The Handbook of Language Variation and Change

Second Edition

Edited by

J. K. Chambers and Natalie Schilling

WILEY-BLACKWELL

The Handbook of Language Variation and Change

Blackwell Handbooks in Linguistics

This outstanding multi-volume series covers all the major subdisciplines within linguistics today and, when complete, will offer a comprehensive survey of linguistics as a whole.

Already published

The Handbook of Child Language
Edited by Paul Fletcher and Brian MacWhinney

The Handbook of Phonological Theory, Second Edition
Edited by John A. Goldsmith, Jason Riggle, and Alan C. L. Yu

The Handbook of Contemporary Semantic Theory
Edited by Shalom Lappin

The Handbook of Sociolinguistics
Edited by Florian Coulmas

The Handbook of Phonetic Sciences, Second Edition
Edited by William J. Hardcastle and John Laver

The Handbook of Morphology
Edited by Andrew Spencer and Arnold Zwicky

The Handbook of Japanese Linguistics
Edited by Natsuko Tsujimura

The Handbook of Linguistics
Edited by Mark Aronoff and Janie Rees-Miller

The Handbook of Contemporary Syntactic Theory
Edited by Mark Baltin and Chris Collins

The Handbook of Discourse Analysis
Edited by Deborah Schiffrin, Deborah Tannen, and Heidi E. Hamilton

The Handbook of Language Variation and Change, Second Edition
Edited by J.K. Chambers and Natalie Schilling

The Handbook of Historical Linguistics
Edited by Brian D. Joseph and Richard D. Janda

The Handbook of Language and Gender
Edited by Janet Holmes and Miriam Meyerhoff

The Handbook of Second Language Acquisition
Edited by Catherine J. Doughty and Michael H. Long

The Handbook of Bilingualism and Multilingualism, Second Edition
Edited by Tej K. Bhatia and William C. Ritchie

The Handbook of Pragmatics
Edited by Laurence R. Horn and Gregory Ward

The Handbook of Applied Linguistics
Edited by Alan Davies and Catherine Elder

The Handbook of Speech Perception
Edited by David B. Pisoni and Robert E. Remez

The Handbook of the History of English
Edited by Ans van Kemenade and Bettelou Los

The Handbook of English Linguistics
Edited by Bas Aarts and April McMahon

The Handbook of World Englishes
Edited by Braj B. Kachru; Yamuna Kachru, and Cecil L. Nelson

The Handbook of Educational Linguistics
Edited by Bernard Spolsky and Francis M. Hult

The Handbook of Clinical Linguistics
Edited by Martin J. Ball, Michael R. Perkins, Nicole Müller, and Sara Howard

The Handbook of Pidgin and Creole Studies
Edited by Silvia Kouwenberg and John Victor Singler

The Handbook of Language Teaching
Edited by Michael H. Long and Catherine J. Doughty

The Handbook of Language Contact
Edited by Raymond Hickey

The Handbook of Language and Speech Disorders
Edited by Jack S. Damico, Nicole Müller, Martin J. Ball

The Handbook of Computational Linguistics and Natural Language Processing
Edited by Alexander Clark, Chris Fox, and Shalom Lappin

The Handbook of Language and Globalization
Edited by Nikolas Coupland

The Handbook of Hispanic Linguistics
Edited by Manuel Díaz-Campos

The Handbook of Language Socialization
Edited by Alessandro Duranti, Elinor Ochs, and Bambi B. Schieffelin

The Handbook of Intercultural Discourse and Communication
Edited by Christina Bratt Paulston, Scott F. Kiesling, and Elizabeth S. Rangel

The Handbook of Historical Sociolinguistics
Edited by Juan Manuel Hernández-Campoy and Juan Camilo Conde-Silvestre

The Handbook of Hispanic Linguistics
Edited by José Ignacio Hualde, Antxon Olarrea, and Erin O'Rourke

The Handbook of Conversation Analysis
Edited by Jack Sidnell and Tanya Stivers

The Handbook of English for Specific Purposes
Edited by Brian Paltridge and Sue Starfield

The Handbook of Language Variation and Change

Second Edition

Edited by

J.K. Chambers and Natalie Schilling

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

This second edition first published 2013

Edition History: Blackwell Publishing Ltd (1e, 2002)

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

Registered Office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Offices
350 Main Street, Malden, MA 02148-5020, USA
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services, and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data

The handbook of language variation and change / Edited by J.K. Chambers and Natalie Schilling. – Second Edition.
pages cm
Includes index.
ISBN 978-0-470-65994-6 (cloth)
1. Language and languages–Variation. 2. Linguistic change. I. Chambers, J. K. editor of compilation. II. Schilling-Estes, Natalie, editor of compilation.
P120.V37H365 2013
417'.7–dc23
2013004420

A catalogue record for this book is available from the British Library.

Cover image: Brian Whelan, *Cityscape*, 36 x 48 ", mixed media on board. © Brian Whelan 2008, www.brianwhelan.co.uk
Cover design by Workhaus

Set in 10/12 pt Palatino by Toppan Best-set Premedia Limited

1 2013

For William Labov

whose work is referred to in every chapter and
whose ideas imbue every page

Contents

Illustrations

Contributors

Sharon Ash
University of Pennsylvania, USA
ash@mail.med.upenn.edu

Guy Bailey
University of Alabama, USA
guy.bailey@ua.edu

Robert Bayley
University of California, Davis, USA
rjbayley@ucdavis.edu

David Britain
University of Bern, Switzerland
britain@ens.unibe.ch

J.K. (Jack) Chambers
University of Toronto, Canada
jack.chambers@utoronto.ca

Patricia Cukor-Avila
University of North Texas, USA
pcavila@unt.edu

Ralph W. Fasold
Georgetown University, USA
fasoldr@georgetown.edu

Crawford Feagin
Arlington, Virginia, USA
feagin@starpower.net

Carmen Fought
Pitzer College, USA
cfought@pitzer.edu

Matthew J. Gordon
University of Missouri at Columbia, USA
gordonmj@missouri.edu

Raymond Hickey
University of Duisburg and Essen, Germany
raymond.hickey@uni-due.de

Tyler Kendall
University of Oregon
tsk@uoregon.edu

Paul Kerswill
York University, England
paul.kerswill@york.ac.uk

Scott F. Kiesling
University of Pittsburgh, USA
kiesling@pitt.edu

Sam Kirkham
Lancaster University, England
s.kirkham@lancaster.ac.uk

Carmen Llamas
University of York, England
carmen.llamas@york.ac.uk

Ronald Macaulay
Pitzer College, USA
ronaldmacaulay@cs.com

Miriam Meyerhoff
University of Auckland, New Zealand
m.meyerhoff@auckland.ac.nz

Lesley Milroy
University of Michigan, USA
amilroy@umich.edu

Emma Moore
University of Sheffield, England
e.moore@sheffield.ac.uk

Dennis R. Preston
Oklahoma State University, USA
dennis.preston@okstate.edu

Robin Queen
University of Michigan, USA
rqueen@umich.edu

Julie Roberts
University of Vermont, USA
julie.roberts@uvm.edu

Gillian Sankoff
University of Pennsylvania, USA
gsankoff@comcast.net

Natalie Schilling
Georgetown University, USA
ns3@georgetown.edu

Edgar W. Schneider
University of Regensburg, Germany
edgar.schneider@sprachlit.uni-regensburg.de

Anna Strycharz
The Meertens Institute, Amsterdam, The Netherlands
anna.strycharz@meertens.knaw.nl

Sali A. Tagliamonte
University of Toronto, Canada
sali.tagliamonte@utoronto.ca

Erik R. Thomas
North Carolina State University, USA
erthomas@ncsu.edu

Walt Wolfram
North Carolina State University, USA
wolfram@gw.ncsu.edu

Preface to the Second Edition

Publication of the first edition of *The Handbook of Language Variation and Change* in 2002 obviously filled a gap in the field. It was widely adopted as a learning, teaching, and reference tool for researchers and students in sociolinguistics, and it was also well used by scholars in numerous related fields seeking authoritative overviews of central topics and methods on language variation and change. The publication of this second edition, slightly more than a decade later, is necessitated by the continuing vigor of the field. In order to ensure that the *Handbook* remains the authoritative source on this vital approach to language study, we prepared a new edition that reflects the state-of-the-art in sociolinguistic studies.

Our goal remains exactly the same: we see the book as a convenient, hand-held repository of the essential knowledge about the study of language variation and change. We have maintained the core structure, the rationale and the focus that made the first edition so successful. The contributors, now as then, are leading researchers in their fields. About three-quarters of the original chapters have been retained but have been updated to reflect developments and new directions in each topic area. The extent of updating is suggested, perhaps, by the fact that two of the revised chapters have undergone title changes to mark new emphases, and three of the original authors have conscripted co-authors to work with them on new developments.

Seven chapters are entirely new, an appropriate reflection of the continuing vitality of the discipline in the intervening decade. Inevitably, some chapters from the first edition were discontinued in order to accommodate the new directions within manageable space limits. Those discontinued chapters remain valid, incisive treatments of their topics, and we expect that many of them will continue to be cited and referenced in their special areas for years to come.

We have invited the authors of the chapters to discuss the ideas – hypotheses, axioms, premises, probabilities – that drive their branch of the discipline, and to illustrate them with empirical studies, their own or others, that not only demonstrate their applications but also their shortcomings and strengths. We expect that these

areas will continue to attract ingenious researchers and engage curious students and other scholars.

After the "informal epistemology," which immediately follows, the book is organized in eight broad subject areas beginning with data collection (Part I). It proceeds through methods for evaluating data (Part II) and categorizing it (Part III). From there, it moves into the main spheres of social influence including the complexities of time (Part IV), social distance and difference (Part V), and communal interactions, individual identities, and their interrelations (Part VI). The pervasive effect of mobility, both geographical and social, has implications for the social uses of language in diverse contact situations (Part VII). We end the book with Walt Wolfram's forward-looking consideration of the ethical and social roles of sociolinguists in the communities they work in (Part VIII), a topic of increasing engagement among responsible scholars.

The contributors of the chapters make a distinguished international roster. Our invitations went to scholars with recognized expertise, either established or potential, with no thought to anything but their insightfulness and mastery of their research areas. As in the first edition, the final reckoning gives an accidental profile of the culture of sociolinguistics: 26 chapters by 30 scholars, 14 women and 16 men, from six nations. These numbers are all the more striking in the historical context. From its inception in a few rather isolated studies on the Atlantic seaboard of the United States, variationist sociolinguistics has spread globally in a few decades and established its stature inexorably among the language sciences. It is our hope that this new edition of *The Handbook of Language Variation and Change* will aid and abet its spread, as the first edition did, and deepen both the understanding of its goals and the appreciation of its results.

In light of the subject matter of the book, the publishers have acknowledged the diverse backgrounds of the contributors by retaining the mixture of US and UK style conventions across their various chapters.

J.K. Chambers and Natalie Schilling

Studying Language Variation

An Informal Epistemology

J.K. CHAMBERS

Societies can obviously exist without language, as witness the social organizations of carpenter ants, honey bees and great apes. But languages cannot exist without societies. Language is quintessentially social, and throughout recorded history, normal human beings have shown unbounded capabilities for social intercourse, conversational interaction, repartee, self-expression, and tale-telling both real and imagined, all governed by intricate sets of conventions normally beneath consciousness.

Before language existed, our hominoid ancestors organized bands for food-gathering and habitats for sheltering their young; and probably, by analogy with the great apes, not much more. In the absence of language, finding daily sustenance and preventing yourself and your young from becoming sustenance for others are pretty much full-time activities. Since survival and propagation can be achieved in the absence of language, it was obviously not survival and propagation that called language into being. Rather, language is the tool for virtually every human aspiration beyond plain survival and propagation.

Sociolinguistics is the study of the social uses of language, in its many guises. In this chapter, I sketch an informal epistemology of sociolinguistics by outlining its historic development as a linguistic discipline (in Section 1), the persistence of social evaluation in language matters (in Section 2), the place of sociolinguistics among the linguistic sciences (in Section 3), and its relation to communicative competence (in Section 4) and to communicative intelligence (in Section 5).

1 Sociolinguistics as a Discipline

Studying the social uses of language proceeds mainly by observing language use in natural social settings and categorizing the linguistic variants according to their

The Handbook of Language Variation and Change, Second Edition.
Edited by J.K. Chambers and Natalie Schilling.

social distribution. The most productive studies have emanated from determining the social evaluation of linguistic variants. These are also the areas most susceptible to scientific methods such as hypothesis-formulation, logical inference, and statistical testing.

Notwithstanding the pervasive effects of the social milieu on the accents and dialects which are its medium, the study of socially conditioned variation in language is relatively recent. Variationist sociolinguistics became an internationally recognized branch of the linguistic sciences in the 1970s. Its effective beginnings as a movement can be quite specifically traced to the early 1960s, when William Labov presented the first sociolinguistic research report at the annual meeting of the Linguistic Society of America (December 1962) and published "The social motivation of a sound change" (Labov 1963). Those events were not the first public airings of socially relevant linguistic studies, as we shall see, but they were far and away the most influential. Unlike the ones that came before, Labov's initiatives inaugurated a discipline. One reason for their success, though probably not the most important one, was the relative maturity of the sociolinguistic framework that Labov had devised. His analyses introduced three striking innovations into the prevailing linguistic culture: (i) correlating linguistic variants with class, age, sex, and other social attributes, (ii) incorporating style as an independent variable, and (iii) apprehending the progress of linguistic changes in apparent time. All three are hallmarks of the sociolinguistic enterprise to this day.

