of speaker identity, [ethnicity] is continually negotiated and reproduced in discourse, and therefore a challenge to code representatively.” (p. 564) They review a number of challenges, including the ambiguity of generic category labels such as ‘Asian American’, the changing nature of census categories, the differing orientation of individual participants to those labels, and the use of diverse forms of questionnaire-based elicitation. Like the present chapter, they recommend maximal coding as well as self-awareness regarding the limits of coding in fully capturing the nature of ethnicity in a community.

In the context of Stockholm, ethnicity and race labels are particularly challenging due to the Swedish “colorblind” approach to both. Asking about ethnic origin is unlikely to receive ethical approval from the country’s National Ethics Review Board, and discussions of “ras” (race) are particularly taboo in part due to the country’s prominent role in Race Biology in the early Twentieth Century. Nonetheless, omitting race as a variable would weaken predictive models and reinforce the erasure that non-white Swedes face. Swedes of color often self-identify using the proxy term “invandrare”, which directly translates as “immigrant”, but is rarely used to refer to white Western European migrants who actually constitute the largest immigrant group in the country. Rather, the term is used for non-white individuals regardless of their actual migration status. Similarly, those who claim to be “svensk” (Swedish) actually often have a non-Swedish parent or grandparent (Young, 2019, p. 85), but are white. To deal with this challenge, the Stockholm participants who discursively referred to Swedes as an outgroup (“they”) or used “invandrare” discursively as an ingroup (“we”) were coded as “invandrare”. Participants who discursively did the opposite were coded as “svensk”. This binary division proved significant *in interaction* with social class, with the pattern noted earlier of polarization between the non-white “invandrare” working class and the white “svensk” working class, with upper social groups in between and uniform across racial lines.

Beyond the challenges of category definitions, it is also always crucial for researchers to be aware that ethnicity is locally mediated. Wong and Hall-Lew (2014) argue persuasively that ethnicity-linked indexicality must always be examined in its regional context. Comparing two Chinese American communities experiencing robust societal change, they find that the use of a specific variant (the raised BOUGHT vowel) does not proceed in an identical manner despite similarities in community age and social change, because the variant has markedly different ambient indexical associations; for example, it indexes ‘stereotypical New Yorker’ in New York but not in San Francisco. Calder and King (2020) similarly compare two African-

American communities and find gender differences in the realization of /s/ in Rochester, New York, but not in Bakersfield, California (a non-urban community where African Americans are a small minority). An analysis of ethnicity independent of region and community ‘stage’ in either of the above studies would have impeded a clear interpretation of the data. Even if a study does not code contextual factors extensively, they must provide future users of a shared corpus (and the producers and readers of the resulting work) with background regional and demographic detail that may otherwise be lost or overlooked. Other chapters in the present volume discuss in more detail the inevitable intersection of ethnicity and race with other social factors.

It is becoming more common to draw on neighborhood statistics on ambient ethnicity, but researchers should not rely exclusively on such data as a proxy for individual social networks. Travis and Sheard (2020) found that census data on languages spoken in neighborhoods did not correlate with the ethnicity of people’s social networks: not surprisingly, people’s social networks did not line up closely with their immediate neighborhoods.

Network transcends neighborhood in this way in numerous other studies. Indeed, ethnic homophily is now one of the most common network measures (e.g. Cheshire et al., 2008; Li Wei, 1994; Matsumoto, 2010; Meyerhoff & Schlee, 2012; Newman, 2010; Wassink, 2016; Wong, 2010; Young, 2019). However, analysts should reflect on whether they are using ethnicity as a proxy for interlocutor speech, that is, assuming that an interlocutor’s ethnicity is a way to measure exposure to a specific speech style. In communities involving migration, these may not be correlated at all. For example, in the London Asian community, network ties designated simply as having ‘South Asian ethnicity’ combined individuals with British Asian accents and with Indian English accents. When analysed separately, the two factors behaved very differently across variables and generations (Sharma, 2017). It is therefore useful to either

separate these two components in coding, or at least provide metadata in a corpus that indicates whether the two align in a given community.

Finally, Hoffman & Walker's (2010) influential composite index for ethnic orientation recognizes this need to go beyond just network measures of ethnicity and additionally assess each individual's personal orientation to their ethnic group. In shifting communities, such identifications and allegiances can change dramatically from one generation to the next.

