1 Introduction

1.1 FOREGROUNDING FIELD METHODS

Empirical data on language in its social context is at the heart of sociolinguistic study. Nonetheless, researchers have lamented that the field methods whereby such data are gathered have always been rather "inconspicuous" (Feagin 2013: 37). Write-ups of research projects focus on results and interpretations – not on how the data from which the findings are drawn were actually obtained. Indeed, Tagliamonte (2006a: 17) maintains that "fieldwork methods may be the best-kept secret of sociolinguistics." In addition, the few publications that do focus on sociolinguistic field methods (e.g. Bailey and Tillery 2004; Bailey, Wikle, and Tillery 1997b; Feagin 2013; Labov 1972a, 1984; Macaulay 2009; Milroy and Gordon 2003; Schilling-Estes 2007; Tagliamonte 2006a; Wolfson 1976) all point to the crucial role that research design and method play in shaping the data and, hence, the conclusions we derive. Further, sociolinguistic research foci have both broadened and become more nuanced, and data-related technology of course continues to transform itself every day. Correspondingly, research methods have become more varied and sophisticated and continue to evolve at a rapid pace.

The purpose of this book is to at last reveal the "secrets" of sociolinguistic fieldwork. It is both a how-to for students and researchers who need to design and conduct studies on language in its social context and a thoughtful exploration of the chief methods in sociolinguistic data collection, including examination of which methods work best for which purposes; evaluation of the strengths and weaknesses, advantages and disadvantages of each approach; and consideration of the theoretical assumptions underlying different methodological approaches. The book covers both quantitative and qualitative methodologies, as well as small- and large-scale studies. However, its chief focus is on how a range of methods and perspectives can be applied to variationist sociolinguistics, or quantitative sociolinguistics, the

1

subfield that has dominated modern sociolinguistic study since its inception in the early 1960s. Thus, in addition to providing practical guidelines, the book also explores several recent and ongoing "turns" in current variationist sociolinguistics, including integrating ethnographic study into traditional survey-based methods; and reconsidering traditional conceptualizations of and methods for studying stylistic variation, including investigating variation as it unfolds in interaction. Also of crucial importance throughout the book are questions of fieldwork ethics, including not only consideration of how to ensure that your research causes no harm but also how your research can actually be of benefit to the communities you study.

The book is intended to be general enough for use by researchers on spoken, signed, and written language variation. However, where research methods become more specialized (e.g. in considering issues of obtaining high-quality recordings and preserving the confidentiality of identifying information), the emphasis is on methods in spoken language research, since sociolinguistics has been most closely focused on spoken language since its inception and since the "secular," everyday linguistic usages that sociolinguists seek have traditionally been concentrated in spoken registers (though, naturally, this has been changing in recent decades as written electronic communications has become increasingly commonplace). In addition, although human communication encompasses not only the linguistic signal *per se* but also gesture, position, facial expression, etc., this book follows the mainstream of variation study in focusing on the verbal channel, though of course it is necessary to concurrently consider the extralinguistic context that shapes and is shaped by the linguistic signal.

Throughout the book, methodological and theoretical approaches and issues will be illustrated with examples from sociolinguistic studies ranging from the first foundational studies in the field, to later groundbreaking and paradigm-shifting works, to important ongoing work in the early decades of the twenty-first century. Illustrations will also come from a range of languages and sociolinguistic fieldwork locales from across the globe. At the same time as I provide breadth, however, I will provide depth of detail by drawing many examples from studies with which I have some degree of personal connection, whether through direct involvement with the research project, through direct contact with the project directors or project team, or in my role as advisor to researchers and students who have sought my guidance in designing and conducting projects on language variation in its social setting.

