Data Collection in Sociolinguistics

Data Collection in Sociolinguistics: Methods and Applications is an accessible, contemporary guide to data collection, a central pursuit of sociolinguistic research. Contributions by veteran as well as up-and-coming sociolinguists cover methods of data collection, whether generating new data or when working with existing data, and tackle important questions of ethics, data sharing, data preservation, and community outreach. Other cutting-edge topics, featured in main chapters and in shorter, reader-friendly vignettes, include the use of public documents, virtual world research, collecting online data, working with video data, and cross-cultural issues in data collection. A comprehensive volume, Data Collection in Sociolinguistics is set to become an indispensable guide for students and scholars interested in both the broad outlines and the finer details of sociolinguistic data collection.

Christine Mallinson is Associate Professor of Language, Literacy, and Culture and Affiliate Associate Professor of Gender and Women’s Studies at the University of Maryland-Baltimore County.

Becky Childs is Associate Professor of English at Coastal Carolina University.

Gerard Ván Herk is Associate Professor and Canada Research Chair in Regional Language and Oral Text at Memorial University of Newfoundland.
Contents

List of Illustrations x
Foreword: Observing the Observers xi
J. K. Chambers
Acknowledgments xv
Copyright Acknowledgments xvi

PART I
Research Design 1

1 Research Design 3
Christine Mallinson

2 Ways of Observing: Studying the Interplay of Social and Linguistic Variation 7
Barbara M. Horvath

Vignette 2a. Multidisciplinary Sociolinguistic Studies 17
Marcia Farr

Vignette 2b. How to Uncover Linguistic Variables 21
Walt Wolfram

Vignette 2c. How to Uncover Social Variables: A Focus on Clans 25
James N. Stanford

Vignette 2d. How to Uncover Social Variables: A Focus on Social Class 29
Rania Habib
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Social Ethics for Sociolinguistics</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>SARA TRECHTER</td>
<td></td>
</tr>
<tr>
<td>Vignette 3a. Responsibility to Research Participants in Representation</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NIKO BESNIER</td>
<td></td>
</tr>
<tr>
<td>Vignette 3b. Conducting Research with Vulnerable Populations</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>STEPHEN L. MANN</td>
<td></td>
</tr>
<tr>
<td>Vignette 3c. Ethical Dilemmas in the Use of Public Documents</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SUSAN EHRLICH</td>
<td></td>
</tr>
<tr>
<td>Vignette 3d. Real Ethical Issues in Virtual World Research</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RANDALL SADLER</td>
<td></td>
</tr>
<tr>
<td>PART II</td>
<td>Generating New Data</td>
<td>63</td>
</tr>
<tr>
<td>4</td>
<td>Generating New Data</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>BECKY CHILDS</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Ethnographic Fieldwork</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>EREZ LEVON</td>
<td></td>
</tr>
<tr>
<td>Vignette 5a. Fieldwork in Immigrant Communities</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>JAMES A. WALKER AND MICHOL F. HOFFMAN</td>
<td></td>
</tr>
<tr>
<td>Vignette 5b. Fieldwork in Migrant and Diasporic Communities</td>
<td>84</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RAJEND MESTHRIE</td>
<td></td>
</tr>
<tr>
<td>Vignette 5c. Fieldwork in Remnant Dialect Communities</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PATRICIA CAUSEY NICHOLS</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>The Sociolinguistic Interview</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>KARA BECKER</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D. VICTORIA RAU</td>
<td></td>
</tr>
</tbody>
</table>
Vignette 6b. Conducting Sociolinguistic Interviews in Deaf Communities
CEIL LUCAS

105

Vignette 6c. Special Issues in Collecting Interview Data for Sign Language Projects
JOSEPH HILL

110

Vignette 6d. Other Interviewing Techniques in Sociolinguistics
BOYD DAVIS

114

7 The Technology of Conducting Sociolinguistic Interviews
PAUL DE DECKER AND JENNIFER NYCZ

118

Vignette 7a. Technological Challenges in Sociolinguistic Data Collection
LAUREN HALL-LEW AND BARTLOMIEJ PLICHTA

127

8 Surveys: The Use of Written Questionnaires in Sociolinguistics
CHARLES BOBERG

131

Vignette 8a. Language Attitude Surveys: Speaker Evaluation Studies
KATHRYN CAMPBELL-KIBLER