Labov's success was partly attributable to the simple fact that the time was ripe. Ancillary investigations into the social uses of language, including studies of discourse, pragmatics, interaction rituals, and subjective evaluation tests, sprang into being around the same time.

Labov recalls feeling considerable trepidation as he prepared to present his results in public for the first time. "In those days . . . , you practically addressed the entire profession when you advanced to the podium," he recalled (in 1997). "I had imagined a long and bitter struggle for my ideas, where I would push the social conditioning of language against hopeless odds, and finally win belated recognition as my hair was turning gray. But my romantic imagination was cut short. They ate it up!" The easy reception may have obscured the revolutionary turn that sociolinguistics represents in the history of language study.

Advances in the nascent discipline came quickly. Labov's methods gained breadth and depth with his own work on the social stratification of English in New York City (Labov 1966) and in a large-scale project based at Georgetown University on the inner-city African-American community in Detroit (Shuy, Wolfram and Riley 1968). The theoretical core of the new discipline was bolstered by a perspicacious statement on its empirical foundations (Weinreich, Labov and Herzog 1968), which stands as the manifesto for the enterprise. Sociolinguistics shucked off any hints of anglocentric provincialism with studies of Montreal French (Sankoff and Sankoff 1973) and Panama City Spanish (Cedergren 1973). It also crossed national boundaries with studies in Norwich, England (Trudgill 1974), Edinburgh, Scotland (Reid 1978) and Belfast, Northern Ireland (Milroy and Milroy 1978). Word about these and other developments spread rapidly, months

and sometimes years before the official publications, through conference presentations, dissertations and working papers.

Enthusiasm for the new discipline was undeniably fanned by the revolutionary zeal that went along with overturning some old pieties. Linguistic heterogeneity had been banned in linguistic orthodoxies from Saussure to Chomsky, and so were its correlates such as social attributes, contextual style and apparent-time. Now they were seen as liberating. "The key to a rational conception of language change – indeed of language itself," Weinreich, Labov and Herzog (1968: 100) declared, "is the possibility of describing orderly differentiation in a language serving a community."

Before sociolinguistics gained a foothold in the second half of the twentieth century, there had been a few maverick precursors. The term "sociolinguistics" had been coined a decade before Labov's inaugural presentation by one Haver C. Currie in 1952, in a programmatic commentary on the notion that "social functions and significations of speech factors offer a prolific field for research." With baptismal zeal, Currie (1952: 28) proclaimed, "This field is here designated *sociolinguistics*." Nothing came of Currie's suggestion, though the name stuck. Years later, Labov expressed misgivings about the word itself. In 1972 (xiii), he wrote: "I have resisted the term *sociolinguistics* for many years, since it implies that there can be a successful linguistic theory or practice which is not social." By then, however, it was too late. Non-social linguistics did not disappear, and the term sociolinguistics, like psycholinguistics, neurolinguistics and other academic derivatives, serves its purpose.

Dialectology is sometimes viewed as a precursor of sociolinguistics but the relationship between them is oblique rather than direct. Systematic dialectology goes back to at least 1876 and thus antedates modern linguistics as well as sociolinguistics. Both dialectology and sociolinguistics are in the broadest sense dialectologies (studies of language variation). However, traditional dialectology embraced the strictures of structural linguistics, concentrating on regional speech patterns of mainly rural, old-fashioned speakers elicited one item at a time (Chambers and Trudgill 1998: 13–31). In terms of intellectual history, sociolinguistics can be viewed as a refocusing of traditional dialectology in response to cataclysmic technological and social changes that required (and facilitated) freer data-gathering methods using larger and more representative population samples (Chambers 2002). In its goals as well as it methods, it is a radical departure.

There is now a branch of sociolinguistic dialectology in which region is one independent variable among the other social and stylistic variables (as in the chapter by Britain in this volume). It is much more beholden to sociolinguistics than to dialectology.

Traditional dialect studies with genuine sociolinguistic bearings are very rare. The exception that proves the rule is Louis Gauchat's study of vernaculars in the Swiss village Charmey (1905 [2008]). Gauchat (1866–1942), professor of philology at University of Zurich, visited the alpine village on several occasions and became acutely aware of social stratification in the local dialect. He was also acutely aware that this variability ran counter to the prevailing wisdom at the time, which held

that the dialect of an isolated village with a virtually immobile population should be homogeneous. "If unity can indeed exist in the speech of a village one would expect to find it in Charmey," Gauchat said (1905 [2008: 228]). Instead, he found "variation in the pronunciation, morphology, syntax, and lexicon" (236). No doubt conscious of his renegade mission (though he never came right out and said so), Gauchat sets down the variation in an analysis rich in insight, thorough in detail and sound in argumentation. He emerges, in hindsight, as the patriarch of variationist linguistics (Chambers 2008). Some six decades before Labov, he correlated linguistic variants with sex, age, and social class, recognized style as an independent variable, and apprehended changes in progress with apparent-time comparisons.

Gauchat anticipated many of the postulates of Weinreich *et al*. (1968) as well as sociolinguistic methods. "Variation in pronunciation among members of a single speech community has not been studied systematically," he says (1905 [2008: 227]), "despite its potential contribution to our understanding of language change." He hoped his study might foment a sociolinguistic revolution, although he was too genteel to put it that way. What he did say was, "My reason . . . for making public these opinions I have formed on the unity of speech in a single community is to encourage dialectologists to undertake similar research in other places." It did not happen. Gauchat's work in Charmey was regarded as eccentric, and no one, not even Gauchat himself in the 30 years of his career that remained, saw fit to follow his lead.

Gauchat was clearly too far ahead of his time. The emergence of the international movement for socially perspicacious linguistic studies was in abeyance for six more decades.

2 Language as a Social Phenomenon

The late-blooming history of sociolinguistics appears paradoxical in view of the obvious social role of language. All societies tolerate and even foster social judgments of language use, and typically integrate them into the communal ethos, most conspicuously in developed nations where they become part of the institutional mandate of schools, government offices and professional societies. So persistent and pervasive are the social judgments of language use that they must be embedded in human nature, perhaps as an adjunct of human communicative competence (discussed below). They have been documented from the beginning of the written record. Thus Sirach, the Old Testament moralist, declared: "When a sieve is shaken, the rubbish is left behind; so too the defects of a person appear in speech. As the kiln tests the work of the potter, so the test of a person is conversation" (Ecclesiasticus 27: 4–5). And Cicero, in 55 BC, enjoins his readers to "learn to avoid not only the asperity of rustic pronunciation but the strangeness of outlandish [that is, regional] pronunciation" (*De Oratore* III, 12).

Value judgments like these, both ancient and modern, have purely social motivation. Linguistically, they are vacuous. This is readily demonstrated by comparing any pair of linguistic variants, such as these grammatical variants:

We don't expect any help from the government.
We don't expect no help from the government.

The two sentences differ in that the second one includes a negative marker on the object noun phrase (*no help*) as well as on the verb phrase (*don't expect*), whereas the first one avoids the double negative by replacing the noun negator with *any*. Notwithstanding this difference, the two sentences convey exactly the same grammatical meaning and everyone who speaks English with even minimal competence recognizes their semantic identity.

The sentences do, however, convey very different social meanings as a direct consequence of the grammatical difference. That is, they carry *socio*linguistic significance. The first, with its standard forms, is emblematic of middle-class or educated speech, while the second is emblematic of working-class or uneducated speech. These differences will also be readily recognized by virtually every speaker of the language.

The perceived superiority of the first sentence is obviously not linguistic, since the two sentences convey exactly the same meaning. Nor is it historical, since the second sentence, the nonstandard double negative, was in fact the standard construction until the fifteenth century, when *any*-suppletion came into the grammar as a competing form and ultimately prevailed as the preferred form. However, recognizing that social evaluations of sentences like these are arbitrary and conventional does not mean that they are inconsequential. On the contrary, people whose speech is judged adversely can suffer socially, occupationally and educationally (as discussed by Preston in this volume).

Because social judgments of linguistic forms have such a continuous and intimate relation to the human condition, it would be natural to expect a fairly long history of inquiry into the sources, functions and significations of language in its social context. Instead, as we have seen, it is relatively recent. Perhaps the social role of language was too commonplace to attract serious inquiry, but more likely it is so integral in language as to escape notice. The classical Greeks missed it entirely. Plato and Aristotle concerned themselves with categorizing linguistic forms, that is, with grammar in the sense discussed in the next section. Neither of them noticed linguistic variation of any kind, and their overwhelming influence on Western thought undoubtedly contributed to the antisocial bias of Western linguistic tradition. According to Kiparsky (1979), the Sanskrit grammarian Pāṇini. (ca. 600 BC) did recognize systematic variability, which he called *anyatarasyām*, but his distinction was trivialized by his successors as meaning "marginal" or "unacceptable," for which Pāṇini had actually used different terms. Pāṇini's followers missed the distinction, and as a result Pāṇini's insight had no impact on tradition.

The only classical scholar who seems to have been aware of the social side of language is the Roman polymath Varro (116–27 BC), who not only recognized linguistic variation (*anomalia*) but also linked it to vernacular language use (*consuetudo*; see Taylor 1975). Varro observed, among other things, the arbitrary nature of linguistic judgments. "The usage of speech is always shifting its position," he

wrote (IX, 17; Kent 1938: 453). "This is why words of the better sort (i.e. morphologically regular forms) are wont to become worse, and worse words better; words spoken wrongly by some of the old-timers are . . . now spoken correctly, and some that were then spoken according to logical theory are now spoken wrongly." One of Varro's insights – *consuetudo loquendi est in motu* – could be emblazoned as the motto of sociolinguistics: "the vernacular is always in motion." Unfortunately, Varro's linguistic treatise, which survives only as a fragment, gave rise to no school of thought. He remains an isolated figure in the history of language study.

Enlightenment authors presupposed the social basis of language. Locke, in *An Essay Concerning Human Understanding* (1690: 101), wrote: "God, having designed man for a sociable creature, made him not only with an inclination, and under a necessity to have fellowship with those of his kind, but furnished him also with language, which was to be the great instrument and common tie of society." But the social uses of the instrument, under the presumption that it was God-given, were apparently deemed to be beyond human scrutiny.

Similarly, twentieth-century linguists dutifully enshrined the social function in their definitions. "Language is defined as the learned system of arbitrary vocal symbols by means of which human beings, as members of a society, interact and communicate in terms of their culture," according to one introductory textbook (Trager 1972: 7). Bloomfield (1933: 42) said, "All the so-called higher activities of man – our specifically human activities – spring from the close adjustment among individuals which we call society, and this adjustment, in turn, is based upon language; the speech-community, therefore, is the most important kind of social group." Firth (1937: 153) said, "speech is social 'magic'. You learn your languages in stages as conditions of gradual incorporation into your social organization. . . . The approach to speech must consequently be sociological."

Yet neither Bloomfield nor Firth nor any of the linguists who shared their structuralist concepts directly studied the social uses of language. Until the advent of sociolinguistics, there were no concentrated attempts at discovering the social significance of linguistic variation. That may be partly explicable in terms of intellectual history. All the social sciences are relatively young. Psychology, sociology, economics, and anthropology had their effective beginnings around the turn of the twentieth century, whereas subject areas less intimately involved with the human condition such as algebra, physics and zoology have ancient origins. Sociolinguistics, as the social-science branch of linguistics (along with developmental psycholinguistics), is a newcomer compared to the branch known as theoretical linguistics, which descends from venerable studies of grammar, rhetoric, and philology.

3 Linguistics and Sociolinguistics

In the development of modern linguistics, the shunting aside of the social significance of language was neither an oversight nor an accident. Saussure, the founder of modern linguistics, noted that "speech has both an individual and a social side,

and we cannot conceive of one without the other" (1916: 8). Inconceivable though it may have been for him to separate the individual and the social aspects, Saussure nevertheless advocated the study of the former without the latter. His famous distinction between *langue*, the grammatical system, and *parole*, the social uses of language, came into being expressly to demarcate what he considered the proper domain of linguistic study:

> But what is *langue*? It is not to be confused with human speech [*parole*], of which it is only a definite part, though certainly an essential one. It [*parole*] is both a social product of the faculty of speech and a collection of necessary conventions that have been adopted by a social body to permit individuals to exercise that faculty. Taken as a whole, speech is many-sided and heterogeneous; straddling several areas simultaneously – physical, physiological, and psychological – it belongs to both the individual and to society; we cannot put it into any category of human facts, for we cannot discover its unity.
>
> Language [*langue*], on the contrary, is a self-contained whole and a principle of classification. As soon as we give language first place among the facts of speech, we introduce a natural order into a mass that lends itself to no other classification. (1916: 9)

Saussure's dismissal of a possible science of *parole* seems curmudgeonly, with hindsight, but he was not alone. Before him, Humboldt had made a similar distinction between what he called a formless *ergon* and a well-formed *energeia*. *Ergon* (or *parole*) was "divided up into an infinity as the sole language in one and the same nation," and *energeia* (or *langue*) was language in the abstract sense, with "these many variants . . . united into one language having a definite character" (1836: 129). After Saussure, Chomsky made a similar distinction between competence, "the speaker-hearer's knowledge of his language," and performance, "the actual use of language in concrete situations," and he went on to say that "observed use of language . . . surely cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline" (1965: 4). With hindsight, Chomsky's dismissal seems not so much curmudgeonly, like Saussure's, as myopic.

