Foreword

Observing the Observers

The ideal data for studying the social uses of language, as all the authors in this book agree, are speech produced in natural circumstances, unmonitored and carefree. Getting access to that kind of speech is challenging, for obvious reasons. For starters, in order to make the speech accessible for study it must be elicited, and eliciting speech usually introduces "foreign" elements into the speech act, including the presence of an outsider (the investigator), recording devices (microphones in plain view), controlled ambience (full attention and relative quiet), and strange tasks interspersed with conversation (reading a text or word list, identifying pictures, and other means to guarantee comparable material from all subjects). The act of observing speech alters its nature.

Sociolinguists seek to observe speech as people use it when they are not being observed. That is the "observer's paradox," and it has been a central preoccupation of sociolinguistic methodology from the beginning (Labov, 1972, p. 61). It is, naturally, a central preoccupation of this book; it is cited explicitly in the chapters and vignettes by Kara Becker, Niko Besnier, Charles Boberg, Becky Childs, Cynthia G. Clopper, Paul De Decker and Jennifer Nycz, Joseph Hill, Laurel MacKenzie, and Sara Trechter. It is implicit almost everywhere.

In the four eventful decades since William Labov gave a name to the observer's paradox, sociolinguists have come up with several ways of neutralizing it. Indeed, one of the rewarding sub-themes of this book is discovering how the experienced fieldworkers who contributed the chapters and vignettes got around it.

One obvious stratagem is diversion. One of the most ingenious examples in my experience was devised by an undergraduate in a course I taught in the 1970s. In those days, the stressed vowel in the word *tomato* had three variants in Toronto: either [ei], the North American variant, or [ɑ], modeled on the British pronunciation, or [æ], a distinctive Canadianism that came into being as a fudge between the other two variants. In order to discover the social correlates of the three variants, my student mounted four pictures on a poster: a cauliflower, a carrot, an apple, and (inevitably) a tomato. He visited department stores frequented by different social classes (following Labov's famous department-store study described, for instance, by Barbara M. Horvath in Chapter 2). He approached shoppers, and, after a friendly introduction, he showed them the poster and asked, "How many of these are vegetables?" If they said "two," he challenged them: "Why not three?" They inevitably answered, "Because a tomato is a fruit, not a vegetable." And, conversely, if they answered "three," he queried their answer, and was told, "Because a tomato is a vegetable, not a fruit (whatever other people might say)." His subjects had no idea, of course, that he was eliciting their pronunciations; they assumed he was challenging their botanical acumen, in which the classification of the tomato is a

well-known point of contention. In a short time, he accumulated hundreds of responses and he was able to show that social class sometimes interacted with age: people under 40 all used the [ei] variant except for a few oddballs from the upper middle class. (Since then, they too have disappeared, and the [ei] variant is nearly unanimous throughout Canada.)

This method has proven practicable for small-scale studies like the *tomato* variable, known as "rapid and anonymous surveys" (discussed by Charles Boberg in Chapter 8, Laurel MacKenzie in Vignette 8c, and Gerard Van Herk in Chapter 10). Nevertheless, the basic idea of framing the interview context so that the subject's attention is fixed on something other than the speech act is one of the key devices for blunting the paradox or, put positively, for eliciting unmonitored speech. Several authors in this book make suggestions and provide models toward that end.

Special communities require specialized methods, and they too are covered incisively in this book. Among them are immigrant communities (discussed by James A. Walker and Michol F. Hoffman in Vignette 5b and Rajend Mesthrie in Vignette 5c, among others), Deaf communities (see Vignette 6b by Ceil Lucas and Vignette 6c by Joseph Hill), LGBT communities (discussed by Lal Zimman in Vignette 3b, Stephen L. Mann in Vignette 3c, and Andrew D. Wong in Vignette 18c), closed enclaves ("clans" in Vignette 2d by James N. Stanford), and moribund dialects and endangered languages (discussed, respectively, by Patricia Causey Nichols in Vignette 5d and D. Victoria Rau in Vignette 6a). There are also data sources that are far removed from unmonitored natural speech but, with suitable precautions, can yield sociolinguistic insights. Prominent among these are the "public" languages of the mass media (called "performed language" by Robin Queen in Chapter 13 and "scripted data" in the vignettes by Tracey L. Weldon, Vignette 13a, and Michael Adams, Vignette 13b). Equally public if less prominent are courtroom transcripts, trial documents, and the like (discussed from different perspectives by Susan Ehrlich in Vignette 3d, Philipp Sebastian Angermeyer in Vignette 11b, and Ronald R. Butters in Vignette 11c).