The necessity for a limited scope is obvious, given the sheer impossibility of representing all of the many types of peoples and places that

have been studied since the inception of modern sociolinguistics. But my personal focus is also driven by personal convictions: first, that sociolinguistic research is best conducted from a dual insider–outsider (i.e. participant–observer) perspective, as I discuss in Chapters 3, 5, and 7; and second, that our best reflections upon fieldwork methods are drawn from projects with which we have at least some degree of "insider" involvement during the fieldwork process. I believe our best insights about field methods come not from reading neat and polished accounts, where process almost always takes a back seat to the final analytical product, but rather from the "messy" and necessarily personal experience of fieldwork itself. My insights have been greatly enriched by connection with a wide array of projects throughout the US and the world, and my hope is to share these insights with the range of readers who stand to benefit from this book as they plan, conduct, and consider their own sociolinguistic and related studies.

As a special aid to the beginning researchers who are one of the chief audiences for this book, I highlight the research experiences of students at Georgetown University in a series of insets on "Students in the field." These students have conducted research on a wide range of communities throughout the world, and they have plenty of knowledge to share, including insight into what worked well, what did not work, and what surprises they encountered. They also share with us their emotional experiences. Sociolinguistic fieldwork can be daunting, but it can also be the most rewarding part of being a researcher, in terms of how much we learn about language and life, and in terms of the personal connections we forge as we engage with communities in a mutual learning process.

The text presumes grounding in introductory sociolinguistics, though it will also be useful for any scholar who wishes or needs to gather data on language in its social context; for example, those working in applied linguistics, linguistic anthropology, historical linguistics, and language endangerment.

1.2 SOCIOLINGUISTIC FIELD METHODS: A BRIEF HISTORY AND OVERVIEW

Variationist field methods have their roots in several traditions of empirical language study, including historical / comparative linguistics; the traditional dialectology (dialect geography) that stems from comparative linguistics; and the American descriptivist / structuralist linguistics of the early twentieth century (Chambers and Trudgill 1998:

13–54; Milroy and Gordon 2003: 1–22). In addition, qualitatively oriented approaches to the interrelation between language and society, namely, linguistic anthropology, the ethnography of communication, and interactional sociolinguistics, have played an important role in shaping quantitative sociolinguistics and have exerted increasing influence as variation analysis has developed over the past four decades (e.g. Milroy and Gordon 2003: 1–22).

The first systematic dialect studies were a natural outgrowth of historical linguists' interest in the interrelations between modern languages and language varieties and what these relations could reveal about historical connections and developments. The earliest dialect geographic studies were Wenker's survey of German dialects (1876–1887), Guillieron's survey of French dialects (1896–1900), and Jaberg and Jud's surveys of the Italian dialects of Italy and southern Switzerland in the first decades of the twentieth century. These studies served as the inspiration for the systematic study of American English dialects, initiated with the inauguration of the Linguistic Atlas of the United States and Canada project in 1930. Concurrently, in the early decades of the twentieth century, American descriptivist linguists inaugurated their own tradition of empirical linguistics with their investigations of the structures of Native American languages. (See Chambers and Trudgill 1998: 13–53 for a detailed discussion of the development of variationist sociolinguistics out of traditional dialectology.)

Because of their interest in linguistic history (i.e. diachronic linguistics), early dialectologists concentrated on obtaining data from speakers whose speech was believed to be most reflective of older speech forms – non-mobile, older, rural males with little formal education. And whereas the focus of American structuralist linguists was chiefly on the interrelation of the elements of language at a given moment in time (i.e. synchronic linguistics), they shared with dialectologists a focus on empirical data and linguistic fieldwork, as well as an interest in how similarities across current languages can be revealing of historical linguistic relationships.

However, it was not until the latter half of the twentieth century that diachronic and synchronic linguistics were truly brought together, with the pioneering work of Uriel Weinreich and William Labov and the advent of modern variationist sociolinguistics (Weinreich, Labov, and Herzog 1968). Like historical linguists and traditional dialect geographers, variationists are keenly interested in language change; however, their focus is not typically on long-completed changes but rather on ongoing language change. The pioneering study in this regard is Labov's 1963 study of Martha's Vineyard, an island off the coast of

Massachusetts (New England, USA), in which Labov demonstrated that, contrary to prior belief, it is indeed possible to observe language change in progress, through studying the synchronic patterning of community variation (1963, 1972c: 1–42). Crucially, in order to fully understand language change (and linguistic systematicity), we must investigate variation as it patterns according to both linguistic and social factors. And the social context of language variation is not limited to its patterning across geographic space but extends also to its patterning across social groups (e.g. age groups, social class groups, ethnic groups) as well as to its social meaning for both groups and individuals (e.g. a particular linguistic feature may be seen as a marker of regional identity or associated with a particular character trait).