Humboldt, Saussure, and Chomsky were obviously right in pointing out that speech, *parole*, is heterogeneous, but they have been proven wrong in dismissing heterogeneity as a viable object of study. From the beginning, the challenge facing sociolinguistics, the science of *parole*, has been to arrive at an understanding of language as, in Weinreich, Labov, and Herzog's phrase, "an object possessing orderly heterogeneity" (1968: 100).

4 Communicative Competence and the Language Faculty

Studying language as *langue* (or *energeia* or competence), as distinct from *parole* (or *ergon* or performance), requires abstracting linguistic data from the real-world variability in which it naturally occurs. Grammarians impose a hypothetical filter

on natural language data to make it invariant, discrete, and qualitative. The filter, called the axiom of categoricity (Chambers 2009: 26–28), has been described in numerous ways. Hjelmslev (1961: 5–6) states it this way: "Linguistics must attempt to grasp language, not as a conglomerate of non-linguistic (e.g. physical, physiological, psychological, logical, sociological) phenomena, but as a self-sufficient totality, a structure *sui generis*." Joos (1950: 701) declared: "We must make our 'linguistics' a kind of mathematics within which inconsistency is by definition impossible."

By contrast, sociolinguists attempt to grasp language as it is used in social situations, which is to say as variant, continuous, and quantitative. *Langue* and *parole* remain useful distinctions today for a reason that Saussure would undoubtedly have found unimaginable, because they now help to define the different objects of inquiry of theoretical linguistics and sociolinguistics. They are separable in theory as natural partitions of the language faculty, or what might plausibly be considered distinct cognitive systems.

Chomsky has put forward the conception of the language faculty as interacting systems conceived, in his words, as "'mental organs' analogous to the heart or the visual system or the system of motor coordination and planning" (1980: 39). Theoretical linguists who adopt the axiom of categoricity are primarily interested in discovering the properties of one of those systems of the language faculty, called GRAMMAR, conceived as a language-specific bioprogram (to use Bickerton's incisive term: 1984, 2008: 110–113). The GRAMMAR is also known as "I-language, 'I' for internal and individual" (Lightfoot 2006: 7), presumably to avoid the ambiguity with "grammar" as a book of rules or language-learner's manual. The internal grammar is "a person's language organ, the system" (Lightfoot 2006: 7). It is made up of, in Chomsky's terms (1980: 55), "a system of 'computational' rules and representations." Attempts at discovering its innate computational properties have led Chomsky and his followers into minute examinations of surface-structure puzzles involving linguistic coreference, scope, and other structural intricacies. They have produced insights into the grammatical processor as "structure-dependent" rather than strictly linear (cf. Hurford 2008: 526) and, in Chomsky's tenacious but disputed stance, as language-specific, not reducible to other, independently motivated, non-language-processing cognitive components.

The GRAMMAR is the module in the language faculty that accounts for the uniquely human attributes of creativity in language production and comprehension, and for the rapidity of language acquisition in infancy. However, it is obviously not autonomous. Linguistic production and comprehension require real-world orientation to express meanings, and the acquisition device requires the stimulus of social interaction to activate learning. Chomsky recognizes the existence of other systems, and he has isolated two of them as follows: "A fuller account of knowledge of language will consider the interactions of grammar and other systems, specifically the system of conceptual structures and pragmatic competence, and perhaps others" (1980: 92). The component that involves real-world orientation Chomsky calls the CONCEPTUAL SYSTEM, and the social stimulus

has its source in what Chomsky calls "pragmatic competence" but is generally called COMMUNICATIVE COMPETENCE.

By the CONCEPTUAL SYSTEM, Chomsky means "the system of object-reference and also such relations as 'agent', 'goal', 'instrument' and the like; what are sometimes called 'thematic relations'" (1980: 54). "Object-reference" includes vocabulary items, the "massive inventory of form-to-meaning mappings" (Hurford 2008: 526) which are the most obvious intermediaries between grammar and the world. The conceptual system reveals uniquely human properties most easily discerned in acquisition. Children master fine semantic distinctions of the sort found in verbs such as *follow* and *chase* relatively early, certainly long before they can consciously define what they mean. These fine vocabulary distinctions recur in all natural languages. One way of explaining this mastery, Chomsky (1988: 31) says, is by postulating that words "enter into systematic structures based on certain elementary recurrent notions and principles of combination." More generally, he says, "The rate of vocabulary acquisition is so high at certain stages of life, and the precision and delicacy of the concepts acquired so remarkable, that it seems necessary to conclude that in some manner the conceptual system with which lexical items are connected is already substantially in place" (1980: 139).

Chomsky's third language module, "pragmatic competence," pertains to, in his words, "knowledge of conditions and manner of appropriate use, in conformity with various purposes. . . . We might say that pragmatic competence places language in the institutional setting of its use, relating intentions and purposes to the linguistic means at hand" (1980: 224–225). This notion has a familiar ring to sociolinguists. It was influentially described by Hymes as "sociolinguistic competence" or COMMUNICATIVE COMPETENCE, as follows:

> Within the social matrix in which [a child] acquires a system of grammar, a child acquires also a system of its use, regarding persons, places, purposes, other modes of communication, etc. – all the components of communicative events, together with attitudes and beliefs regarding them. There also develop patterns of the sequential use of language in conversation, address, standard routines, and the like. In such acquisition resides the child's sociolinguistic competence (or, more broadly, communicative competence), its ability to participate in its society as not only a speaking, but also a communicating member. (1974: 75)

Hymes adds, "What children so acquire, an integrated theory of sociolinguistic description must be able to describe."

Like the other organs of the language faculty, COMMUNICATIVE COMPETENCE develops early and rapidly in normal children with little or no tutoring. Since most of the conventions governing communicative events are beneath consciousness, explicit teaching is impossible in any case. Evidence for COMMUNICATIVE COMPETENCE as an entity independent of GRAMMATICAL COMPETENCE (and presumably the other organs of the language faculty) can be found, for instance, in pathologies in which people are forced to function with one in the absence of the other.

The independence (or modularity) of COMMUNICATIVE COMPETENCE is revealed in certain neurological disorders in which it is disturbed and disrupted. People suffering from what is called "semantic-pragmatic disorder" tend to interrupt the conversational flow with inappropriate or ill-timed assertions, fail to follow topics, introduce what appear to be digressions or non-sequiturs, and speak out of turn (Mogford-Bevan and Sadler 1991). Typically, their speech is phonologically and grammatically well-formed, and not infrequently their speech is remarkably fluent. Clinical researchers usually rely on standardized tests as diagnostic tools, but people with semantic-pragmatic disorder tend to score within normal ranges because of their grammatical fluency. As a result, descriptions of semantic-pragmatic disorder in the psycholinguistic literature often appear to be cursory and vague.

Recognizing it as a sociolinguistic disorder might persuade clinicians to use sociolinguistic observation and analysis for its description. In any event, what malfunctions in the people who are afflicted with the disorder is their communicative competence. They speak grammatically but they cannot carry a conversation. Just as myxedema proves the existence of the thyroid gland in the endocrine system (if proof were needed), so semantic-pragmatic disorders prove the existence of communicative competence as an autonomous module in the language faculty.

5 Interdependence of Language and Communication

Though communicative competence was admitted fairly late into the Chomskyan conception, it has taken on an increasingly important role in conceptualizations of the language faculty. As Brooks and Ragin (2008: 514) point out, "Language is not merely the product of a language-ready brain; it is a cultural product of a community of practitioners."

Language and its social context are inseparable, but for Chomsky it does not follow that they are interdependent. "It is true, a virtual tautology, that the study of communication takes into account social context," he says (2011: 266). "It is also uncontroversial that the study of the mechanisms that we put to use [in GRAMMAR] typically ignores social context, and quite rightly so." Communication, in Chomsky's schema, is the "externalization by the SM [sensory-motor] system" and, in his view, it "appears to be a secondary property of language" (2011: 275). The language faculty developed many millennia after the sensory-motor system, he claims, and "language use is only one of many forms of communication" (2011: 275). In Chomsky's view, language use seems to be no more important in communication than are gestures, facial expressions, and eye-gaze cues (assuming he would include these sensory-motor reflexes among the "many forms of communication"), and in his conception none of these impinges upon or affects in any way the innate, unalterable language faculty.

Alternative conceptions suggest that he is wrong on both counts. The growing consensus finds grounds for interdependence of grammar and communication. Primate-like sensory-motor systems undeniably pre-dated language, but their uniquely human adaptation to accommodate speech must have occurred more or less simultaneously with the development of the language faculty. Evolution of grammatical competence devoid of communicative competence is inconceivable.

Chomsky's insistence on autonomy rests largely on claims about the poverty of stimulus – claims that primary linguistic data are not rich enough for children to learn grammar, and therefore require "information that is available to children independently of experience, represented in some fashion in the genetic material" (Lightfoot 2006: 9). Arguments for the poverty of stimulus are largely theoretical rather than empirical, consisting of, for instance, assertions that English speakers never violate the Empty Category Principle (i.e., never utter sentences like *How many mechanics did he wonder whether fixed the cars?*) even though they never hear sentences that would allow them to infer the existence of the principle (Chomsky 2011: 265). The principle must therefore be genetically encoded in the language faculty.

Because the stimulus plays such a crucial role in the Chomskyan conception, other linguists are trying to come to grips with what exactly the stimulus consists of. The best of these inquiries are amassing evidence that increasingly demystifies Chomsky's assumptions. Gleitman *et al*. (2007: 566), for instance, show that "young learners are quite adept at taking visual perspective in object labeling tasks; by the time they're 18 months old, young children will inspect a speaker's attentional state upon hearing a novel label." The astounding proliferation of lexical items by toddlers is thus partly accounted for by their hitherto undiscovered sensitivity to visual cues, and not only for concrete, visible objects. Gleitman *et al*. say, "Adults may be using this information [gaze cues] rapidly and expediently to arrive at increased communicative alignment, and children may be able to utilize the caretaker's gaze direction patterns in complex language-learning tasks such as verb learning and syntactic interpretation" (2007: 566).

These results also demonstrate that, contra Chomsky, communication modes seem to be hierarchic, with nonverbal, primitive eye gaze deployed in the service of language development. They also appear to demonstrate interdependence between language and communication. Gleitman *et al*. (2007: 566) say, "If we are right, the unconscious, rapid, and incremental speech machinery is not wholly or even predominantly conception first and speech only thereafter . . . ; rather, the representations constructed by the visual-attentive and linguistic-conceptual systems may be integrated all along the line."

None of this impugns in any way the innate human endowment, the "language-ready brain," that enables rapid acquisition and communicative creativity. Innateness is, as Chomsky has insisted from the start, a truism, a given. What it calls into question is Chomsky's antisocial conceptualization of it. In many linguistic fields, we are coming to grips with what it is that is essentially human; one of the common thrusts seems to be that it is more sociable than we formerly believed.

The cognitive infrastructure includes "communicative intelligence," described by De Ruiter and Levinson (2008: 518) as a "specific type of intelligence . . . to encode and decode the communicative intentions behind any type of potentially communicative behavior, linguistic, nonverbal or otherwise." Communicative intelligence is intuitively convincing, seemingly part of the common experience of every normal human being, and it is gaining credence in several experimental paradigms. Something like communicative intelligence is presumably implicated in what Gleitman *et al*. call "increased communicative alignment," that facilitates extrapolation from concrete objects to more abstract language-learning tasks. Galantucci (2005) showed that paired subjects who are under pressure in problem-solving tasks quickly develop communication systems in the hitherto unfamiliar modes at their disposal. Noordzij *et al*. (2009) show that subjects improvising communication strategies in spontaneous problem-solving tasks show activation in the same well-defined brain region, stimulating "inferential processes" used in language processing but evidently not exclusive to it.

Communicative intelligence is rooted in socialization and is the common cognitive inheritance of all human beings. "Languages . . . are created and filtered by brains that are biologically endowed with communicative intelligence," according to De Ruiter and Levinson (2008: 518). "Together with the vocal/auditory apparatus, this cognitive adaptation for communication makes possible the cultural evolution of spoken languages. . . . Without such specialized structures, the speed and flexibility with which language (in multiple modalities) is used, learned and changed, even within one generation, would not be possible."

The social basis of language was admitted into discussions of the language faculty belatedly, as its study was also admitted belatedly into the discipline of linguistics. In both instances, its importance was quickly recognized. Ignoring it now, in either domain, would be unthinkable.