142

Vignette 8b. Cultural Challenges in Online Survey Data Collection
NAOMI S. BARON

147

9 Experiments
CYNTHIA G. CLOPPER

151

PART III

Working with and Preserving Existing Data

163

10 Working with and Preserving Existing Data
GERARD VAN HERK

165

11 Written Data Sources
EDGAR W. SCHNEIDER

169
Vignette 11a. Accessing the Vernacular in Written Documents  
FRANCE MARTINEAU  
179

Vignette 11b. Adapting Existing Data Sources: Language and the Law  
PHILIPP SEBASTIAN ANGERMEYER  
183

Vignette 11c. Advances in Sociolinguistic Transcription Methods  
ALEXANDRA D’ARCY  
187

Vignette 11d. Transcribing Video Data  
CÉCILE B. VIGOUROUX  
191

12 Data Preservation and Access  
TYLER KENDALL  
195

Vignette 12a. Making Sociolinguistic Data Accessible  
WILLIAM A. KRETZSCHMAR, JR.  
206

Vignette 12b. Establishing Corpora from Existing Data Sources  
MARK DAVIES  
210

Vignette 12c. Working with “Unconventional” Existing Data Sources  
JOAN C. BEAL AND KAREN P. CORRIGAN  
213

13 Working with Performed Language: Movies, Television, and Music  
ROBIN QUEEN  
217

Vignette 13a. Working with Scripted Data: A Focus on African American English  
TRACEY L. WELDON  
228

Vignette 13b. Working with Scripted Data: Variations among Scripts, Texts, and Performances  
MICHAEL ADAMS  
232

14 Online Data Collection  
JANNIS ANDROUTSOPoulos  
236
PART IV
Sharing Data and Findings 251

15 Sharing Data and Findings 253
CHRISTINE MALLINSON

16 Community Activism: Turning Things Around 258
ARAPERA NGAHÄ

17 Sociolinguistic Engagement in Schools: Collecting and Sharing Data 269
ANNE H. CHARITY HUDLEY

Vignette 17a. Beyond Lists of Differences to Accurate Descriptions 281
LISA GREEN

Vignette 17b. Linguistic Flexibility in Urban Zambian Schoolchildren 286
ROBERT SERPELL

Vignette 17c. Engagement with Schools: Sharing Data and Findings 290
DONNA STARKS

18 Sociolinguistics in and for the Media 293
JENNIFER SCLAFANI

Vignette 18a. Media Interest in Sociolinguistic Endeavors 304
SCOTT F. KIESLING

Vignette 18b. Sociolinguistics on BBC Radio 308
CLIVE UPTON

Vignette 18c. Media, Politics, and Semantic Change 314
ANDREW D. WONG

19 Conclusion 318
CHRISTINE MALLINSON, BECKY CHILDS, AND GERARD VAN HERK

List of Contributors 319
Index 325
Vignette 6a
Cross-cultural Issues in Studying Endangered Indigenous Languages

D. Victoria Rau

My research on Yami, a Philippine Batanic language spoken by 4,000 speakers on Orchid Island, which lies off the southeastern coast of Taiwan, began in 1994 with a personal invitation from Maa-neu Dong, who was seeking an Austronesian linguist to compile a dictionary of her mother tongue. In response, I conducted a sociolinguistic survey of this indigenous language, following an SIL method (Blair, 1990; Grimes, 1995) to gather basic word lists, texts for intelligibility tests, information on bilingual ability, language use, and language attitudes.

In the process of analyzing the word lists, several linguistic variables emerged as potential candidates for Labovian-style sociolinguistic studies, including one similar to the centralization of diphthongs on Martha’s Vineyard (Labov, 1972). The Yami diphthongs (ay) and (aw) were undergoing vowel raising on the island (e.g., mangay ~ mangey ‘go’, arow ~ arow ‘day, sun’), with an isogloss separating the more progressive northeast from the more conservative southwest. However, even though most coding of the variants in the word list was completed within three years of the initial trip, I lacked adequate understanding of both the linguistic structure of Yami and the social structure of the speech community to be able to analyze the social stratification of the change. As it was an endangered minority language, no comprehensive grammatical sketch of Yami was available at the time. Not until our *Yami Texts with Reference Grammar and Dictionary* was published (Rau & Dong, 2006) did we feel we were ready to write up the centralization of diphthongs (Rau, Chang, & Dong, 2009). Furthermore, during 2005–2009 we were able to put our Yami materials online for language conservation, including documentation (http://yamiproject.cs.pu.edu.tw/yami), e-Learning (http://yamiproject.cs.pu.edu.tw/elearn), and an online dictionary (http://yamibow.cs.pu.edu.tw).