Because variationists recognize that language variation must be studied across the entire social spectrum, their target populations are much broader than those of traditional dialect geographers, encompassing a full range of age, gender, socioeconomic, and ethnic groups. In addition, because they are focused on ongoing change rather than historic language forms, variationists often center their studies in urban rather than rural locations and may focus on younger populations as well. Some of the many important studies of language variation in urban settings include such pioneering works as Labov's (1966) study of New York City English; Wolfram's (1969) study of African American English in Detroit, Michigan; Cedergren's (1973) study of Spanish in Panama City; Trudgill's (1974) study of Norwich, England; and Macaulay's (1977) work in Glasgow, Scotland. In addition, there have been important ongoing studies of, for example, Montreal French (Sankoff, Sankoff, Laberge, and Topham 1976; Thibault and Vincent 1990; Vincent, Laforest, and Martel 1995) and Philadelphia English (Labov 1994), as well as re-studies of Martha's Vineyard (Blake and Josey 2003; Pope, Meyerhoff, and Ladd 2007), Panama City (Cedergren 1973, 1984), Norwich (Trudgill 1988), and New York City (r) (Labov 1994), among others. Finally, recent and ongoing studies of language variation and change in urban settings increasingly are taking language and dialect contact into account, as evidenced, for example, in Horvath's (1985) study of the socioethnic varieties in Sydney, Australia; Kerswill's studies of dialect contact, dialect leveling, and new dialect formation in southeast England (Kerswill 1996, 2013; Kerswill and Williams 2000, 2002); and Walker and Hoffman's ongoing studies of the highly multiethnic and multilingual city of Toronto, Canada (e.g. Hoffman and Walker 2010).

Variationist sociolinguistics also differs from traditional dialectology and American descriptivist linguistics in terms of what type of speech

data is sought, and hence on primary methods of data collection. Dialectologists traditionally relied on lengthy questionnaires designed to elicit information on the use of a variety of lexical, phonological, and grammatical features, sometimes directly (e.g. "Have you ever heard the term 'snap beans' used for the beans that you break in half to cook? If yes, how often would you use that term …?") but more often indirectly ("What word would you use for the beans that you break in half to cook?"; see Bailey, Tillery, and Wikle 1997a). Questionnaires were distributed by mail or by traveling fieldworkers across very large areas (often across a whole country), and there tended to be very few respondents (often a single respondent) from each location within the wider survey area. And again, the social spectrum was typically narrow, with most informants being those who presumably represented the oldest speech forms, especially in European dialect surveys.

The American structuralists also relied on elicited language forms, though their interest was in piecing together the linguistic systems of undocumented or little-studied languages rather than in areal or social variation within well-known languages. Thus, their elicitations focused on obtaining information on the interrelation between linguistic elements rather than isolated items, through such tasks as judgments of "same" or "different," to determine minimal pairs, and sentence permutation tasks, to determine interrelations such as the subject and verb forms constituting the person–number paradigm. Again, though, the number of informants was limited – not just per location but per entire language – and there was typically little or no information on intra-language variation.

Again, because their focus is on intra-language variation, including social variation within populations as well as social and geographic variation across social groups, variationists gather data from many speakers in each community of study. In addition, their data must include many tokens of each of the variable linguistic features they are interested in, not just one or two examples of each, in order to gain a full picture of the linguistic and social factors that condition regularly patterned variability. Finally, variationist methodology has long been rooted in Labov's vernacular principle, which states:

> the style which is most regular in its structure and in its relation to the evolution of the language is the vernacular, in which the minimal attention is paid to speech (Labov 1972a: 112).