6 The Sociolinguistic Enterprise

For social interaction to work, both the content of speech and its form must be suited to the speakers and their interlocutors in a particular social context. Sociolinguistic analysis has revealed that our main resources come from modulating linguistic elements in subtle (and clearly unteachable) ways, selecting, so to speak, a particular vowel variant with a certain frequency in a particular situation or a past tense variant or other structural variant in appropriate contexts.

The variants we choose with such casual virtuosity range along a continuum from standard to nonstandard and stigmatized. No linguistic principle can explain the social evaluation attached to any of them. As Varro observed two millennia ago, "words of the better sort," that is, morphologically regular forms, are sometimes the socially stigmatized forms. There is also no linguistic principle behind their distribution in the speech of different social groups in the community, or the relative frequency of their use from one generation to the next.

It is these aspects that underlie the age-old mystery of language change, which is irrepressible and inexorable in spite of the fact that it is, in a common-sense view, both dysfunctional and otiose – dysfunctional in so far as it impedes communication in the long run, and otiose in so far as the changes neither improve nor degrade the language as a communicative medium. The root causes seem to be nothing more profound than social convention.

Variation is socially motivated, and pinpointing the motivations and giving them empirical substance remains perhaps our greatest challenge. We are gaining an understanding of human communicative competence. Every chapter of this book provides evidence, in its own way, of how people respond to social evaluations of their speech, which are always shifting, usually tediously but sometimes rapidly, and almost always tacitly. *Consuetudo loquendi est in motu*.

The wonder of it is that it attracted virtually no conscious investigation for centuries and indeed millennia – much longer, for instance, than metaphysical speculations about free will or grammatical taxonomies of verb conjugations. It is surely a measure of how deeply ingrained our communicative competence is in all our activities that it could lay hidden so long from consciousness, and a measure as well of how deeply embedded it is in our human nature.

REFERENCES

Bickerton, Derek (1984) The language bioprogram hypothesis. *The Behavioral and Brain Sciences* 7: 173–187.

Bickerton, Derek (2008) *Bastard Tongues*. New York: Hill and Wang.

Bloomfield, Leonard (1933) *Language*. New York: Holt, Rinehart and Winston.

Brooks, Patricia J. and Ragin, Sonia (2008) Prolonged plasticity: necessary and sufficient for language-ready brains. *Behavioral and Brain Sciences* 31: 514.

Cedergren, Henrietta (1973) The interplay of social and linguistic factors in Panama. PhD dissertation, Cornell University.

Chambers, J.K. (2002) Dialectology. In Neil J. Smelser and Paul B. Baltes (eds.), *International Encyclopedia of the Social and Behavioral Sciences*. Amsterdam: Elsevier Science.

Chambers, J.K. (2008) Louis Gauchat, Proto-Variationist. *Historiographica Linguistica* 35: 215–225.

Chambers, J.K. (2009). *Sociolinguistic Theory: Linguistic Variation and Its Social Significance*. 3rd edn. Malden, US, and Oxford, UK: Wiley-Blackwell.

Chambers, J.K. and Trudgill, Peter (1998) *Dialectology*. 2nd edn. Cambridge, UK: Cambridge University Press.

Chomsky, Noam (1965) *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam (1980) *Rules and Representations*. New York: Columbia University Press.

Chomsky, Noam (1988) *Language and Problems of Knowledge: The Managua Lectures*. Cambridge, MA: MIT Press.

Chomsky, Noam (2011) Language and other cognitive systems: What is special about language? *Language Learning and Development* 7: 263–278.

Currie, Haver C. (1952) A projection of socio-linguistics: the relationship of speech and its social status. *The Southern Speech Journal* 18: 28–37.

De Ruiter, Jan Peter and Levinson, Stephen C. (2008) A biological infrastructure for communication underlies the cultural evolution of languages. *Behavioral and Brain Sciences* 31: 518.

Firth, J.R. (1937) *The Tongues of Men*. London: Watts.

Galantucci, B. (2005) An experimental study of the emergence of human communication systems. *Cognitive Science* 25: 737–767.

Gauchat, Louis (1905 [2008]) Phonetic unity in the dialect of a single village. Sarah Cummins (trans.). *Historiographica Linguistica* 35: 227–274. Original: L'unité phonetique dans le patois d'une commune. *Festschrift Heinreich Morf: Aus Romanischen Sprachen und Literaturen*. 175–232. Halle: M. Niemeyer.

Gleitman, Lila R., January, David, Nappa, Rebecca, and Trueswell, John C. (2007) On the *give* and *take* between event apprehension and utterance formulation. *Journal of Memory and Language* 57: 544–569.

Hjelmslev, Louis (1961) *Prolegomena to a Theory of Language*. Francis J. Whitfield (trans.). Madison: University of Wisconsin Press.

Humboldt, Wilhelm von (1836 [1972]) *Linguistic Variability and Intellectual Development*. George C. Buck and Frithjof A. Raven (trans.). Philadelphia: University of Pennsylvania Press.

Hurford, James R. (2008) Niche-construction, co-evolution, and domain-specificity. *Behavioral and Brain Sciences* 31: 526.

Hymes, Dell (1974) *Foundations in Sociolinguistics: An ethnographic approach*. Philadelphia: University of Pennsylvania Press.

Joos, Martin (1950) Description of language design. *Journal of the Acoustical Society of America* 22: 701–708.

Kent, Roland G. (1938) *Varro on the Latin Language with an English Translation*, vol. 2. London: William Heinemann; Cambridge, MA: Harvard University Press.

Kiparsky, Paul (1979) *Pāṇini as a Variationist*. Cambridge, MA: MIT Press.

Labov, William (1963) The social motivation of a sound change. *Word* 19: 273–309.

Labov, William (1966) *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Labov, William (1972) *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William (1997). How I got into linguistics, and what I got out of it. Undergraduate address 1987, revised. http://www.ling.upenn.edu/~labov/Papers (accessed February 4, 2013).

Lightfoot, David (2006) *How New Languages Emerge*. Cambridge, UK: Cambridge University Press.

Locke, John (1690) *An Essay Concerning Human Understanding*. New York: Dolphin Books.

Milroy, James and Milroy, Lesley (1978) Belfast: change and variation in an urban vernacular. In Peter Trudgill (ed.), *Sociolinguistic Patterns in British English*. 19–36. London: Edward Arnold.

Mogford-Bevan, K. and Sadler, J. (1991) *Child Language Disability*, vol. II: *Semantic and Pragmatic Difficulties*. Clevedon and Philadelphia: Multilingual Matters.

Noordzij, Matthijs, Newman-Norlund, Sarah E., De Ruiter, Jan Peter *et al*. (2009) Brain mechanisms underlying human communication. In Jennifer S. Beer (ed.), *Frontiers in Human Neuroscience*, 3: 1–13.

Reid, Euan (1978) Social and stylistic variation in the speech of children: some evidence from Edinburgh. In Peter Trudgill (ed.), *Sociolinguistic Patterns in British English*. 158–171. London: Edward Arnold.

Sankoff, David and Sankoff, Gillian (1973) Sample survey methods and computer assisted analysis in the study of grammatical variation. In Regna Darnell (ed.), *Canadian Languages in the Social*

Context. 7–64. Edmonton: Linguistic Research Inc.

Saussure, Ferdinand de (1916) *Course in General Linguistics*. Charles Bally and Albert Sechehaye (eds.), Wade Baskin (trans.). New York and Toronto: McGraw-Hill.

Shuy, Roger, Wolfram, Walter A., and Riley, William K. (1968) *Field Techniques in an Urban Language Study*. Washington, DC: Center for Applied Linguistics.

Taylor, Daniel J. (1975) *Declinatio: A Study of the Linguistic Theory of Marcus Terentius Varro*. Amsterdam: John Benjamins.

Trager, George L. (1972) *Language and Languages*. San Francisco, CA: Chandler.

Trudgill, Peter (1974) *The Social Differentiation of English in Norwich*. Cambridge, UK: Cambridge University Press.

Weinreich, Uriel, Labov, William, and Herzog, Marvin I. (1968) Empirical foundations for a theory of language change. In Winfred P. Lehmann and Yakov Malkiel (eds.), *Directions for Historical Linguistics: A Symposium*. 95–188. Austin TX: University of Texas Press.

Part I Data Collection

1 Entering the Community

Fieldwork

CRAWFORD FEAGIN

While the ultimate goal of sociolinguistic research is to resolve questions of linguistic importance, such as how language change comes about, nothing of that sort can be accomplished without first entering a community in order to collect data which will help provide the basis for any such answers. The central problem in collecting sociolinguistic data has been described by Labov as the Observer's Paradox: "our goal is to observe the way people use language when they are not being observed" (1972a: 61). Sociolinguistic fieldwork of all kinds, whether recorded interviews, participant observations or street-corner quizzes, must be geared to overcome this problem. In this chapter, I consider several well-established methods. I begin with a section on "Planning the Project," dealing with preliminary considerations for designing and conducting a sociolinguistic survey. The heart of the chapter, as indeed of field research, is the second section on the "Sociolinguistic Interview," the Labovian protocol for selecting informants and eliciting different styles of speech. I then consider some other elicitation methods used in sociolinguistics: participant observation and rapid and anonymous observations. While telephone surveys have been fruitful in the past (see Labov *et al*. 2006; Ash 2000), today they have limited use because of the general shift to cell phones, eliminating the use of area codes or telephone directories in identifying likely participants. Long-distance surveys today can utilize internet phone services such as Skype or other internet technologies (e.g. web-based surveys for gathering self-reports of linguistic production and/or information on linguistic perceptions and attitudes; see Schilling 2013); they will not be discussed here. Instead, I will focus on face-to-face methods.

The Handbook of Language Variation and Change, Second Edition.
Edited by J.K. Chambers and Natalie Schilling.

1 Planning the Project

Although the methods involved are presented here as if they were sequential, in practice the various phases of fieldwork and other aspects of research are cyclical, or perhaps spiral. Investigation in one area will influence what can be done in another. An interview might provide insights about the community that can be incorporated into the protocol and produce a much better interview with subsequent informants. For instance, in my work in Anniston, Alabama (Feagin 1979), one teenager mentioned a recent snowstorm, an unexpected and exciting phenomenon in that part of the world, and so in later interviews I asked the rest of the teenagers about it. As a result, I came away with excited accounts of sledding on garbage-can tops and cookie sheets, wearing improvised boots made from plastic bags, and skidding dangerously over slippery roads. My interview protocol for the older people already included questions about a tornado that had hit Anniston 20 years before; the snowstorm provided similarly dramatic stories from an incident in the recent memories of the teenagers.

Similarly, sometimes in the course of an interview, investigators might discover an unexpected grammatical form or phonological realization. They must be attentive and flexible in order to pursue the newly discovered linguistic feature for that community.

As an aid to planning, a small-scale pilot project along the general lines of the main research will indicate more precisely what might be feasible goals and procedures. A larger consideration is that collecting data is only an intermediate goal. The ultimate goal is linguistic.

The hypothesis that motivates the project will influence how to go about collecting the data. Again, in my own work in Anniston, I hypothesized that over the three-and-a-half centuries of close contact, African-American speech would have influenced European-American grammar in the South. I therefore set out to elicit data from the white community that was parallel to Labov's African-American data from Harlem (Labov *et al*. 1968; Labov 1972b). Even though it turned out in large part that my hypothesis was not correct, nonetheless it was important to try to get parallel data so that a comparison would be possible.

An important guideline for fieldworkers at the planning stage is that a close analysis of a small amount of data is better than an unfinished grandiose project. With that in mind, I concentrated on the extreme generations (teenagers and grandparents) and extreme social classes (local working class and upper class), and the older rural working class (with no younger counterpart). More than that I could not handle, though ideally I would have liked to include the middle class and the middle aged, not to mention the local African-American community. However, examining only the two urban classes plus the older rural working class, using adolescents and grandparents in the city and elderly people from the country, and keeping the sample balanced in terms of gender, I was able to see change progressing through the community.

A rule of thumb in disciplines that require fieldwork is that one third of the project time will be spent in fieldwork, one third in analysis, and the final third

in writing up the work. Though far from scientific, this rule provides an effective reminder of the point that time required for analysis and writing increases in a ratio of about 2:1 for each hour of data elicitation.

Competent fieldworkers have included a wide range of personality types. Because fieldwork requires face-to-face interaction, it is usually assumed that gregarious persons do best, and it seems likely that they would have an advantage, at least in getting started. Shy people might find this sort of work excruciating, especially in the beginning. However, shy people have sometimes proven highly successful in conducting interviews and obtaining data, for the simple reason that people often open up when talking to quiet people, perhaps because they find them unthreatening and perhaps because the lack of interruptions encourages them to speak at length (Schilling, personal communication).

1.1 Library research

Once the community has been selected for research, the next step is to get a perspective on the community itself – linguistic, demographic, and historical. Information on local speech, major industries, labor, religious institutions, communications, movement of peoples, and the historical development of the area can aid in understanding local society.

