

## 4 Generating New Data

*Becky Childs*

In Part II of this volume, “Generating New Data,” we take an in-depth look at methods and processes in creating data and corpora for sociolinguistic analysis. The chapters and vignettes in this section address the various ideological frameworks and applied methods for creating and dealing with new sociolinguistic data, and readers will note the predominant theme of creating data that speaks best to one’s research questions and strengths as a researcher.

In Chapter 5, Erez Levon discusses ethnographic data collection. He presents a comprehensive chronological overview of the use of ethnographic data collection in sociolinguistic research and walks readers through the strengths and weaknesses of the method, with reference to specific studies. Focusing on four guiding principles for conducting ethnographic fieldwork, the chapter moves readers through the important processes of accessing a community, interacting with participants, collecting data, and then, importantly, leaving the community. Levon provides specific examples from his own research in Israel to illustrate approaches and methods to be considered when conducting ethnographic research.

The vignettes that follow the chapter on ethnographic data collection present three studies in communities from around the world that have used an ethnographically informed framework for data collection. Vignette 5a, by James A. Walker and Michol F. Hoffman, looks at fieldwork in immigrant communities. They begin by identifying the characteristics of immigrant communities and then work through their study, which looks at the place of immigrant English varieties in the large metropolitan city of Toronto. Walker and Hoffman offer suggestions for ways to effectively collect a corpus in an immigrant community, from using in-community fieldworkers to university-affiliated fieldworkers, and focusing throughout on the importance of building rapport with a community. In Vignette 5b, Rajend Mesthrie describes his work in a migrant and diasporic community in South Africa. He recounts a range of experiences often encountered when engaged in ethnographic study, including how to locate speakers of language variety under study to the strange and at times humorous interactions and missteps that can occur when working with people in a community that is not your own. In Vignette 5c, “Fieldwork in Remnant Dialect Communities,” Patricia Nichols describes her work in the Gullah community in the sea islands of South Carolina. Having spent a significant amount of time situating herself in

the community before she began data collection, Nichols' experience as explained in this vignette illustrates the importance of gaining community insight prior to initiating a research study.

Chapter 6, "The Sociolinguistic Interview," by Kara Becker, looks intensely at the construct that has been the standard for data elicitation in the field. Becker carefully breaks down this method for data collection, beginning the chapter with a strict definition, moving on to look at the usefulness of this elicitation tool, and highlighting the utility of the sociolinguistic interview, especially when the research question involves an examination of different speech styles. Becker concludes her chapter with a reminder that any elicitation tool must work in hand in hand with one's research question and that the sociolinguistic interview is well suited for any study whose central questions fall within the Labovian variationist paradigm.

Four vignettes accompany Chapter 6, each showing the application of the sociolinguistic interview in a specific study. Vignette 6a, "Cross-cultural Issues in Studying Endangered Indigenous Languages," by D. Victoria Rau, looks at the use of the sociolinguistic interview in an endangered indigenous language, Yami. Rau worked within the Labovian variationist paradigm and adopted a traditional sociolinguistic methodology for data collection, using word lists, texts for intelligibility tests, tests of bilingual ability, and information about language use and language attitudes. While some of these methods were successful, others had to be adapted to fit the community, and Rau recounts these adaptations and the rationales. Finally, Rau leaves us with a four-step approach to data collection, especially if the purpose is to produce materials. In Vignette 6b, Ceil Lucas explains the process of conducting a sociolinguistic interview in Deaf communities and the concerns that linguistic work in communities that have historically had direct ties to education and educational policy may bring up. These considerations are also echoed in Vignette 6c, where Joseph Hill takes a close look at two major issues that arise in collecting sign language data. Specifically, he points to the role of the observer's paradox and the signer's sensitivity to the interlocutor's audiological status and ethnicity. While these issues are of specific concern in the Deaf community, issues of this type (interlocutor ethnicity, age, and social class, among others) are common concerns that ideally are considered before data collection begins. Vignette 6d, "Other Interviewing Techniques in Sociolinguistics," by Boyd Davis, takes a critical look at the sociolinguistic interview, provides alternative methods for data collection, and reminds that there is no "one size fits all" method that will answer any research question.

Although methodological issues such as interview style are important considerations in sociolinguistic data collection, we cannot overlook the methodological concerns related to the actual recording process. In the past 20 years, technology has changed the way in which data collection is done. Reel-to-reel recorders gave way to cassette tape recorders, and soon afterwards the era of digital recording began. In Chapter 7, Paul De Decker and Jennifer Nycz look at the technology of conducting sociolinguistic interviews. With the goal of obtaining a significant corpus of data of high enough quality for sociophonetic analysis, the authors cover appropriate recording equipment (including minimum

requirements), technical concerns (e.g., sampling rate, group recordings), and interview storage. After reading this chapter, fieldworkers will be ready to conquer the “technical” aspects of the interview process. Vignette 7a, by Lauren Hall-Lew and Bartłomiej Plichta, gives real-life examples of recording challenges as well as practical considerations for equipment use and choice when in the field. The experiences and advice in this vignette are invaluable and are important to consider before entering the field.

While spoken language data has been a primary focus of many studies in sociolinguistics, the use of written language data, especially data collected by surveys, has also had an important place in the field. In Chapter 8, “Surveys: The Use of Written Questionnaires in Sociolinguistics,” Charles Boberg looks at the role of surveys in sociolinguistics and the strengths and weaknesses of this method. Sociolinguistic surveys have the advantage of gathering a large number of participants with a relatively quick response and collection time; however, as Boberg points out, they are limited in the type of data that they can collect. Pointing to specific survey-based studies, especially those on Canadian English, Boberg shows how surveys can be used and discusses how different methodological choices can drive the selection of variants for a survey.

Vignette 8a, by Kathryn Campbell-Kibler, extends the survey method beyond linguistic variable collection to speaker attitude and evaluation study. She shows how speaker evaluation studies are a specialized form of survey that can work alongside other sociolinguistic methods, in order to provide a more robust picture of a language variety and speaker group. Campbell-Kibler moves us through the various steps of setting up a speaker evaluation study, from stimuli to tasks and context, all the while focusing on the goal of having accurate and understandable results. Vignette 8b, by Naomi S. Baron, looks at online surveys as a method for data collection, a viable and attractive method for data collection for many researchers. Baron discusses her work in collecting online data in communities from around the world and points out variation in cultural considerations. Most importantly, she notes that we must consider the responses of our participants when we design surveys, as we do not all have the same cultural assumptions.

Closing this second part of the book, Chapter 9 looks at the use of experiments as data-generating resources. Cynthia G. Clopper begins the chapter by noting that experiments and the data that they create can either stand alone or be used alongside data generated by other collection methods or collected from other sources. Focusing on both production and perception experiments, Clopper demonstrates the utility of each method and, more importantly, the types of research questions that they can answer. She walks us through the ways that experiments can be used in the field and covers the advantages of experiments, while also cautioning us as to the disadvantages of using this approach (including limitations on the naturalness of data and having access to equipment of the quality needed) that must be thought through.

The chapters and vignettes in Part II work together to provide an overview of the process of collecting sociolinguistic data, which can be overwhelming in the choices that are available to researchers and the detail needed to consider each

methodological approach. Allowing the research question to drive the primary data collection method and even supplemental data collection will help researchers decide on a framework that is best suited to their study. With an understanding of the technological necessities for conducting ideal recordings and the potential missteps that can occur and that do occur when working with real language and real speakers, researchers will also be aware of what awaits us in the field and ready to respond, even when a miscue happens.