To fulfill the need for long stretches of connected, unselfconscious speech, Labov devised the so-called **sociolinguistic interview** – a casual interview designed to approximate the flow of an everyday relaxed

conversation, as well as to focus speakers' attention on what they are talking about rather than on their speech itself (Labov 1972a; Wolfram and Fasold 1974). Interviewer questions are kept brief, and interviewees are encouraged to talk at length about topics of interest to them rather than asked to provide concise information on particular language features. Many researchers have modified the basic format of the sociolinguistic interview, and everyone tailors it to suit each individual community under study. In addition, some variationists have questioned the focus on vernacular, unselfconscious speech and have demonstrated how more selfconscious styles lend valuable insight into the patterning of variation and change (Coupland 2001a, 2001b, 2007; Eckert 2000: 213–228; Milroy and Gordon 2003: 49–51; Schilling-Estes 1998). Nonetheless, the sociolinguistic interview as originally conceived remains the most important item in the variationist sociolinguist's fieldwork toolkit. This vital tool will be discussed at length in Chapter 3, including its strengths and limitations in terms of investigating a range of speech styles.

Case study: Investigating change through studying systematic variation – Labov's pathbreaking study of Martha's Vineyard

In the early 1960s, Labov conducted a pioneering study of Martha's Vineyard, an island off the coast of Massachusetts, in the Northeastern US (1972c: 1–42). His focus was on the apparent increase in the production of the /ay/ and /aw/ diphthongs with centralized rather than low nuclei, a development which seemingly was taking place at the same time as the island was being transformed from a traditional, self-sustaining maritime community to a magnet for vacationers and new residents from mainland communities. An increase in centralization would be unusual, since historically the centralized forms developed into today's diphthongs with low central nuclei, and we do not typically expect a community to reverse a language change, or "turn its back on the history of the English language" (Labov 1972c: 25).

In order to investigate whether the suspected movement away from [aɪ] and [aʊ] toward increasing [əɪ] and [əʊ] was indeed taking place, Labov conducted tape-recorded interviews with numerous islanders, representative of important community social groups. Through subsequent quantitative analysis (i.e. counting all tokens of centralized and non-centralized /ay/ and /aw/ in the recordings and noting the linguistic contexts and

social factors surrounding each occurrence), Labov was able to demonstrate that the centralization of /ay/ and /aw/ showed regular patterning according to linguistic factors such as following environment. For example, speakers were more likely to produce centralized nuclei for /ay/ before voiceless than voiced obstruents, so that [rəɪt] is more likely than [rəɪd].

In addition, he demonstrated that younger speakers showed more centralization than older speakers. This observation, coupled with data from earlier dialect geographic studies (and general information on the history of English), confirmed that the centralization of /ay/ and /aw/ was indeed on the rise in the island community.

Finally, Labov noted a correlation between speakers' orientation to Martha's Vineyard and usage levels for centralized /ay/ and /aw/: Those who positively valued the local community and its traditional ways showed higher usage levels for the centralized variants than those with more neutral or negative feelings toward the island and its traditional inward focus and isolation from outside forces.

Labov's study thus illustrates several key elements of variation theory and its inhering methodology:

1) We can investigate language change in progress through investigating the systematic patterning of language variation in a given population at a single moment in time.
2) To fully investigate the patterning of language variation, we must take into account both the linguistic and social factors that condition and / or correlate with the variable features in question.
3) The social factors that correlate with / condition language change include not only readily observable demographic factors such as age, occupation, and ethnicity, but also, and perhaps most importantly, attitudinal and identificational factors such as orientation toward one's local community.

1.3 ENRICHING QUANTITATIVE SOCIOLINGUISTICS WITH QUALITATIVE DATA / METHODS

Though essentially quantitative in nature, variationist sociolinguistics shares much in common with the qualitatively oriented disciplines of linguistic anthropology, the ethnography of communication, and interactional sociolinguistics, since all are focused on the interrelation of

