

Christine Mallinson, Becky Childs, Gerard van Herk (eds.). 2018. *Data Collection in Sociolinguistics*, 2nd edition. London and New York: Routledge. 325 pp.

Reviewed by Costin-Valentin Oancea*

The volume under review, *Data Collection in Sociolinguistics*, is edited by Christine Mallinson, Becky Childs and Gerard van Herk and published by Routledge. The book is structured into four parts: “Research design”, “Generating new data”, “Working with and preserving existing data”, and “Sharing data and findings”. At the end of the book the editors include an “Index” (pp. 319-325).

Part I, “Research Design” (pp. 1-64), tackles two of the most important pillars regarding data collection in sociolinguistics: research design and ethics. In chapter 2, “Ways of observing: Studying the interplay of social and linguistic variation”, Barbara M. Horvath presents the frameworks and methods used in sociolinguistic research. The author makes reference to different disciplines (e.g. sociology, geography, psychology, anthropology) which have shaped modern-day sociolinguistics. As she puts it, “from geography they borrow regional studies, maps and the concept of place, and from sociology they borrow community studies, social survey methods, and social network analysis.” Horvath also discusses the general requirements for data collection in quantitative, as well as qualitative sociolinguistic research.

In vignette 2a, “Multidisciplinary sociolinguistic studies”, Marcia Farr presents the pros and cons of interdisciplinary and multidisciplinary research and provides as an example her own study of transnational Mexican families.

In vignette 2b, “How to uncover linguistic variables”, Walt Wolfram discusses the importance as well as the necessity to choose and analyse linguistic variables that are not part of the canonical set, but which might shed light on sociolinguistic variation. Two cases are presented: *a*-prefixing in Appalachian English and the *call oneself* construction found in African American English.

Vignette 2c, “Studying difficult to study variables”, by J. Daniel Hasty, focuses on one morphosyntactic variable which is hard to find – the double modal in the medical consultation (in which doctors talk to their patients face to face). He uses a corpus containing over 45,000 fully transcribed and searchable audio recordings collected and maintained by Verilogue Inc.

Vignette 2d, “How to uncover social variables: A focus on clans”, by James N. Stanford, describes social variables in reference to the indigenous Sui people of Guizhou Province, China. Anthropological information abounds and the author also provides a few suggestions to help uncover locally meaningful social variables: (i) be engaged with the community and personally involved in local life as much as possible; (ii) let go of prior assumptions; (iii) depend on the insights of cultural insiders.

The last vignette, “How to uncover social variables: A focus on social class”, by Rania Habib, highlights the intricate relation between rural social uniformity and urban social uniformity, and presents a study on the variables (q) and (e) in child and adolescent language in the village of Oyun Al-Wadi in Syria.

Chapter 3, “Social ethics for sociolinguistics”, by Sara Trechter, raises an important issue: “At what point does an observer become a quasi community member?”. She makes reference to Wolfram’s (1998) “Principle of Linguistic Gratitude”, which urges the researcher to share his/her expertise and knowledge with the host research community. Trechter advocates that researchers give back to the community and she does this by drawing on her experience as a member of the Linguistic Society of America.

* “Ovidius” University, Constanța, oancea_costin@yahoo.com.

In vignette 3a, “Responsibility to research participants in representation”, Niko Besnier presents his research in Central Pacific. Even though gossip is quite a well-researched topic in sociolinguistics, the author shows that it can be a sensitive one for the participants.

Vignette 3b, “Working with transgender communities”, by Lal Zimman, offers guidance to academics who are interested in pursuing this line of research. Zimman urges the reader to keep an open mind and get rid of preconceived ideas, as well as to be cautious with the language we use when we talk to transgender people.

In vignette 3c, “Conducting research with vulnerable populations”, Stephen L. Mann mentions the conundrums he encountered in his research on gay and queer men in the Southern United States of America.

In vignette 3d, “Ethical dilemmas in the use of public documents”, Susan Ehrlich talks about her research on the discourse of women who have been complainants in rape trials. Ehrlich succeeds in providing a solution as to how to protect participants from misrepresentations.

The last vignette, which ends part one, “Real ethical issues in virtual world research”, by Randall Sadler, tackles the ethical dilemmas which can appear when collecting data online. The author provides different tips on how to use pseudonyms, obtain informed consent, etc.