A survey of previous linguistic work must be carried out, both on the linguistic aspects you intend to study and on any previous research concerning the local language variety. Earlier work on the local variety, regardless of its quality, can be useful for time depth or for pinpointing interesting problems.

First-hand accounts of fieldwork can be found in Labov (1966), Feagin (1979), Milroy (1980), Dayton (1996), and Eckert (2000) for linguistics, and in Whyte (1943, 1984) and Liebow (1967) for ethnography. Such personal accounts are rarely published, but dissertations often include them in chapters on methodology. More general discussions may be found in Labov (1972a, 1984), Wolfram and Fasold (1974), Milroy (1987), Romaine (1980), Baugh (1993), Milroy and Gordon (2003), Di Paolo and Yaeger-Dror (2011), and Schilling (2013). For sociolinguistic fieldwork in non-Western societies where the investigator is clearly an outsider, see Albó (1970), Harvey (1992), Wald (1973), and Bowern (2008). Obviously, a different set of problems arises when the fieldworker is a foreigner, of different ethnicity, and not a native speaker of the language. While addressed to researchers doing basic linguistic fieldwork (rather than sociolinguistic research) in non-Western languages (frequently in remote areas), Samarin (1967) provides an overview of linguistic fieldwork, though now somewhat dated. Bowern (2008) is a more recent resource.

1.2 Ethnography

Along with gathering linguistic data, it is important to study the community itself in situ. While material collected from library research must not be overlooked if it is available, the researcher in the field must begin by observing the physical layout of the place, who lives where, who associates with whom, and in what situations

particular people associate with each other. While this type of research can be seen in Fischer (1958) and more elaborately in Labov (1963), subsequent studies have become more sophisticated and more detailed, culminating in Eckert's intricate study of a suburban Detroit high school (Eckert 2000). It is through a thorough knowledge of both the structure and dynamics of the local community that the patterning and social meanings of language variation and change in the speech community can be fully understood. While some linguists have criticized sociocultural investigations as outside the competence of linguists who are not specialists in sociology or anthropology (Bailey 1996), the only way some aspects of language behavior can be understood and analyzed is through such an undertaking.

It was through such a study that Labov was able to show that younger people on the island of Martha's Vineyard who had decided to remain on the island after their high school years were picking up the fishermen's pronunciation of (ay) and (aw), regardless of their social class, while those who had decided to leave the island for further education and employment were shifting toward mainland speech norms (Labov 1963). Similarly, Eckert (2000) was able to show that the social division between "jocks" (middle class) and "burnouts" (working class) in suburban high schools played a role in transmitting urban Detroit features into suburban teenage speech. See Eckert (2000: Chapter 3) for a valuable account of the process of studying the ethnography of a community.

1.3 Linguistic variables

In a quantitative study of linguistic variation, acquaintance with previous work and perhaps a pilot study should help to narrow the focus of the project. In practical terms, however, this does not always take place right at the beginning. What needs to be isolated before analysis can begin, and preferably before data-gathering begins, is a selection of linguistic variables to be studied. As with fieldwork more generally, though, the process is iterative, and it may turn out that the variables one originally sets out to study are not of great sociolinguistic interest, and more important features may be revealed as fieldwork progresses.

The linguistic variable, a concept originating with Labov (1963, 1966), is a linguistic entity which varies according to social parameters (age, sex, social class, ethnicity), stylistic parameters (casual, careful, formal), and/or linguistic parameters (segmental, suprasegmental). Usually the social and stylistic variation will be coordinated in some way, so that the casual speech of an accountant will be similar to the formal speech of a plumber – though that remains to be seen in the course of the investigation.

The linguistic variable can be found at all linguistic levels: most common are phonological, such as, for example, (r) might be realized as [ɹ] or as [ə] in a community which has been r-less and is becoming r-ful; morphophonological as in (ing), the English present participle marker which has two common pronunciations, standard [ɪŋ] and casual [ɪn]; morphological as in the realization of the past tense form of *dive* either as *dived* or as *dove*; syntactic as in the realization of negated *be* variously as *ain't, isn't, 's not, is not*; or lexical as in the use of either *hero* or

grinder as the word to designate a particular kind of sandwich. The most frequently studied variables are phonological and morphological.

The main criterion for determining the set of variants of a single variable is that the referential meaning must be unchanged regardless of which variant occurs. (This can present a problem when dealing with grammar, as pointed out by Lavandera (1978) and Romaine (1981).) The selection of one variant from the set will generally be motivated by either social or stylistic considerations. See Wolfram (1993) and Guy (1993) for discussions of some of the problems connected with settling on the variable(s) to be investigated.

1.4 Recording equipment

To name particular types of recording equipment would not be useful, because technology changes so rapidly. However, it is crucial that researchers use recording equipment meeting the technical specifications needed to produce sound of high enough quality for potential acoustic phonetic analysis (whether or not this is the immediate goal of the study) as well as high-quality external microphones. Some types of equipment have abiding advantages. For example, the lavaliere (lapel) microphone improves the quality of the sound and minimizes the speaker's attention to the recording mechanism. Also crucial is selection of recording location. Clearly, quiet locations are better than noisy ones; however, sometimes fieldworkers must sacrifice sound quality in favor of enhanced interactional quality when quiet locations that are comfortable to interviewees are not available. In addition, researchers should be aware that some types of noise that wreak havoc on audio recordings are practically unnoticeable to the untrained ear – for example, the noises emitted by electronic equipment (including computers), kitchen appliances, and espresso machines (despite the many other advantages of conducting interviews in comfortable public locations like coffee shops). The reader is referred to Cieri (2011) for excellent, detailed advice on selecting locations for interviews and choosing microphones and recording equipment. See also Schilling-Estes (2007) for good discussion and advice about videotaping interviews.

The main point is to get the best equipment possible given the practical constraint of expense. Recording fidelity is the primary consideration, and after that come ease of use, flexibility, weight, and other factors. Field recordings can be useful for many years, for purposes unplanned. In my case, tape recordings intended only for a study of grammar have since been used for work on phonology, both using impressionistic phonetic transcription and computer-assisted vowel analysis.

1.5 Institutional Review Board approval

Before heading off to the field, it is necessary to fulfill the requirements of the IRB – the Institutional Review Board – also called Ethical Review Board (ERB) or Independent Ethics Committee (IEC) at your institution and/or granting agency. In the US IRB approval is required for all research involving human subjects. Each

institution has its own requirements, so it is advisable to obtain and complete IRB approval forms early and allow enough time for your project to be reviewed and accepted. A crucial component of the approval processes involves preparing an Informed Consent Form which will have to be signed by each study participant (or legal guardian, in the case of children under 18).

1.6 Self-presentation of the fieldworker

Having selected the community and investigated the locale, culture, and speech, and having the approval of the IRB, the investigator finally has to actually go there and find people to talk to. This is a rather stressful position to be in, from all accounts. Eckert (2000) describes the nightmares she had before beginning her work in the Detroit suburbs. Entering any community carries with it certain responsibilities for respecting the privacy and customs of local people. Most often, this is not a great problem because researchers tend to investigate cultures with which they have some personal familiarity. It is a much greater problem, obviously, in a culture and language that is not native to the investigator. In these situations, Samarin (1967: 19) recommends that the researcher undertake meticulous planning to deal with the pressures, being aware of the problems that might arise and arranging for breaks in order to get away from the locale from time to time.

Often, cultural alienation is not a factor. My own fieldwork, for instance, took place in my home town, where I had lived until I was 15, and where both my mother and grandfather had grown up. My role there, while conducting fieldwork between 1969 and 1973 and again in 1990 and 1991, was both as a visitor in the town, staying with my grandparents, and as a researcher working on my dissertation, carrying out interviews. On my side of town I was known to the people I interviewed as a friend's granddaughter or cousin, but on the other side of town I was a complete stranger doing research. I told people that I was working on a book on growing up in the town, and how it was changing over time, especially for the teenagers. I said I wanted to record speech in the interests of accuracy, so I would get the dialog right. As a former resident with kinship ties in the town, I attended church with my family, visited friends, and took my grandmother to her club meetings. I also attended revival meetings and visited a church on the other side of town, which helped me learn about the life and culture outside of my own experience and to meet older people who were members of the church I visited.

I was careful to dress suitably according to local custom, always wearing a skirt and stockings to interview older people and to attend classes at the high school, but sometimes wearing blue jeans and sitting on the floor when interviewing teenagers, explaining that I needed to watch the level on the tape recorder while we were talking. In this way I was showing respect to my elders and solidarity with the younger group. With teenagers, I generally took along sodas and chips, which helped make the interview less formal, though the crunch of potato chips sometimes can be heard on the recording.

In reporting on his research in a small town in North Carolina, Hazen (2000) explains that before beginning his fieldwork he had married a woman from the

community, which gave him entree. However, as a native of suburban Detroit, he was not as well acquainted with the culture as he might have liked, though this also allowed him to assume the role of a student of that culture and ask questions that only an outsider could ask.

Albó (1970) describes in detail his entry into rural communities in Andean Bolivia where his identity as a priest proved advantageous. He was sometimes asked to bless houses, which gave him an opportunity to observe the living standards of the families. This contributed to his understanding of the degree of modernization of the household, giving insight into the relationship between the borrowing of linguistic forms and of material culture. It also gave him opportunities to line up interviews. Similarly, Harvey (1992), whose research was in Southern Peru after it became a dangerous area for outsiders, was considered the adopted daughter of a local family, which gave her a place in the community, allowing her to observe both language and culture.

Both Whyte (1943) and Liebow (1967) emphasize that it is never possible to completely fit in, nor is it necessary or even advisable. As white middle-class men carrying out ethnographic research among working-class men, one group white and the other black, they report that they were able to lower the barriers between their subjects and themselves but not to remove them. Liebow uses the image of the chain-link fence: you can see through it, but it remains a barrier. The researcher can become a friend, and even find a role in the community, but skin color, class affiliation, speech, or education may all set the investigator apart, which may of course result in less than ideal conditions for collecting maximally naturalistic speech data and maximally informed community understandings but which can also serve as a protection in some situations.

2 The Sociolinguistic Interview

The classic method of sociolinguistic research is the one-on-one recorded conversational interview (Labov 1972a, 1984; Wolfram and Fasold 1974). Recording has the obvious advantage of permanency, so that it is possible to return to the recording again and again, either for clarification or for further research. A second major advantage is that the recording permits the researcher to fulfill the Principle of Accountability (Labov 1972c: 72), so that all occurrences as well as non-occurrences of the variable in question can be identified and accounted for. In this way statistical manipulations of the data can show whether the occurrence of a variant is happenstance or patterned, and, if patterned, to what degree in contrast to its occurrence in the speech of others of varying social characteristics – age, sex, social class, ethnicity – and across speech styles. This, then, is the primary method of quantitative sociolinguistics.

Variations on this classic sociolinguistic interview approach include interviewing two or more speakers together (Feagin 1979), or even breaking down the one-on-one interview structure with pairs of interviewers. Labov used group interviews in his work with Harlem street gangs (Labov *et al*. 1968; Labov 1972b),

with one lavaliere microphone per person, and a multitrack recorder, while Wolfram, Hazen and Schilling-Estes (1999) reported great success in using pairs or even teams of interviewers in their work in Ocracoke, North Carolina. These variations on the one-on-one interview are intended to reduce the formality of the interview, turning it into a more natural social event.

The sociolinguistic interview – regardless of the variations on it – does carry some disadvantages. The interview as a speech event is a special genre (Wolfson 1976), so the naturalness and certainly the informality of the recorded speech can be called into question, regardless of efforts to make the speaker feel comfortable with the situation. The use of lavaliere microphones may remove the microphone from view, but the recording device is always there. However, despite our fears that recorded sociolinguistic interviews may be less than fully "natural," Eckert (2000) reminds us that speakers are adaptable in both research and non-research contexts, and they can just as readily shape their speech to naturally fit a conversational sociolinguistic interview as any other speech event in which they find themselves in the course of their daily interactions. Furthermore, recording devices are becoming less obtrusive as technology continues to advance, and non-research-related recording situations are more commonplace.

The interview method works best for frequently occurring variables, especially phonological and morphological, and certain syntactic forms, such as negation. But many syntactic structures, including interrogatives, double modals, and special auxiliaries such as perfective *done*, do not occur frequently enough in interviews to provide sufficient data for analysis. Moreover, the interview is problematic for discourse studies and ethnomethodology (Briggs 1986).

2.1 Selecting speakers

The earliest community-based research in sociolinguistics, Labov's work in Martha's Vineyard (1963), used a judgment sample, selecting subjects to fill preselected social categories, all locally born and raised adults and teenagers. His categories crisscrossed geographic area, profession, and ethnicity. It is interesting that in this early study gender was not considered a separate social variable, though only men were used for acoustic analysis. In his New York study a few years later, Labov was able to base his subject selection on a previous random survey by the Mobilization for Youth, a project of the School of Social Work at Columbia University, which had conducted a random-sample survey of the Lower East Side. Labov used their demographic data to select natives of the area or people who had arrived by age five, as well as people from across a range of social strata. This was, then, a stratified random sample in that it selected a stratified sample from what had originally been a random sample. In his third major project (Labov *et al*. 1968; Labov 1972b), Labov worked with teenage boys who were members of street gangs. This represents an early – possibly the earliest – study of language variation through social networks.