3.4 Social network

We do not describe social network coding in detail here, for reasons of space and also because recommendations for how to code social networks apply equally to communities in flux and those that are relatively stable (Sharma & Dodsworth, 2020). Whether a speech community is changing or not, researchers must aim for maximal coding of multiple dimensions to identify those relevant for their community. For example, Sharma (2017) noted major generational changes in network size and network type during her ethnography. Her methods therefore included new metrics of size and diversity of network and avoided eliciting a fixed number of named ties.

Transnational or wider out-group activity may play a particularly central role in changing communities (e.g., Bortoni-Ricardo, 1985; Dubois & Horvath, 1998; Li Wei, 1994; Matsumoto, 2010; Milroy, 1987). In Sharma's community, a transnational index (comprised of frequency of visits to South Asia, personal communication with ties in South Asia, origin of spouse/partner, and extent of South Asian work ties) showed a sharp generational divide in transnational activity, with older Gen 2 individuals consistently at the high end of the index with a dramatic decline in such activity among the younger Gen 2, which had overwhelmingly low index values even when some of them had very high levels of daily linguistic and cultural engagement with Punjabi culture within the UK (Sharma, 2014).

3.5 Other factors: bilingualism and cultural alignment

Many further social factors are covered in other chapters in the present volume. Here, in closing, we touch briefly on two factors: degree of bilingual language use and cultural affiliation.

In situations of migration and contact, bilingualism can be a decisive factor in the adoption of foreign variants. Once again, although levels of bilingual language use can be quantified and compared etically across communities, their measurement usually requires emic detail. In Sharma's West London research, information was collected on bilingual (English and Punjabi) language use from all participants using a 15-category scale of interlocutor type devised through ethnographic observation of locally relevant categories, including grandparents and "uncles and aunties", a term for acquaintances and relatives of parents' generation (Sharma & Sankaran, 2011; cf. Gal's 1978 inclusion of language used for praying). Given important variation among interlocutors, even within the family, her coding was based on interlocutor and speech task (Gal, 1978) rather than simply domains of interaction (Blom & Gumperz, 1972). Further questions established whether degrees of bilingualism had changed over the participants' lifespan.

While bilingualism taps into an individual's regular use of languages in their social network, cultural allegiance or affiliation can be more removed from such material ties. In the case of London, a person might have dense ties to their Asian family and may live in an Asian neighborhood but may align in taste and cultural preference with non-Asian culture.

Hoffman & Walker (2010) explore a wide range of survey question modules to tap into this dimension of cultural preference alongside more material network exposure. Sharma's project also gathered detailed cultural practice information (taste in music, TV, radio; participation in cultural activities) alongside network detail. But an exploration of correlations between linguistic practice and these fine cultural tastes revealed a statistical quirk that sociolinguists

rarely address. This is the risk that a factor that measures style choices such as music genres or clothing may be endogenous, that is to say, lacking in true independence from the dependent variable. In the case of Sharma's data, deep involvement with Punjabi bhangra music showed extremely high correlation with use of Asian-style post-alveolar /t/, but these seemed likely to be a linked set of self-presentation practices together driven by other independent factors. This example resembles other such correspondences between linguistic and non-linguistic style in the literature (Eckert 1996; Mendoza-Denton 2008). If we take seriously Eckert's (2000) proposal that linguistic practice is part of wider stylistic practice, then it may not be strictly accurate to treat cultural practice variables as independent.

That said, correlations between wider cultural practices—ideally established as independent of linguistic style—can help to cross-validate other predictive models, thereby strengthening (or weakening) their explanatory power. Adli (2013) proposes that “lifestyle can uncover sociolinguistically relevant differences in less heterogeneous subpopulations.” (p. 508) He found that French university students who were culturally active and politically critical preferred a *wh*-movement variant (*A qui elle prête sa carte bancaire?*) while those who were sports oriented preferred *wh*-in-situ (*Elle prête sa carte bancaire à qui?*), a finding that echoes the jock-burnout dichotomy in Eckert (2000). In such coding, it may also be worth noting Bourdieu's (1980) distinction between implicit tastes, which are deeply embodied and tied to inculcated, class-linked input in early childhood, and explicit “surface-level” tastes that are tied to conscious aspirations and identity work.