When a “cross-cultural” investigation involves a less commonly studied, endangered indigenous minority language, practical goals of language conservation should take precedence over theoretical sociolinguistic goals. Researchers may also need to accept the frustrating reality that it is never possible to interpret the limited data as quickly, accurately, and adequately as when studying dominant languages. Below, I describe some of the sociolinguistic issues raised during my work with the Yami speech community.
Urban Dialectology vs. Endangered Indigenous Language Studies

The goals of and data collection methods for a variationist sociolinguistic study of a minority language differ from those of urban dialectology. A typical Labovian-style study seeks to address the question of social motivation of linguistic change. A valid and reliable variationist study usually has several prerequisites. First, the researcher needs to have native or near-native command of the target language, whether the researcher personally conducts sociolinguistic interviews, hires a local interviewer to match the local speech style (Trudgill, 2010), or uses other supplemental techniques (Wolfram, 2011). Second, there are usually grammatical sketches, dialect studies, or records of historical linguistics in the target language to serve as a basis for comparison. Third, the linguistic variable to be investigated has to provide sufficient stratified data in the subsystem to meet Labov’s (1972) principle of accountability and provide Tagliamonte’s (2009) three lines of evidence for VARBRUL analysis. This makes a dominant language a perfect candidate for Labovian-style variation studies.

Endangered indigenous languages are a different story (Rau, 2011). It takes much longer to develop a basic understanding of the language before studies of variation can even be attempted. Data collection is usually restricted to word lists and narratives, as the researcher’s proficiency in the language is limited. In addition, the range of linguistic variables is also compressed, as the consultants who assist in data transcription may edit out some “variations,” both to make the language look more “standard” and because the transcriber naturally transcribes in their own dialect.

Methodological Differences

How did my methods of sociolinguistic data collection on Orchid Island differ from the Labovian method? My initial data were gathered in 1994 as part of a sociolinguistic survey to establish a relationship with the community. The word lists were transcribed phonetically, but the narrative data were transcribed phonemically by my Yami consultant, who came from the non-centralized /ay/ and /aw/ dialect area. To gather more narrative data, my consultant and I went back to the community when we were commissioned to do a Yami dictionary project in 1998–2000 and language documentation project in 2005–2007. We managed to glean enough data from the same speakers who had contributed word list and narrative data in 1994 for a variation study.

Unfortunately, as the language is not being transmitted to the younger generation (Lin, 2007), we cannot test our hypothesis of change in progress. Nor did we ever have a chance to conduct a trend study or panel study (Sankoff & Blondeau, 2007), since it took us over a decade to process and understand the data gathered in 1994. To study the two diphthongs, my Yami consultant had to go back to the previously transcribed texts to recode the pronunciation of /ay/ and /aw/. Lacking sociolinguistic interviews, we treated narratives as “informal style” in contrast with the “formal” word list reading style, as defined by the degree of
attention paid to form. In our later study of word order variation, we found that narratives and conversations could be further distinguished by word order variation (Chang & Rau, 2011), but this insight came too late for our 2009 study.

Advice

On the basis of my experience with the Yami community, I recommend using a four-step approach to data collection for the purpose of producing useful sociolinguistic materials, following the principle of linguistic gratuity (Wolfram, Reaser, & Vaughn, 2008):

1. Conduct a sociolinguistic survey, gathering word lists and narratives.
2. Write a reference grammar and teaching materials as part of a language conservation effort to “give back” to the community under study.
3. Identify potential linguistic variables to contribute to both practical and theoretical issues. For example, our variation study of the two diphthongs (Rau et al., 2009) has provided the theoretical underpinning for orthography development in Yami. A study of Yami word order (Chang & Rau, 2011) has led us to understand how narrative and conversation styles can account for word-order variation between VS and SV. A recent study of the variation between path verbs and manner verbs in Yami (Rau, Wang, & Chang, 2012) has increased our understanding of motion events in cognitive linguistics.
4. Prepare to write a user-friendly socio-grammar (Nagy, 2009). A useful pedagogical grammar with a focus on language use suitable for indigenous language teacher training programs would be highly appreciated, as a linguist’s grammar is usually perceived as incomprehensible or irrelevant to indigenous teachers who are teaching their language in a school setting for language conservation.

As studies on minority speech communities require a lifelong commitment, I hope this “been there, done that” account will give researchers a firm and practical foundation when collecting cross-cultural data.

References