Trudgill (1974: 20–30), who followed soon after with a study of Norwich, England, relied on a quasi-random sample taken from four ward voter registration

lists. The names from the voter lists were chosen randomly, but the wards were not random but were selected "so that they had, between them, social and economic characteristics that were, on average, the same as those of the city as a whole" (1974: 22).

My own work in Anniston, Alabama, was based on a judgment sample, filling pre-selected cells on a number of criteria. First, speakers were chosen because they were native-born or had arrived by age five. Second, preference was given to those whose parents were from the area. Though I did not know of the literature on networks at the time, I often selected subjects who were "a friend of a friend," using the resources of my family and their acquaintances for contacts. I began with friends of a younger cousin, then moved on to friends of my grandparents. Later, when I wanted to work in another section of town, I began with a home economics teacher who turned out to be an acquaintance and an admirer of my grandfather. She welcomed me into her classes where I was able to observe, and in some cases (with permission) to record the students and make appointments with them for interviews. Twenty years later, in 1990, I followed the same procedures to find teenage subjects on both sides of town. Luckily enough, the new home economics teacher said that if the earlier one, who had been her own teacher, had let me visit her class, it was all right with her. The now-retired home economics teacher was still in touch with the students I had interviewed 20 years before, and through her I was able to find those students again, most of whom still lived in the area.

When the Milroys were selecting informants in Belfast (Milroy 1980; Milroy and Milroy 1985), they were forced to rely on the "friend of a friend" method for contacts because of the sectarian problems in the city, and especially in the working-class neighborhoods in which they intended to conduct their research. Their methods auspiciously introduced the concept of the network to sociolinguists. (See Milroy and Llamas, this volume, on social networks.)

Generally, researchers must use common sense to select subjects not by some pre-ordained "social-science" formula but according to the prevailing conditions of the setting they are working in, as well as their research goals. Thus, in selected subjects for their study of Ocracoke, Wolfram *et al*. (1999) chose ancestral islanders whose families had been on the island for at least several generations, because the purpose of the study was to recover, as far as possible, the traditional dialect that was rapidly eroding in the face of incursion into the island community by tourists and new residents from a range of dialect areas. In a quite different vein, Eckert (2000) selected high school students of opposing ideologies and styles, known as "burnouts" and "jocks," basically working-class and upper-middle-class adolescents, because she was studying the dynamics of adolescent speech and culture in the school setting.

One danger with selecting informants by pre-selected categories is that results can be self-fulfilling or circular. For a more general community study, Horvath (1985) gathered speech data from a stratified judgment sample in Sydney, Australia, and analyzed it using principal components analysis, a statistical technique which grouped speakers into clusters according to their linguistic similarities, and

in that way revealed what the sociolinguistic groupings of Sydney were, based entirely on speech, rather than on preconceived notions about class membership, sex, or other social groupings.

Except for studies that take a special interest in the language of children (as with Roberts, this volume), it is better to avoid speakers younger than adolescents, since there is the possibility of confounding phonological or grammatical development with local variation.

The two genders must be kept fairly even numerically in order to prevent a confounding of gender differences with other distinctions. Many studies have demonstrated gender differences in language, beginning with Fischer's (1958) study of (ing) which showed that boys in a small New England town were more likely to use the [ɪn] variant than girls.

Attention must also be given to social class (as in Ash, this volume), as well as its interaction with age and gender. The older members of any class usually have the most conservative phonology; teenage working-class boys and girls are often the leaders in innovation, with certain items being more characteristic of one gender than the other. Eckert (2000) elaborates a striking example of highly innovative teenagers who show gender- and social group-based differentiation in their usage patterns for new linguistic features. In regard to grammar, the higher classes will usually use a local variety of the standard; the older members of the working class will maintain older forms which have become nonstandard and which may be obsolete in other places, while the younger speakers may still use those forms, but may also show innovative forms. For example, an older working-class woman in Anniston used *clim* as the past participle of *climb*, a form which existed in seventeenth- and eighteenth-century English but which since has become obsolete.

Ethnicity often provides a striking correlate with linguistic variation. Wolfram *et al*. (1999) and Rickford (1985) have shown that African Americans and European Americans living together on isolated islands, of the same socioeconomic background, education and age, show consistent differences in their speech, both in phonology (on the Outer Banks) and in grammar (Sea Islands and Outer Banks).

2.2 *Sample size*

The next question to be resolved is how many speakers are needed. The question depends most directly on the number of independent variables. If you are interested in comparing the speech of working-class men and women of the same age, say, 30 years old, then you have subjects in only two cells: 30-year-old women and 30-year-old men. If you expand the study to include men and women of 60 as well, the number of cells doubles to four. If you expand to include both working-class and middle-class subjects, it doubles again to eight cells. Obviously, each cell must be filled with enough subjects to provide confident generalizations about the social group.

How many subjects should fill each cell? The simple answer is: the more the better. In practice, sociolinguistic analysis requires isolating and classifying dozens and sometimes hundreds of tokens from each subject. It bears little resemblance

to the sampling carried out in many kinds of social sciences for the purposes of opinion polls or voter preferences. As a rule of thumb, five persons per cell is often adequate, assuming the cells are well-defined in terms of local social categories (Guy 1980). I followed this rule in my Anniston study, where cells consisted of the independent variables of age/sex/social class/locale (urban/rural); so, for instance, I had to locate and interview at least five older rural working-class male informants.

2.3 Interview protocols and questionnaires

There are two main types of sociolinguistic interviews. The most influential one, modeled on Labov's work, uses a set of questions to elicit as much free conversation as possible, with some reading tasks designed to elicit a range of styles. Another way of going about it is simply to let the conversation flow (Briggs 1986; Hazen 2000). This more open-ended type of interviewing is intended to reduce the distance between interviewer and subject, making the interaction more naturalistic.

For the more structured interview, protocols may be found in appendices of several reports (Labov 1966; Feagin 1979; Labov 1984; Horvath 1985; Wolfram *et al*. 1999, to name a few). The chief goal is to obtain large quantities of speech that is as relaxed and naturalistic as possible; often, too, researchers will design protocols to sample other speech styles as well, for example reading styles considered to be more "formal" than spoken conversational speech. Some researchers, however, have considered the conceptualization of style as a unidimensional "formal–informal" continuum to be problematic (see Schilling, this volume), and so will focus solely on conversation rather than including readings as well.

Sociolinguistic interviews usually begin by asking subjects about themselves – year and place of birth, parents' birthplace, schooling (speaker's and parents'), occupation (their own or that of their parents or spouse). Questions like these often yield a relatively formal or self-conscious speaking style, known as Interview Style, as will discussion of school or the workplace (see Sankoff and Laberge 1978). Such questions invite self-conscious responses by asking the subjects to reflect on their histories and their accomplishments. However, in some circumstances, asking about school activities may elicit informal and spontaneous speech, for example, if directed to subjects deeply and personally involved in those activities. Thus Eckert's teenaged subjects become very animated when talking about activities, groups, and characters in their school, as did mine (Eckert 2000; Feagin 1979). This distinction is crucial in planning the interview protocol, since it is not really topic per se that correlates with degree of self-consciousness but rather extent and type of involvement with the topic area. People tend to be least self-conscious when talking about subjects with which they are intimately involved, while the most self-conscious speech comes from asking people to talk about their credentials.

In the opening section on demographics, asking the subjects to list the houses they have lived in can lead to a discussion of the neighborhood where

the speakers grew up, and that can lead to discussing childhood friends and describing rules for various games, jump-rope rhymes, and so on. Here the speaker will probably switch to a less formal, more conversational style. It is difficult to monitor one's speech when recalling and reciting such rhymes as "Fatty, fatty two-by-four, can't get through the bathroom door."

Asking the subjects about their first dates or how they met their spouses sometimes elicits a flood of speech, at least in the European-American context. Labov's best known question has to do with the danger of death: "Have you ever been in a situation where you were in serious danger of being killed, where you thought to yourself, *This is it* . . . What happened?" (Labov 1972a: 113).

While sometimes this elicits an outstanding narrative, it seems to work better in New York City than anywhere else. My speakers in Alabama, asked the same question, generally responded, after a pause, "No." Others have had similar experiences – Trudgill (1974) in Norwich, England, Chambers (1980) in Toronto, and Milroy (1980) in Belfast. In Anniston, after the danger-of-death question proved unsuccessful, I discovered that the question "Have you ever heard of anybody seeing a ghost around here?" often elicited long elaborate narratives of local mayhem and murder from older working-class speakers. Similarly, with his Canadian subjects in Toronto, Chambers discovered he could elicit passionate speech by saying, "People keep saying we're getting more and more American. Do you think that's true?"

The interview, obviously, must be adjusted for local conditions. Familiarity with local customs helps develop questions such as "When did you get your first gun?" in the southern United States, or "What were you doing when that tornado hit back in 1954?" There is no simple formula for eliciting relatively unmonitored, casual styles. The best advice is for researchers to know their regions, especially the tensions in the community, when planning the interview protocol.

2.4 More formal styles: Reading passages, word lists, minimal pairs

The use of written materials in the interview protocol depends on the focus of the research. Presenting subjects with a reading passage, word list, and minimal pair list can certainly be useful for research oriented toward phonology, because the researcher can ensure that the same words, involving particular phonological contrasts or certain variables in particular contexts, are recorded for every subject. In studying syntax, having the speaker read sentences while being recorded can produce valuable results, if they are used to elicit judgments on grammaticality or acceptability. The speakers can be asked who would use such a sentence, even if they themselves would not. If reading is a problem, as it often is for the oldest rural subjects either through poor eyesight or through illiteracy, having subjects repeat sentences read by the interviewer can also be a source of information. Wolfram and Fasold (1974) discuss repetition tasks and some of the information they can yield. In my own work, I started out using word lists and sentences, but dropped them, since I was concentrating on grammar alone. However, judgments

on sentences proved to be useful, as ancillary evidence. Now that I am using the same recordings to work on phonology, I am very much aware that it would have been helpful to have kept the word list to observe style shifting, and to get an idea of what might be considered more self-conscious speech.

Word lists and reading passages that have been used successfully may be found in the appendix to Labov (1966), Trudgill (1974), and elsewhere. See also Labov (1984) for a description of various field experiments and references to their use. Each community and each set of variables requires its own materials, but looking at previous models can be helpful.

2.5 *During the interview*

In conducting the sociolinguistic interview, it is important for researchers to give interviewees plenty of space to elaborate on topics of interest and to not waste too much time recording their own voices. Perhaps the most embarrassing moment for novice fieldworkers is the discovery, on listening to interviews they have made, that their own contributions limited what the subject might have offered by interjecting friendly asides or interrupting the flow of the subject's conversation. The resulting interviews sometimes preserve hard evidence of misguided sociability. Nonetheless, as Milroy (1987) notes, it is important to remember that interviews are exchanges, and interviewers do have to make contributions to get quality conversation in return. Keeping the attention and interest of the speaker during the interview is obviously important, and that makes it hard for the researcher to limit back-channelling. It is natural to respond to what the speaker says, to offer your own opinions and to bring up parallel experiences. And whereas sometimes interviewers can get a bit carried away, providing a reasonable amount of co-conversation can be valuable indeed. Breaking my self-imposed silence in a second interview with one of my subjects, comparing notes with the speaker on some experiences we shared, I discovered that the speaker's phonology and grammar altered at that point, with more local vowels – more breaking and shifting – and nonstandard grammar where there had been little or none before.

Thus, while controlling the inclination to take the floor, the interviewer must provide signs of involvement – both verbal and nonverbal (for example, maintaining eye contact, if culturally appropriate) – at the same time keeping a watchful eye on the recording equipment and a dutiful ear on the production of the desired variables.

2.6 *After the interview*

Whether or not to provide monetary compensation to informants is subject to debate (as in Whyte 1984: 361–365). While I have never paid speakers for interviews, others have and do. This may be a community-specific issue. Researchers are often graduate students working on doctoral dissertations – unpaid or poorly paid themselves, so that most of them rely on an exchange of services, such as giving rides, if the researcher has a car, helping with schoolwork, or writing letters,

as did Dayton (1996). As Whyte points out (1984) paying speakers can change the nature of the enterprise, even compromising the possibility of further research by making it much too expensive for others following after.

As noted above, another very important matter must be addressed before leaving the speaker: The person interviewed must sign an Informed Consent Form, indicating their understanding of the basic purpose of the research project and recording and their permission for the interview to be used for research purposes. The wording must be approved by the researcher's Institutional Review Board and any other relevant organizations or agencies (for example, appropriate school officials, if recording teenagers in a school).

Finally, as detailed in Wolfram, this volume, most sociolinguists feel strongly that they must give back to their communities of study, partly in exchange for community members' having shared with them their voices, life stories and life experiences, and partly because, no matter what community members have given them, sociolinguistic researchers feel a scientific and ethical obligation to share the linguistic knowledge they have gained through community studies with as wide an audience as possible, including not only academic audiences but also research communities and the general public.