This chapter has focused on social factors rather than linguistic variables and internal factors, but we close with a few brief observations on the handling of linguistic variables. Needless to say, in changing communities, well-established linguistic variables may not always be the main variables of interest. Nevertheless, Hoffman & Walker (2010) make a strong case for the usefulness of examining both participation in wider changes-in-progress as well as community-

internal variants. If established variables are examined, it is also crucial to bear in mind that these may not always be governed by the ‘usual’ internal factors, especially if language contact is involved.

4 Conclusions and looking ahead

The examples in this chapter offer a roadmap for variationist sociolinguists working on changing communities who wish to produce cross-comparable analyses while avoiding an overly templatic approach. The preceding discussion has made a number of recommendations for capturing details of a changing social context:

- Following Hall-Lew and Wong (2014), we recommend maximal coding of social factors. Field practices should include coding multiple, transparent sub-components of factors rather than single, selective, or highly derived indexes;
- It is imperative to store rich metadata on as many relevant aspects of community social practices and history as possible, with comprehensive sociological description of key properties of the community, whether or not these are recorded in the form of coded variables;
- Etic factors need to be fine-tuned based on emic or ethnographic understanding of local realizations of social network (e.g. attention to sub-types of Asian ties), class (e.g. effects of public housing or schooling policy), gender, ethnicity, religion, local politics, and other local social categories;
- Researchers must be alert to potential changes in the relative influence of these multiple dimensions on social organization in a changing community;
- Age and generation are key variables for changing communities, but the relevant groupings, sometimes with specific ‘break points’ transforming lived experience, can only be identified through independent sociological and historical work;

- Coding protocols must document each factor carefully in a shared corpus to avoid misinterpreted factors and erroneous comparisons to other data;
- Intersectionality of social factors may be changing over time and requires attention in coding and documentation in protocols;
- Changing social groups can develop multi-dimensional speech repertoires, so sampling speech beyond sociolinguistic interview can be crucial for a complete picture of language practices.

The last of the points above relates to linguistic sampling rather than social factors. One of the clearest conclusions of recent research on changing communities has been the need to go beyond individual linguistic variables and interview speech and instead consider whole repertoires (Benor, 2010; Boyd et al., 2015; Sharma, 2011). Other chapters in the present volume explore these additional approaches to repertoire, situation, accommodation, and attitudes in further detail. Tapping into a wide range of speech settings was in fact strongly endorsed from the earliest days of variationist methodology: “the methods... described for overriding the constraints of the formal interview are only substitutes for the real thing and give us only fragments of the vernacular. A more systematic approach to recording the vernacular of everyday life is to allow the interaction of natural peer group itself to control the level of language produced.” (Labov, 1972, p. 115).

Gathering speech data beyond the interview is not always feasible, but even including broad ethnographic observations in metadata for shared corpora—for example, noting the prevalence of certain kinds of bidialectalism in the community—is useful for later researchers who may be increasingly removed from the field site.

While ethical constraints sometimes restrict the data we can collect, researchers should also think beyond their own projects, particularly in an age where a commitment to data sharing and transparency is on the rise. It can be productive to reflect on how our data will be used

decades from now, and so to document socio-demographic and linguistic detail as extensively as possible.

References

- Abd-El-Jawad, H. R. (1987). Cross-dialectal variation in Arabic: Competing prestigious forms. *Language in Society*, 359–367.
- Adli, A. (2013). Syntactic variation in French Wh-questions: A quantitative study from the angle of Bourdieu's sociocultural theory. *Linguistics*, 51(3), 473–515.
- Alford, R. R. (1962). A suggested index of the association of social class and voting. *Public Opinion Quarterly*, 417–425.
- Andersson, R. (1996). The geographical and social mobility of immigrants: Escalator regions in Sweden from an ethnic perspective. *Geografiska Annaler: Series B, Human Geography*, 78(1), 3–25.
- Benor, S. B. (2010). Ethnolinguistic repertoire: Shifting the analytic focus in language and ethnicity. *Journal of Sociolinguistics*, 14(2), 159–183.
- Bigelow, L., Gadanidis, T., Schlegl, L., Umbal, P., & Denis, D. (2020). Why are wasteyutes a ting? *University of Pennsylvania Working Papers in Linguistics*, 26(2), 11–20.
- Blom, J.-P., & Gumperz, J. J. (1972). Social meaning in Linguistic Structure: Code-Switching in Norway. In J. J. Gumperz & D. Hymes (Eds.), *Directions in Sociolinguistics* (pp. 407–434). Oxford: Basil Blackwell.
- Bortoni-Ricardo, S. M. (1985). *The urbanization of rural dialect speakers – A sociolinguistic study in Brazil*. Cambridge: Cambridge University Press.
- Bourdieu, P. (1980). The Aristocracy of Culture. *Media, Culture & Society*, 2, 225–254.
- Boyd, Z., Elliott, Z., Fruehwald, J., Hall-Lew, L., & Lawrence, D. (2015). An evaluation of sociolinguistic elicitation methods. In The Scottish Consortium for ICPHS 2015 (Ed.),

- Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: University of Glasgow.
- Calder, J., & King, S. (2020). Intersections between race, place, and gender in the production of /s/. *University of Pennsylvania Working Papers in Linguistics*, 26(2), 31–38.
- Cheshire, J., Fox, S., Kerswill, P., & Torgersen, E. (2008). Ethnicity, friendship network and social practices as the motor of dialect change: Linguistic innovation in London. *Sociolinguistica*, 1–23.
- Cheshire, J., Kerswill, P., Fox, S., & Torgersen, E. (2011). Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics*, 15(2), 151–196.
- Chirkova, K., Stanford, J. N., & Wang, D. (2018). A long way from New York City: Socially stratified contact-induced phonological convergence in Ganluo Ersu (Sichuan, Chin). *Language Variation and Change*, 30(1), 109–145.
- D’Arcy, A., & Tagliamonte, S. (2018). What’s age got to do with it? Problematizing the temporal dimension for linguistic explanation. *New Ways of Analyzing Variation 47*, New York, New York, USA. October 18-21, 2018.
- Dodsworth, R., & Benton, R. A. (2017). Social network cohesion and the retreat from Southern vowels in Raleigh. *Language in Society*, 46(3), 371–405.
- Dubois, S., & Horvath, B. M. (1998). Let’s tink about dat: Interdental fricatives in Cajun English. *Language Variation and Change*, 10(3), 245–261.
- Eckert, P. (1996). Vowels and nailpolish: The emergence of linguistic style in the preadolescent heterosexual marketplace. In J. Ahlers, L. Bilmes, M. Chen et al. *Gender and belief systems*. Berkeley, Berkeley women and language group.
- Eckert, P. (1998). Age as a Sociolinguistic Variable. In F. Coulmas (Ed.), *The Handbook of Sociolinguistics*. Malden, MA: Blackwell.

- Eckert, P. (2000). *Language variation as social practice: The linguistic construction of identity in Belten High*. Oxford: Wiley-Blackwell.
- 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.
- Engzell, P., & Ichou, M. (2020). Status loss: The burden of positively selected immigrants. *International Migration Review*, 54(2), 471–495.
- Experian Ltd & InsightOne Nordic AB. (2013). *Mosaic Sweden E-Handbook: The Classification of Swedish Consumers*. Stockholm: InsightOne Nordic AB.
- Forsberg, H. (2018). School competition and social stratification in the deregulated upper secondary school market in Stockholm. *British Journal of Sociology of Education*, 39(6), 891–907.
- Fox, S. P., & Sharma, D. (2017). *The language of London and Londoners* (P. Heinrich & D. Smakman, Eds.). London: Routledge.
- Fruehwald, J. (2017). Generations, lifespans, and the zeitgeist. *Language Variation and Change*, 29(1), 1–27.
- Gal, S. (1978). Peasant men can't get wives: Language change and sex roles in a bilingual community. *Language in Society*, 7(1), 1–16.
- Goldthorpe, J. H. (2000). *On sociology: Numbers, narratives, and the integration of research and theory*. Oxford: Oxford University Press.
- Grama, J., Travis, C. E., & Gonzalez, S. (2020). Ethnolectal and community change over time: Word-final (er) in Australian English. *Australian Journal of Linguistics*, 40(3), 346–368.
- Hall-Lew, L. (2009). *Ethnicity and phonetic variation in a San Francisco neighborhood* [PhD Thesis]. Department of Linguistics, Stanford University.