Besides speech, there are also data sources in written materials. As Edgar W. Schneider says in Chapter 11, in his discussion of written material, writing "represents a secondary encoding of speech." The language we write is a kind of abstraction of the language we speak, hemmed in as it is by spelling conventions and stylized formatting. Nevertheless, written records existed for a millennium or more before audio recordings. Comparative linguistics, the most vital branch of language studies until the early 20th century, made monumental advances in genetic classification based almost entirely on classical texts. Those materials and other written documents continue to yield insights, and those insights are all the more astute now that we have deeper understandings of spoken vernaculars. Knowing the dynamics of living languages enriches our understanding of ancient processes on the understanding that linguistic processes were the same in nature and kind hundreds of years ago as the ones we now observe. Blurring the line between spoken and written language, the ways people communicate on the internet provides sociolinguists with yet another type of data to examine. Of course, the task of collecting online language data brings its own assorted challenges, many of which are discussed by Jannis Androutsopoulos in Chapter 14 as well as in vignettes that cover virtual worlds (Randall Sadler, Vignette 3e), Facebook status updates (Ruth Page, Vignette 14a), and Twitter data (Steven Coats, Vignette 14b).

One type of written documentation that has long proved useful in supplementing our linguistic knowledge is the written questionnaire, discussed by Charles Boberg in Chapter 8. The obvious limitation of asking people to tell us what they say is that they can only tell us what they *think* they say, which is not always the same thing. Subtle phonetic distinctions often require some training to recognize, and rare syntactic constructions sometimes strike users as strange even though they themselves may use them.

Self-administered questionnaires definitely work best on well-defined and easily discernible features. Countering this limitation, as Boberg makes clear, is their efficiency. Written surveys can cover a large territory with great density in a short time.

One of the minor cavils in Boberg's account is illustrated from my Dialect Topography survey, and I cannot resist showing that subsequent information gives it a rather more positive spin than was originally evident. Open-ended questions, Boberg notes, sometimes "elicit an overwhelming variety of minority responses." As a case in point, he cites a question about the schoolyard prank now widely known as a *wedgie*. True enough, when the question was first posed in the early 1990s, the responses were (almost) "overwhelming": specifically (as discussed in detail in Chambers, 2017) there were four main responses and at least a dozen minor ones—and almost everybody over 50 left it blank. What a mess, we thought at the time. But when we replicated the survey 10 years later, the results were stunningly different: this time, there was only one word for it. Almost everyone (93%) called it *wedgie*, including many old-timers. What had happened in the 10-year interval is that the *wedgie* had entered general consciousness. It had previously existed in the semi-literate subculture of grade-schoolers, but suddenly it was known to almost everyone. The word *wedgie* showed up in dictionaries, and it was called that by teachers, parents, and some grandparents as well as by schoolchildren. The shift from the profusion of responses in the first survey to the focusing of the later one documents "a prototypical standardizing change" (2017), one of the best yet documented. The real-time evidence of the second survey illuminated the profusion of minority responses that formerly seemed overwhelming. Without it, Boberg may be right in saying that open-ended questions may sometimes yield more information than we know what to do with.

The sheer volume of data when we study language as it is used by real people in real situations was one of the chronic problems of dialectology. As Kretzschmar, Schneider, and Johnson (1989, p. v) put it some years ago: "The development of dialect studies, whether geographical or sociolinguistic, has always been hampered by a superfluity of data." This statement appeared in one of the pioneering introductions to computer applications in dialectology, and so it set up the problem of "superfluity" in the context of its solution. Data-handling is no longer the overriding problem that it once was. Our discipline has made striking advances in storing, manipulating, and processing data. These aspects are to some extent inseparable from other matters and find their way into virtually every chapter of the book. They come to the fore especially in Chapter 7 on technology, by Paul De Decker and Jennifer Nycz, and the vignettes associated with it, and in Chapter 12, by Tyler Kendall, on preserving and accessing data, and the vignettes that go with it as well.

This collection provides a balanced, judicious, forward-looking summation of the ways in which we collect, access, and process the data that are the foundation of our enterprise. In its format and its tone, it has the feeling of a symposium involving a select group of sociolinguists sharing their personal experiences as well as their collective wisdom. It is an invaluable sourcebook for researchers and students and also for veteran fieldworkers in the diverse situations we face on entering the community.

<div align="right">J. K. Chambers</div>

References

Chambers, J. K. (2017). Cracking the code: *Wedgies* and linguistic respectability. In B. E. Evans, E. Benson, & J. Stanford (Eds.), *Language regard: Methods, variation and change*. New York: Cambridge University Press.

Kretzschmar, W. A., Jr., Schneider, E. W., & Johnson, E. (Eds.) (1989). Computer methods in dialectology. Special edition of *Journal of English Linguistics*, 22.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.