2.7 Ethics

Surreptitious recordings, made by planting a recording device where it will capture ambient conversations without the knowledge or consent of the participants, are often illegal and are considered unethical – and pointless – by the vast majority of sociolinguists. In their favor, of course, is the elimination of the Observer's Paradox, but in purely practical terms, apart from ethics, sound quality is usually so poor that it is a waste of time, and discovery by the community can lead to serious repercussions. The legal aspects of surreptitious recordings have been discussed by Larmouth *et al*. (1992), who review state and federal laws of the United States, defend the use of such recordings, and illustrate their points with examples of real or possible situations and their legal outcomes.

Harvey (1992) made covert recordings of drunken speech because it was central to her research, and she states that, while she found it distasteful, she would do it again (1992: 80). She considers surreptitious recordings as no more unethical than researchers not being entirely open about their research agenda with speakers, as in my telling speakers that I was interested in what it was like growing up in Anniston, Alabama, rather than saying outright that I was interested in their grammar.

Most researchers consider that surreptitious recording violates the privacy of the subjects. Even in open recording, it is usually necessary to respect the privacy of subjects by disguising their identities. Some researchers use alpha-numeric codes for speakers, but a better system is to use pseudonyms that preserve clues to ethnic background and other essential traits, so that someone with a German name would be given a German pseudonym, and the same style of naming. Using carefully constructed pseudonyms rather than mysterious codes renders analysts' jobs easier and also results in more readable text. Recorded discussions of illegal

activities or private matters should be treated as confidential, regardless of the informant's attitude toward such things at the time.

3 Participant Observation

Because the effect of recording on the interview can never be completely eradicated and because interviews are entirely unsuitable for obtaining certain kinds of data, participant observation has come to be seen as a complementary method of data collection in variationist sociolinguistics. This entails living and participating in the community in some function other than as a linguist, while at the same time observing and noting particular types of linguistic data. Such observations are frequently used to supplement material collected from interviews, as by Labov *et al*. (1968) and Feagin (1979), but they can also be used as the primary source of data, as in Rickford (1975), Mishoe and Montgomery (1994), and Dayton (1996).

Participant observation is especially useful for studying infrequent grammatical items such as questions, modals, and particles, where recorded interviews will not capture these forms. Either the discourse constraints are such that the question/answer format or the extended narrative of the interview do not allow the forms, or the forms are too rare to make an interview worthwhile. For such variables, participant observation becomes necessary. It is crucial to remember that both participation and observation are crucial: The researcher must immerse him- or herself in the community as far as possible while at the same time maintaining some measure of outside, "observer" status.

One of the best discussions of the rationale for using participant observation as well as one of the most complete descriptions of this method as employed in variationist sociolinguistics is found in Dayton (1996: Chapter 2). Here Dayton relates how she, a white woman, became a member of an African-American working-class community in Philadelphia. She first lived in that neighborhood for two years simply as a graduate student, not participating in the life there. Then she lived as a participant observer for four and a half more years, becoming a block chairman, organizing clean-ups, volleyball games, and generally entering into the local African-American life in that block.

The participant observer studying forms not likely to surface in sociolinguistic interviews will write down their data rather than make audio recordings. Dayton managed to write down most of the data for her study within an hour of hearing it. She seldom attempted to store and remember more than three items at a time. Mishoe and Montgomery (1994), who collected their corpus of double modals through participant observation, report that they wrote items down within a minute of hearing them.

This technique has certain advantages over the recorded interview in that the researcher becomes an insider, in so far as possible, and can in this way overcome the Observer's Paradox. In order to do this the researcher must reach the point of understanding the communicative and interactional norms of the speech community and participating in the informal social ties and exchange relationships that hold the community together (Dayton 1996: 71).

In the course of her study, Dayton collected 3,610 tokens of African-American tense/mood/aspect markers (Dayton 1996: 55), probably the largest corpus of these grammatical forms. Her observations also included the more general social context as well as the linguistic context of the use of these markers.

The drawback of participant observation is that researchers cannot write down all the tokens of the variable they might hear. There is an inevitable selectivity in the linguistic record. The selectivity means that the data cannot be quantified, so that it is impossible to provide information on the relative frequency of the variable. In addition, there is no permanent record of the speakers, so that it is not possible to return to the source of the evidence. Here the question of accuracy and reliability naturally arises. Counterbalancing that, it permits the study of rare forms, otherwise undocumentable. And the perceptual saliency of the items can abet the accuracy of the observations. In another context, Wolfram suggests that socially marked items are the most transparent differences, and as such they rank high on a "continuum of linguistic trustworthiness" (Wolfram 1990: 125; similarly Dayton 1996: 68–80).

4 Rapid and Anonymous Observations

While participant observation is a very time-consuming and labor-intensive way to overcome the Observer's Paradox, another, faster technique is "rapid and anonymous observation," first described by Labov (1966, 1972c). By this method, the variable under study is embedded in the answer to a question that can be posed to strangers. Labov, in a famous example, asked sales clerks in department stores, "Where are the women's shoes?" The respondents replied, "Fourth floor." What Labov was interested in was the pronunciation of (r) in the words *fourth* and *floor*. Labov selected a range of stores, from luxury (Saks Fifth Avenue) to bargain basement (Kleins), and was able to confirm that sales clerks tend to speak in a manner that reflects the clientele. The clerks at Saks were r-ful as are upper-middle-class New Yorkers, while those at Kleins were r-less, like working-class New Yorkers. Labov was able to capture 528 tokens of *fourth floor* from 264 subjects in approximately 6.5 hours.

The simplicity of this study has encouraged replications of it in New York and many other places, either studying (r) or other variables. For example, in some communities, the question "Excuse me. Could you tell me what time it is?" (at the right time of day) will produce many tokens of *five* or *four*. This type of study obviously sacrifices knowledge of the background of the speaker in favor of the naturalness of the speech.

5 Life after Fieldwork

Whatever methods the researcher uses, when the fieldwork is finally completed, any sense of relief evaporates rapidly as the reality of analysis of all that data

dawns. Analysis, of course, moves the sociolinguist onto an entirely different level, with its own problems and its own rewards (as the following chapters in this volume make clear). The crucial first step, the fieldwork, becomes subordinated to finding, expressing and disseminating the substantive results of the project. Many sociolinguists firmly maintain that the more successful the fieldwork, the less noticeable it is in the final analysis and that fieldwork draws attention to itself mainly when the researcher has to concede that there are gaps in the data, flawed elicitations, or results that require caution in the interpretation. However, as variationists increasingly incorporate ethnographic and social constructionist viewpoints into their work, more attention is being given to how data and analyses are inevitably shaped by research methods, research contexts, and researchers themselves, and so the fieldwork process most likely will not remain quite as backgrounded as it traditionally has been. For the moment, though, the sociolinguist's prowess as fieldworker is often a private source of professional pride that only occasionally seeps into the public domain when sociolinguists gather informally at conferences and meetings. Inconspicuous it may be, fieldwork is the bedrock of the sociolinguistic enterprise, and it is crucial for novice researchers and advanced scholars to understand the methodological underpinnings of even the most theoretically sophisticated analyses.

REFERENCES

Albó, Xavier (1970) Social constraints on Cochabamba Quechua. PhD dissertation, Cornell University.

Ash, Sharon (2000) Sampling strategy for the Telsur/Atlas Project. http://www.ling.upenn.edu/phonoatlas/sampling (accessed February 5, 2013).

Bailey, Charles-James N. (1996) *Essays on Timed-based Linguistic Analysis*. Oxford: Oxford University Press.

Baugh, John (1993) Adapting dialectology: The conduct of community language studies. In Dennis Preston (ed.), *American Dialect Research*. 167–192. Philadelphia: John Benjamins.

Bowern, Claire (2008) *Linguistic Fieldwork: A Practical Guide*. New York/Basingstoke, Hampshire: Palgrave McMillan.

Briggs, Charles L. (1986) *Learning How to Ask: A Sociolinguistic Appraisal of the Role of the Interview in Social Science Research*. Cambridge: Cambridge University Press.

Cieri, Christopher (2011) Making a field recording. In Marianna Di Paolo and Malcah Yaeger-Dror (eds.), *Sociophonetics: A Student's Guide*. 24–35. New York: Routledge.

Chambers, J.K. (1980) Linguistic variation and Chomsky's "homogeneous speech community." In Murray Kinloch and A.B. House (eds.), *Papers from the Fourth Annual Meeting of the Atlantic Provinces Linguistic Association*. 1–32. Fredericton: University of New Brunswick.

Dayton, Elizabeth (1996) Grammatical categories of the verb in African-American Vernacular English. PhD dissertation, University of Pennsylvania.

Di Paolo, Marianna and Yaeger-Dror, Malcah (2011) Field methods: gathering data, creating a corpus, and reporting your work. In Marianna Di Paolo and Malcah Yaeger-Dror (eds.), *Sociophonetics: A Student's Guide*. 7–23. New York: Routledge.

Eckert, Penelope (2000) *Linguistic Variation as Social Practice: The Linguistic Construction of Social Identity in Belten High*. Oxford: Blackwell.

Feagin, Crawford (1979) *Variation and Change in Alabama English: A Sociolinguistic Study of the White Community*. Washington, DC: Georgetown University Press.

Fischer, John L. (1958) Social influences on the choice of a linguistic variant. *Word* 14: 47–56.

Guy, Gregory R. (1980) Variation in the group and in the individual. In William Labov (ed.), *Locating Language in Time and Space*. 1–36. New York: Academic Press.

Guy, Gregory R. (1993) The quantitative analysis of linguistic variation. In Dennis Preston (ed.), *American Dialect Research*. 223–249. Philadelphia: John Benjamins.

Harvey, Penelope (1992) Bilingualism in the Peruvian Andes. In Deborah Cameron, Elizabeth Frazer, Penelope Harvey, M.B.H. Rampton and Kay Richardson (eds.), *Researching Language: Issues of Power and Method*. 65–89. London: Routledge.

Hazen, Kirk (2000) *Identity and Ethnicity in the Rural South: A Sociolinguistic View through Past and Present Be*. Publication of the American Dialect Society 83. Durham, NC: Duke University Press.

Horvath, Barbara (1985) *Variation in Australian English: The Sociolects of Sydney*. Cambridge: Cambridge University Press.

Labov, William (1963) The social motivation of a sound change. *Word* 19: 273–309.

Labov, William (1966) *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Labov, William (1972a) Some principles of linguistic methodology. *Language in Society* 1: 97–120.

Labov, William (1972b) *Language in the Inner City*. Philadelphia: University of Pennsylvania Press.

Labov, William (1972c) *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William (1984) Field methods of the project on linguistic change and variation. In John Baugh and Joel Sherzer (eds.), *Language in Use*. 28–53. Englewood Cliffs, NJ: Prentice-Hall.

Labov, William, Ash, Sharon, and Boberg, Charles (2006) *Atlas of North American English: Phonetics, Phonology, and Sound Change*. Berlin: Mouton de Gruyter.

Labov, William, Cohen, Paul, Robins, Clarence, and Lewis, John (1968) A study of the Non-Standard English of Negro and Puerto Rican speakers in New York City. Cooperative Research Project No. 3288. Office of Education, US Department of Health, Education, and Welfare.

Larmouth, Donald W., Murray, Thomas E., and Ross Murray, Carin (1992) *Legal and Ethical Issues in Surreptitious Recording*. Publication of the American Dialect Society, 76. Tuscaloosa: University of Alabama Press.

Lavandera, Beatriz (1978) Where does the sociolinguistic variable stop? *Language in Society* 7: 171–83.

Liebow, Elliot (1967) *Talley's Corner: A Study of Negro Streetcorner Men*. Boston, MA: Little, Brown.

Milroy, James and Milroy, Lesley (1985) Linguistic Change, Social Network and Speaker Innovation. *Journal of Linguistics* 21: 339–384.

Milroy, Lesley (1980) *Language and Social Networks*. Oxford: Basil Blackwell.

Milroy, Lesley (1987) *Observing and Analyzing Natural Language*. Oxford: Blackwell Publishers.

Milroy, Lesley and Gordon, Matthew J. (2003) *Sociolinguistics: Method and Interpretation*. Oxford: Blackwell.

Mishoe, Margaret and Montgomery, Michael (1994) The pragmatics of multiple modal variation in North and South Carolina. *American Speech* 69: 3–29.

Rickford, John (1975) Carrying the new wave into syntax: The case of Black

English BIN. In Ralph W. Fasold and Roger W. Shuy (eds.), *Analyzing Variation in English*. 162–183. Washington, DC: Georgetown University Press.

Rickford, John (1985) Ethnicity as a sociolinguistic boundary. *American Speech* 60: 99–125.

Romaine, Suzanne (1980) A critical overview of the methodology of urban British sociolinguistics. *English World-Wide* 1: 163–198.

Romaine, Suzanne (1981) On the problem of syntactic variation: A reply to Beatriz Lavandera and William Labov. *Working Papers in Sociolinguistics 82*. Austin, TX: Southwest Educational Development Laboratory.