- Hall-Lew, L., & Wong, A. W. (2014). Coding for demographic categories in the creation of legacy corpora: Asian American ethnic identities. *Language and Linguistics Compass*, 8(11), 564–576.
- Hoffman, M. F., & Walker, J. A. (2010). Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change*, 22(1), 37–67.
- Hollingshead, A. B. (1975). *Four factor index of social status* [unpublished manuscript].
- Hollingshead, A. B., & Redlich, F. C. (1958). *Social class and mental illness: Community study*. Hoboken: John Wiley & Sons Inc.
- Holmlund, H., Häggblom, J., Lindahl, E., Martinson, S., Sjögren, A., Vikman, U., & Öckert, B. (2014). *Decentralisering, skolval och fristående skolor: Resultat och likvärdighet i svensk skola* [Decentralization, school choice, and charter schools: Results and equivalency in the Swedish school]. Stockholm: Institute for Evaluation of Labour Market and Education Policy.
- Hua, X., Meakins, F., Algy, C., & Bromham, L. (2021). Language change in multidimensional space: New methods for modelling linguistic coherence. *Language Dynamics and Change* (12): 78-123.
- Kerswill, P. (1994). *Dialects converging: Rural speech in urban Norway*. Clarendon Press.
- Kerswill, P. (2018). Dialect formation and dialect change in the Industrial Revolution: British vernacular English in the nineteenth century. In L. Wright (Ed.), *Southern English Varieties Then and Now* (pp. 8–38). Berlin and Boston: De Gruyter Mouton.
- Kerswill, P., & Sebba, M. (2011). From London Jamaican to British youth language: The transformation of a Caribbean post-creole repertoire into a new Multicultural London English. *2011 Summer Conference of the Society for Pidgin and Creole Linguistics, July–August, University of Ghana, Accra*.

- Kotsinas, U.-B. (1988). Stockholmspråk i förändring [Stockholm's language in change]. In G. Pettersson (Ed.), *Studier i svensk språkhistoria* [Studies in Swedish language history], vol. 1, pp. 133–147, Lund University Press.
- Labov, W. (1972). Some principles of linguistic methodology. *Language in Society*, 1(1), 97–120.
- Labov, W. (1994). *Principles of linguistic change, Volume 1: Internal factors*. Oxford: Blackwell.
- Labov, W., Rosenfelder, I., & Fruehwald, J. (2013). One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language*, 30–65.
- Lentin, A. (2008). Europe and the Silence about Race. *European Journal of Social Theory*, 11(4), 487–503.
- Li Wei. (1994). *Three generations, two languages, one family: Language choice and language shift in a Chinese community in Britain*. Toronto: Multilingual Matters.
- Lightfoot, D. (1991). *How to set parameters: Arguments from language change*. Cambridge: MIT Press.
- Lightfoot, D. (1997). Catastrophic change and learning theory. *Lingua*, 100(1–4), 171–192.
- Matsumoto, K. (2010). The role of social networks in the post-colonial multilingual island of Palau: Mechanisms of language maintenance and shift. *Multilingua - Journal of Cross-Cultural and Interlanguage Communication*, 29(2), 133–165.
- Meakins, F. (2008). Land, language and identity The socio-political origins of Gurindji Krioll. In M. Meyerhoff & N. Nagy (Eds.), *Social Lives in Language* (pp. 69–94). Amsterdam: John Benjamins Publishing.
- Mendoza-Denton, Norma. (2008). *Homegirls: Language and cultural practice among Latina youth gangs*. Maiden, MA: Blackwell.