language and society / culture. Central to anthropological / ethno-
graphic approaches is a concern for coming to understand cultures
and communities from the perspectives of their members, through
long-term participant-observation in community life. In addition to
gathering copious amounts of language data, variationists too seek
community understandings, though often the variationist focus is
more linguistic than cultural, since language change and linguistic
systematicity remain central concerns. Labov's Martha's Vineyard
study is a model not only in its careful linguistic analysis but also in
its concern for uncovering the sociocultural meanings of the linguistic
features under study for the people who use them. Other, larger scale
sociolinguistic studies in urban contexts have sometimes sacrificed
sociocultural depth for linguistic and social breadth, as well as for
replicability across communities, and relied on pre-imposed social
categorizations and meanings, for example by dividing the community
into "typical" socioeconomic groups based on "typical" measures of
income, education, occupation, etc., rather than uncovering more
nuanced local categorizations (e.g. Eckert 2000: 16–25, 2005; Rickford
1986). However, variationist sociolinguistics has never really strayed
far from its ethnographic roots and its focus on the local as well as the
global, as evidenced, for example, as early as Labov's studies of the local
meaning of features of African American Vernacular English among
Harlem youth in the 1960s and continuing with Milroy and Milroy's
studies of the local meanings connoted by use of traditional vernacular
forms among members of various types of social networks in Belfast,
Northern Ireland, in 1975–1981 (1985). As our studies have progressed,
there has been ever-increasing recognition that the social meaning and
hence regular patterning of language variation (and the course of
language change) is best understood by combining broad-brush surveys
using sociolinguistic interview methods with in-depth ethnographic
analysis of local sociocultural meanings, practices, and categories.

A model study exemplifying the synergistic union of quantitative
variationist and qualitative ethnographic methods is Penelope Eckert's
Linguistic Variation as Social Practice (2000). Eckert's work on adolescent
language patterns in a suburban Detroit (Michigan) high school is a
thorough and careful quantitative analysis of language variation and
ongoing change utilizing the best practices (including field method
techniques) of variationists. Crucially, though, it is grounded in exten-
sive ethnographic observation of the high school under study and of
American high school and adolescent culture more generally, and so
Eckert is able to insightfully examine interrelations between patterns
of variation and change and *local* social categories, meanings, and

practices. For example, she is able to show that particular features associated with a current change in progress correlate more closely with kids' *own* social categories (i.e. the school-oriented "jocks" and street-oriented "burnouts") than their parents' socioeconomic class groups, as defined by "objective" measures imposed from outside community perspectives. In addition, she shows how language variation may play a greater role in demarcating which practices the kids engage in than the groups to which they belong. Thus, for example, among the jock and burnout boys in her study, usage levels for the newest vowel pronunciations are more closely correlated with the activity of "cruising" (i.e. traveling frequently to potentially risky areas of Detroit) than with membership in the jock or burnout social categories (pp. 150–153).

At the same time that variationists are broadening their language studies by including solid ethnographic studies of community sociocultural perspectives, they are also looking more deeply into their data by investigating how variants pattern in unfolding discourse. In this way, they gain understanding not only of the local community meanings of language features but of individual and interpersonal meanings as well. In addition, through such study we come to understand that linguistic variants are not simply reflective of membership in certain social groups or of particular attitudinal postures but that people actively utilize particular features to create and shape individual and group identities and (language) styles, as well as project particular stances – that is, relationships to what is being said and to one's interlocutors (see Section 4.3 for more on the notion of "stance").

Further, just as identities and styles are not fixed but fluid, so too are the meanings of variables malleable and multifaceted. For example, in order to examine the patterning and meaning of one of the most noticeable features of one of the most distinctive dialects of the US, "Pittsburghese," the variety associated with Pittsburgh, Pennsylvania, and the immediate vicinity, Johnstone and Kiesling (2008) combined quantitative study of the production and perception of monophthongal vs. diphthongal /aw/ with analysis of how /aw/ was used in connected speech (interviews and other data-gathering tasks), as well as how people talked *about* /aw/. (They also conducted historical investigation of the rise of Pittsburghese as an entity in its own right.) Their findings revealed that, far from remaining fixed, the meaning of /aw/ has changed drastically – from a highly localized but quite unnoticeable feature of Pittsburgh speech, to a marker of regionality to be avoided in "correct" speech, to a stereotype that can be used to perform the Pittsburgh dialect or Pittsburgh personas.