The second part of the book, “Generating new data” (pp. 65-162), discusses at length the intricate methods and stages in collecting data and using different corpora for sociolinguistic analysis.

Chapter 5, “Ethnographic fieldwork”, by Erez Levon, highlights the ethnographic data collection method. The author objectively presents the strengths and weaknesses of this method, and provides relevant examples. The chapter guides the reader through the four main principles for conducting ethnographic fieldwork: (i) accessing a community; (ii) interacting with participants; (iii) data collection; (iv) leaving the community.

In vignette 5a, “The joy of sociolinguistic fieldwork”, John R. Rickford provides, through personal examples, the actual joys of conducting sociolinguistic research. He stresses the idea that the research participants share with the researcher the highs, lows and small delights of their lives and this might have a profound effect on the fieldworker.

In vignette 5b, “Fieldwork in immigrant communities”, James A. Walker and Michol F. Hoffman start their discussion by identifying the reasons of focusing on an immigrant community. They also present different techniques of entering a community, e.g. “the friend of a friend technique”. Included here are ways of collecting a corpus by means of different methods.

In vignette 5c, “Fieldwork in migrant and diasporic communities”, Rajend Mesthrie shares with the reader his experiences in working with a migrant and diasporic community in South Africa, the KwaZulu-Natal.

Vignette 5d, “Fieldwork in remnant dialect communities”, by Patricia Causey Nichols, starts with a description of the study of remnant dialect community provided by Wolfram (2004: 84), who claims that such a community “retains vestiges of earlier language varieties that have receded among speakers in the more widespread population.” The author presents a quantitative research design which concentrates on morphosyntactic features of Gullah. Nichols succeeded in identifying and meeting the people’s agenda (for five months) which resulted in a wealth of qualitative data, to further strengthen the quantitative analysis.

In vignette 5e, “Linguistic landscape and ethnographic fieldwork”, Jackie Jia Lou gives the reader insight into how research on linguistic landscape can be carried out. She argues that a researcher needs to decide upon a unit of analysis, geographic boundaries, length of engagement, etc.

Chapter 6, “The sociolinguistic interview”, by Kara Becker, focuses on the most important tool in variationist sociolinguistics. She starts by defining the sociolinguistic interview, and then she moves on to the utility of the sociolinguistic interview, in the Labovian variationist paradigm.

Vignette 6a, "Cross-cultural issues in studying endangered indigenous languages", by Victoria Rau, discusses the use of the sociolinguistic interview with reference to an endangered language, i.e. Yami, a Philippine Batanic language spoken by 4,000 speakers on Orchid Island (Taiwan). The author ends the vignette with a four-step approach to data collection.

Vignette 6b, "Conducting sociolinguistic interviews in deaf communities", by Cecil Lucas, describes the process of conducting a successful sociolinguistic interview in a Deaf community. The author argues that a very important component of data collection is the selection of the subjects. Another important aspect mentioned is the impact that new technologies have on how sign language data can be collected.

In vignette 6c, "Special issues in collecting interview data for Sign Language projects", Joseph Hill explains the role of the Observer's Paradox as well as the signer's sensitivity to the interlocutor's ethnicity.

Vignette 6d, "Other interviewing techniques in sociolinguistics", by Boyd Davis, highlights alternative methods used for collecting data.

Chapter 7, "The technology of conducting sociolinguistic interviews", by Paul De Decker and Jennifer Nycz, delves into the technical aspects of the sociolinguistic interview. They advise the researcher to use digital recorders, and provide relevant examples of different types of recorders. The presentation of these "technicalities" is done in a remarkably objective way, with the purpose of obtaining high quality data.

Vignette 7a, "Technological challenges in sociolinguistic data collection", by Lauren Hall-Lew and Bartłomiej Plichta, offers personal examples of recording problems as well as important tips for equipment use and choice in the field.

Chapter 8, "Surveys: The use of written questionnaires in sociolinguistics", by Charles Boberg, explains the role of surveys in sociolinguistic research as well as the pros and cons of this method. The author focuses on survey-based studies, more precisely on Canadian English and argues that one of the advantages of surveys is the quantity and one disadvantage is the quality.

In vignette 8a, "Language attitude surveys: Speaker evaluation studies", Kathryn Campbell-Kibler shows that the stimuli represent the heart of the experiment. The author highlights that speaker evaluation studies are actually a specialised form of survey which can work together with other sociolinguistic methods.