- Meyerhoff, M., & Schlee, E. (2012). Variation, contact and social indexicality in the acquisition of (ing) by teenage migrants 1. *Journal of Sociolinguistics*, 16(3), 398–416.
- Milroy, L. (1987). *Language and Social Networks* (2nd ed.). Oxford: Basil Blackwell.
- Nagy, N. (2018). Linguistic attitudes and contact effects in Toronto's heritage languages: A variationist sociolinguistic investigation. *International Journal of Bilingualism*, 22(4), 429–446.
- Neergaard, A. (2017). The Swedish Model in transition: Trade unions and racialised workers. In A. Ålund, C.-U. Schierup, & A. Neergaard (Eds.), *Reimagining the Nation: Essays on twenty-first-century Sweden* (pp. 85–117). Frankfurt am Main: Peter Lang.
- Newman, M. (2010). Focusing, implicational scaling, and the dialect status of New York Latino English 1. *Journal of Sociolinguistics*, 14(2), 207–239.
- Otheguy, R., Zentella, A. C., & Livert, D. (2007). Language and dialect contact in Spanish in New York: Toward the formation of a speech community. *Language*, 770–802.
- Pharao, N., & Maegaard, M. (2017). On the influence of coronal sibilants and stops on the perception of social meanings in Copenhagen Danish. *Linguistics*, 55(5), 1141–1167.
- Pike, K. L. (1967). *Etic and emic standpoints for the description of behavior*. Amsterdam: Mouton & Co.
- Platt, L. (2005). The intergenerational social mobility of minority ethnic groups. *Sociology*, 39(3), 445–461.
- Poplack, S., & Malvar, E. (2007). Elucidating the transition period in linguistic change: The expression of the future in Brazilian Portuguese. *Probus*, 19(1), 121–169.
- Quist, P. (2000). Ny københavnsk 'multietnolekt'—Om sprogbrug blandt unge i sprogligt og kulturelt heterogene miljøer [New Copenhagen 'multiethnolect'—On language use among youth in linguistically and culturally heterogeneous milieus]. *Danske Talesprog*, 1, 143–211.

- Rickford, J. R. (1986). The Need for New Approaches to Social Class Analysis in Sociolinguistics. *Language & Communication*, 6(3), 215–221.
- Satyanath, S. (2015). Language variation and change: The Indian experience. In D Smakman & P. Heinrich (Eds.), *Globalising Sociolinguistics: Challenging and Expanding Theory* (pp. 107–122). London: Routledge.
- Sharma, D. (2011). Style repertoire and social change in British Asian English. *Journal of Sociolinguistics*, 15(4), 464–492.
- Sharma, D. (2014). Transnational flows, language variation, and ideology. In M. Hundt & D. Sharma (Eds.), *English in the Indian diaspora* (pp. 215–242). Amsterdam: John Benjamins.
- Sharma, D. (2017). Scalar effects of social networks on language variation. *Language Variation and Change*, 29(3), 393–418.
- Sharma, D., & Dodsworth, R. (2020). Language Variation and Social Networks. *Annual Review of Linguistics*, 6, 341–361.
- Sharma, D., & Sankaran, L. (2011). Cognitive and social forces in dialect shift: Gradual change in London Asian speech. *Language Variation and Change*, 23(3), 399–428.
- Smakman, Dick, & Heinrich, P. (2015). *Globalising sociolinguistics: Challenging and expanding theory*. London: Routledge.
- Stanford, J. N. (2016). A call for more diverse sources of data: Variationist approaches in non-English contexts. *Journal of Sociolinguistics*, 20(4), 525–541.
- Stroub, K. J., & Richards, M. P. (2013). From resegregation to reintegration: Trends in the racial/ethnic segregation of metropolitan public schools, 1993–2009. *American Educational Research Journal*, 50(3), 497–531.
- Travis, Catherine, & Sheard, E. (2020). Measuring ethnic orientation across corpora and communities. *2020 Conference of the Australian Linguistics Society*.

- Wacquant, L. (2008). *Urban outcasts: A comparative sociology of advanced marginality*. Cambridge: Polity.
- Wassink, A. (2016). *Ethnic homophily, localness and participation in regional linguistic change: The social networks of minority ethnicity speakers in Washington State* [unpublished manuscript].
- Wolfram, W. (1969). *A Sociolinguistic Description of Detroit Negro Speech, Urban Language Series, No. 5*. Center for Applied Linguistics.
- Wong, A. (2007). Two vernacular features in the English of four American-born Chinese. *University of Pennsylvania Working Papers in Linguistics*, 13(2), 217–230.
- Wong, A. W. (2010). New York City English and second generation Chinese Americans. *English Today*, 26(3), 3–10.
- Wong, A. W., & Hall-Lew, L. (2014). Regional variability and ethnic identity: Chinese Americans in New York City and San Francisco. *Language & Communication*, 35, 27–42.
- Yaeger-Dror, M., & Cieri, C. (2014). Introduction to the Special Issue on Archiving Sociolinguistic Data. *Language and Linguistics Compass*, 8(11), 465–471.
- Young, N. (2019). *Rhythm in late-modern Stockholm – Social stratification and stylistic variation in the speech of men* [PhD Thesis], Department of Linguistics, Queen Mary, University of London. <http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-178897>.

The project leading to this publication has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 892963