Samarin, William (1967) *Field Linguistics: A Guide to Linguistic Field Work*. New York: Holt, Rinehart and Winston.

Sankoff, David and Laberge, Suzanne (1978) The linguistic market and the statistical explanation of variability. In David Sankoff (ed.), *Linguistic Variation: Models and Methods*. 239–250. New York: Academic Press.

Schilling, Natalie (2013) *Sociolinguistic Fieldwork*. Cambridge: Cambridge University Press.

Schilling-Estes, Natalie (2007) Sociolinguistic fieldwork. In Robert Bayley and Ceil Lucas (eds.), *Sociolinguistic Variation: Theories, Methods, and Applications*. 165–189. Cambridge: Cambridge University Press.

Trudgill, Peter (1974) *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Wald, Benji (1973) Variation in the system of tense markers of Mombasa Swahili. PhD dissertation, Columbia University.

Whyte, William F. (1943 [1993]) *Street Corner Society: Structure of an Italian slum*. 4th ed. Appendix A: On the Evolution of Street Corner Society. Chicago: University of Chicago Press.

Whyte, William F. (1984) *Learning from the Field: A Guide from Experience*. Beverly Hills, CA: Sage Publications.

Wolfram, Walt (1990) Review article: Re-examining Vernacular Black English. *Language* 66: 121–133.

Wolfram, Walt (1993) Identifying and interpreting variables. In Dennis R. Preston (ed.), *American Dialect Research*. 193–221. Amsterdam: John Benjamins.

Wolfram, Walt and Fasold, Ralph (1974) *The Study of Social Dialects in American English*. Englewood Cliffs, NJ: Prentice Hall.

Wolfram, Walt, Hazen, Kirk, and Schilling-Estes, Natalie (1999) *Dialect Change and Maintenance on the Outer Banks*. Publication of the American Dialect Society 81. Tuscaloosa: University of Alabama Press.

Wolfson, Nessa (1976) Speech events and natural speech. *Language in Society* 5: 189–209.

2 Data in the Study of Variation and Change

TYLER KENDALL

From its modern beginnings with the work of William Labov (1963, 1966), the sociolinguistic study of language variation and change has centered on the empirical investigation of naturalistic talk, primarily gathered in the field (see Feagin, this volume) and elicited through the sociolinguistic interview (Labov 1966, 1972). Other data types and methodologies have also been found to be useful for investigating language variation and change – such as the use of written records (Schneider, this volume), general public corpora (Bauer 2002), attitudinal data (Preston, this volume), and, increasingly, experimental laboratory-based work (e.g. Campbell-Kibler 2007) – but data obtained through conversational, sociolinguistic interviews remain the bread and butter of sociolinguistic research. Yet, what do we actually mean when we talk about "data" in the study of language variation and change? How does speech became data? Every act of analysis involves interpretation and abstraction and the process of going from actual, naturalistic talk in and of the community to the sort of quantitative data that sheds light on the nuanced "orderly heterogeneity" of language (Weinreich, Labov, and Herzog 1968) is often under-discussed in the literature, although with further consideration it is quite complex.

This chapter considers the nature of spoken language data and how it is treated in variationist research. Following somewhat from Kendall (2008),[1] I focus on the meta-theoretical questions of just what variationist data are and how modern computer-based techniques can enhance sociolinguists' connection to and use of their data. I also illustrate how the ways we conceptualize our data and interact with them impacts our analyses and our understanding of the very task of studying language variation and change.

The Handbook of Language Variation and Change, Second Edition.
Edited by J.K. Chambers and Natalie Schilling.

1 A Brief History of Data in Sociolinguistics

Since its inception, sociolinguistics has been driven by an interest in natural, authentic language data and, over the history of the discipline, scholars have recorded a huge amount of speech. Some of the earliest modern projects alone, such as Wolfram's (1969) work on African American English in Detroit, collected many hundreds of hours of audio over the course of their fieldwork. The treatment of these recordings – from the more meta-theoretical question of how they have been conceptualized as the data upon which sociolinguistic descriptions and theories are made, to the more methodological issues of how they have been managed and preserved – has changed over the years. I here consider how sociolinguists have treated and described these data, their actual speech recordings, during the past 50 years.

Research reports in the early days of sociolinguistic variation studies (e.g. Labov 1966; Wolfram 1969; Sankoff and Sankoff 1973; Trudgill 1974) tended to publish thorough accounts of their methods, ranging from explications of their sampling techniques – how and why they chose the informants they did – to discussions of their interviewing strategies and even of training their fieldworkers. These methodological reports were an important and necessary step in establishing sociolinguistics as a credible and quantitatively oriented social science and served to aid future scholars by explicitly sharing "the knowledge of the problems [the researchers] faced and the solutions [they] tried" (Sankoff and Sankoff 1973: 12).

Many of these early reports included detailed information about their recordings. For instance, Shuy, Wolfram, and Riley (1968) committed an entire volume to their field methods for the Detroit Dialect Study (Wolfram 1969). Through it, they provide a more thorough account of the treatment of that project's data and recordings than most other projects have, explaining and commenting upon a range of issues from determining a sampling system to designing the format of the files used for coding their data. Similarly, Sankoff and Sankoff (1973) provide a thorough overview of the field and laboratory methods for their sociolinguistic study of Montreal French, including a discussion of their computerized transcription system and an overview of their complete database, which they enumerate in detail:

> (i) 120 reels of taped interviews (2 copies);
> (ii) 64 boxes, most of them full, of computer cards containing transcriptions, about 100,000 cards in all;
> (iii) computer printouts (in several copies) in readable format;
> (iv) in addition, we are presently storing corrected transcriptions on a master computer tape. Thus, to date, 40 interviews, over 20 boxes of cards, are now stored on a single reel of tape at the Centre de Calcul. (Sankoff and Sankoff 1973: 42)

Much of the work following this first generation of sociolinguistic research did not address its data to such a detailed degree. It may be that the level of

description provided by Sankoff and Sankoff became viewed as unnecessarily detailed. Yet, the earliest projects had a world of options surrounding them about what to analyze and even what to consider as their data for analysis. They had to be explicit about each step of their work as they abstracted from real-world speech events to filtered-down quantitative variable data. Detailed accounts of their data and research steps were needed. As these studies found success by yielding robust findings and indicating just how systematic language variation actually is, it became less important to dwell on many of the technical details. Just as fieldwork moved from rigorous, large-scale random sampling techniques (Labov 1966; Wolfram 1969) to network studies and "friend of a friend" sampling techniques (cf. Milroy and Gordon 2003) as it became clear that the systematicity of language variation was discoverable from smaller scale (and more locally sensitive) studies (cf. Eckert 2005), the fine attention to reporting each step of one's analysis process also became less important. In a sense, the data of variationist analysis often jumps from the actual recordings of conversational sociolinguistic interviews to spreadsheets of variable instances.

It is important also to appreciate that recordings during the first decades of modern sociolinguistics were expensive. Before the advent of lightweight tape recorders and now ultra-lightweight solid-state digital recorders, recording equipment was large, cumbersome, and costly. As recording technologies became more accessible via inexpensive and ubiquitous equipment and storage media, sociolinguists' discussions of their methods could focus less and less on the practicalities of recording and the details of the actual, physical recordings. A review of many popular sociolinguistic textbooks shows that their discussions of methods often skip from data acquisition to data analysis and/or to demographic and theoretical issues pertaining to analyzing language in relation to social attributes of speakers (as in Wolfram and Fasold 1974; Milroy 1987; Milroy and Gordon 2003). There are numerous robust discussions of issues like how to choose informants, how to elicit and obtain "good" speech, and how to analyze the resulting sociolinguistic variable data. However, almost across the board these discussions neglect issues in how the speech recordings should be organized, stored, preserved, and so on.

Of course, there are some exceptions in the literature. Poplack (1989) discussed the Ottawa-Hull French Project's data archive and methodology in detail, a project with a goal to improve methodologies inherent in working with large sets of data for sociolinguistic analysis. Her methodologically focused paper responded to the fact that:

> One area in which development has been sporadic at best is in the construction of major sociolinguistic databases. The trade-off between sociological representativeness and ethnographic thoroughness has resulted in insufficient data from a large sample of speakers, or masses of data of questionable generalizability from a few speakers. Efforts to increase quantity or authenticity of recordings are still marked by losses in the quality of the data obtained. And even as a database reaches respectable size, its accessibility is concurrently hampered by the uneconomical effort needed to search it systematically in studies of individual variables. (1989: 413)

Poplack's paper provides a thorough treatment of many of the steps, from determining a sample population, to interviewing and recording that population, to organizing the resulting collection, and to developing a computer-based corpus of the recordings.

Much recent work in sociolinguistics has returned to dealing explicitly and thoroughly with its data. Tagliamonte's (2006) textbook, *Analysing Sociolinguistic Variation*, has an entire chapter, "Data, data, and more data," that reviews a wide range of data management tasks, from labeling and organizing interviews into a coherent corpus to transcribing the data and working with computerized transcripts and recordings. Schilling's (2013) book *Sociolinguistic Fieldwork* also discusses a range of important data management and preservation questions. Meanwhile, many funding agencies, such as the National Science Foundation in the US, have recently instituted policies about the management, preservation, and dissemination of data collected under funded research.[2] Likely these kinds of policies will make the explicit treatment of data a larger part of sociolinguistic research endeavors in the coming years. My own work (e.g. Kendall 2008, 2011) has attempted to explore the consequences of our relationships with our data, how the decisions we make – for example, when organizing our data, when transcribing, and so on – impact the kinds of questions that we can ask and the answers that we obtain. In Section 3 I consider this point more thoroughly, but first we consider the status of "corpora" in work on language variation and change.

2 Sociolinguistics, Corpora, and Data Sharing

Several publications consider the ways that standard, publically available corpora can be used to examine language variation and change (Bauer 2002; Baker 2010; Kendall 2011). I do not consider the sociolinguistic analysis of public, or "conventional" (Beal, Corrigan, and Moisl 2007a), corpora in this chapter but rather the fact that sociolinguists are increasingly discussing and thinking about their own data as corpora, a reconceptualization that has potential benefits for the variationist endeavor. Viewing our recording collections and data as corpora – as coherent, self-contained, representative samples of a language variety (see below for a fuller definition) – positions us to be more explicit about just what counts as the data used in a particular project, what is included, what is not, and how we (and others) access them. It also better supports the model of replicable research to which all scientific research should strive.

Corpus linguists primarily view corpus linguistics as a methodology rather than a theoretical stance (cf. McEnery and Wilson 2001; McEnery, Xiao, and Tono 2006; Gries 2009; Kendall 2011) and as such can offer complementary research methods and practices to the investigation of language variation and change. However, corpus linguists differentiate corpora proper from other, less systematically developed collections of language data. Corpora are often defined as involving an explicit focus on:

- *Representativeness* and *balance* – a corpus should accurately represent the full language variety it purports to contain; further, it should be balanced across the proportions of linguistic and social categories that comprise the variety.
- *Machine-readability* – a corpus should be machine-readable (which typically means electronic text).
- A particular (large) *size* – many descriptions of "what makes a corpus a corpus" do not explicitly argue for a size requirement, but in reading the corpus linguistic literature one notes a common focus on size measured in number of words and a growing interest in large corpora.

While corpus linguists pay a great deal of attention to the notions of representativeness and balance, these are often taken to be more of an ideal than a strict requirement. It is, of course, often not possible to represent a language variety precisely in a corpus. Gries writes:

> If I know that dialogs make up 65 percent of the speech of adolescent Californians, approximately 65 percent of my corpus [of adolescent Californian speech] should consist of dialogue recordings. This example already shows that this criterion is more of a theoretical ideal: How would one measure the proportion that dialogs make up of the speech of adolescent Californians? (2009: 8)

Even though corpus linguists accept that these criteria are difficult to meet in actuality, many sociolinguistic recording collections do not attempt to meet the sampling criteria or size to be considered "corpora" by many corpus linguists. (Instead, they might be termed "archives" or "databases" by these scholars.) But, terminology aside, thinking about sociolinguistic data in terms of corpora can benefit both the long- and short-term life of the data. For instance, thinking about sociolinguistic fieldwork as corpus creation (in the terms spelled out in, say, McEnery *et al*.'s 2006 introductory text about corpus-based language study) can lead to better-organized and more manageable data collections.

One lesson in particular to take from the corpus linguistic literature is that it is beneficial to build and organize data collections with the goal that an unfamiliar researcher could make sense of the data without you. It may be the case that you do not plan to share your recordings with anyone else (or that you are unable to, see below), but if you return to your data collection five or 10 years – or even six months – in the future you may find that you approach the data as would a total stranger, for example, no longer remembering how to interpret the file-naming conventions or directory structure.

One of the biggest hurdles to overcome in building sociolinguistic corpora may not actually be technical but rather about who will be able to access the recordings. In some cases, the question of whether or how to share the recordings and other data generated over the course of a research project and who to share with (direct collaborators, other researchers, the general public) is something that each researcher must consider for her- or himself. In other cases, as indicated above, funding agencies or other institutions might enforce a data-sharing or data-accessibility plan, or, as discussed below, human subjects concerns might prevent