Vignette 8b, "Cultural challenges in online survey data collection", by Naomi S. Baron, tackles online survey tools (stand alone products like SurveyMonkey or surveys embedded in Facebook) which facilitate the collection of larger and more diverse samples.

Vignette 8c, "Dialect surveys and student-led data collection", by Laurel MacKenzie, describes various approaches to large-scale data collection in which she involves her students.

The second part ends with Chapter 9, "Experiments", by Cynthia G. Clopper. The author discusses production as well as perception experiments in sociolinguistics. She shows the effectiveness of each method and the types of research questions that they answer.

The third part of the book, "Working with and preserving existing data" (pp. 163-252), provides an in-depth analysis of the challenges of adapting data to the needs of sociolinguists.

Chapter 11, "Written data sources", by Edgar W. Schneider, looks into the usage of written data for sociolinguistic analysis. The author deems writing a cultural artefact. Different methodological issues and concerns relevant in such an investigation are under scrutiny. The chapter ends with a list of reasons for investigating written data sources.

Vignette 11a, "Accessing the vernacular in written documents", by France Martineau, focuses on features of written vernacular.

In vignette 11b, "Adapting existing data sources: Language and the law", Philipp Sebastian Angermeyer presents different data sources in forensic linguistics, and considers also the ethical component regarding the use of the data.

Vignette 11c, “Issues in forensic linguistic data collection”, by Ronald R Butters continues the discussion started by Angermeyer and talks about the goals of forensic linguistics.

The discussion swiftly changes to “Advances in sociolinguistic transcription methods”, by Alexandra D’Arcy, which is the topic of vignette 11d. The author presents state-of-the-art software programs used in transcribing data (e.g. CLAN, ELAN, EXMARaLDA, Praat, Transcriber, Transana, VoiceMaker). The author also stresses the idea that orthographic files can be time-aligned with audio files.

Vignette 11e, “Transcribing video data”, by Cécile B. Vigouroux, outlines the “protocol” which has to be followed when transcribing data. Vigouroux argues that if the researcher decides to video record rather than audio record, then all that visual information must be included somehow in the transcription.

Chapter 12, “Data preservation and access”, by Tyler Kendall, reviews several of the issues involved in preserving and maintaining access to sociolinguistic data. The author also raises an important question, namely, how can other researchers use relevant sociolinguistic data?

In vignette 12a, “Making sociolinguistic data accessible”, William A. Kretzschmar Jr. guides the reader through the intricacies of making our own audio or video recordings available to future generations of researchers, taking into account the policies for the protection of human subjects. Such policies urge the destruction of such data.

Vignette 12b, “Establishing corpora from existing data sources”, by Mark Davies, continues the discussion commenced by Kretzschmar. Davies makes reference to two of the most important corpora in the English language – the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA), created in less than a year, but which was based on pre-existing materials and data.

In vignette 12c, “Working with “unconventional” existing data sources”, Joan C. Beal and Karen P. Corrigan share their experience of working with data collected at various times using different methods and methodologies in order to produce the Newcastle Electronic Corpus of Tyneside English.

Chapter 13, “Working with performed language: Movies, television, and music”, by Robin Queen, tackles the recent media discussions of “vocal fry”, starting from a quote by Dahl (2011: 12) who states that “more college women speak in creaks, thanks to pop stars.” The remainder of the chapter delves into the ways in which performed media can represent a good source of data about language.

In vignette 13a, “Working with scripted data: A focus on African American English”, Tracey L. Weldon draws on her experience and on the experience of other researchers and focuses on one of the most important dialects in the US, i.e. African American Vernacular English.

Vignette 13b, “Working with scripted data: Variations among scripts, texts, and performances”, by Michael Adams, investigates the issue of negotiation and change between original scripts and released materials. The author stresses the idea that scripted works in mass media seem to be stable and reliable sources of linguistic evidence.

Chapter 14, “Online data collection”, by Jannis Androutsopoulos, reviews research on computer-mediated communication (CMC) in linguistics and presents the characteristics of online language (text). Included here is a discussion of data sampling in the Computer-Mediated Discourse Analysis Framework. The chapter ends with a note on research ethics.

Vignette 14a, “Sociolinguistic approaches to storytelling in Facebook status updates”, by Ruth Page, stresses the importance of storytelling practices posted online. Page provides examples from Facebook, Twitter as well as Wikipedia.

The discussion continues in vignette 14b, “Collecting data from Twitter”, by Steven Coats, with an emphasis on automated data extraction, and different technical information regarding the online platform Twitter.

The last part of the volume, “Sharing data and findings” (pp. 253-318), tackles concepts and various techniques used to collect sociolinguistic data from a wide variety of places and populations, as well as sharing the results and findings with the community.

Chapter 17, ‘Sociolinguistic engagement in schools: Collecting and sharing data’, by Anne H. Charity Hudley, surveys the methods used to conduct successful research in schools. The author identifies models of sociolinguistic engagement for researchers who want to collect data and share data with those in schools.

Vignette 17a, “Beyond lists of differences to accurate descriptions”, by Lisa Green, sketches the development of language used by children who grew up in non-mainstream American English speech communities.

In Vignette 17b, “Linguistic flexibility in urban Zambian schoolchildren”, Robert Serpell illustrates the complexity of the linguistic repertoire which children need in order to reach full communicative competence.

Vignette 17c, “Engagement with schools: Sharing data and findings”, by Donna Starks, explores some of the problems that researchers might encounter while collecting data in schools, and briefly mentions the Pasifika Languages of Manukau Project.

Chapter 18, “Sociolinguistics in and for the media”, by Jennifer Sclafani, puts forward the idea of engagement with the media. The author brings forth two case studies, the 1996 Ebonics controversy and the situation of terminology focused on immigrants working illegally.

In vignette 18a, “Media interest in sociolinguistic endeavors”, Scott F. Kiesling presents a few remarks on the use of *OMG* and *WTF* as well as the status of *dude*.

Vignette 18b, “Sociolinguistics on BBC Radio”, by Clive Upton, describes the BBC Voices Project (2004-2007) and some of the findings.

In Vignette 18c, “Media, politics, and semantic change”, Andrew D. Wong analyses the semantic change of the Chinese label *tongzhi* ‘comrade’ is under scrutiny. The author developed a corpus of articles from the *Oriental Daily News* (ODN), the most widely circulated newspaper in Hong Kong.

The last vignette, “Engaging local and mass media on issues of language policy”, by Phillip M. Carter, emphasizes ways in which local mass media can be involved in issues of language policy. The author makes reference to his work in Miami, Florida.

The book *Data Collection in Sociolinguistics*, edited by Christine Mallinson, Becky Childs and Gerard van Herk, represents an excellent tool for anyone interested in sociolinguistic variation and change as well as fieldwork. The case studies presented here offer a fresh and interesting perspective to the even changing field of sociolinguistics. The authors focus on their personal experience, tackling the conundrums they faced as well as their successes in the field. The communities investigated range from Gullah, to Zambia and Yami. Each part of the volume delves into a particular topic relevant to the field of (socio)linguistics. For achieving all of these, the editors as well as the authors who contributed to this volume deserve congratulations.

References

- Dahl, M. n.d. More college women speak in creaks, thanks to pop stars. *The Body Odd*. http://bodyodd.msnbc.msn.com/_news/2011/12/12/9393348-more-college-women-speak-in-creaks-thanks-to-pop-stars.
- Wolfram, W. 1998. Scrutinizing linguistic gratuity: A view from the field. *Journal of Sociolinguistics* 2: 271-279.
- Wolfram, W. 2004. The sociolinguistic construction of remnant dialects. In C. Fought (ed.), *Sociolinguistic Variation: Critical Reflections*, 84-106. New Work: Oxford University Press.

All rights reserved; partial or total reproduction of this text, its multiplication in any form and by any means – electronic, mechanical, photocopying, recording or otherwise –, public delivery via internet or other electronic net, storage in a retrieval system for commercial purposes or charge-free, as well as other similar actions committed without the written permission of the publishers represent violations of the intellectual property rights legislation and are punishable under the criminal and/or civil law in force.

Tiparul s-a executat sub c-da nr. 4521 / 2019
la Tipografia Editurii Universității din București
B-dul Iuliu Maniu, 1-3, Complex Leu
Tel.: 0799 210 566,
E-mail: tipografia.unibuc@unibuc.ro, mihaela.stancu@unibuc.ro