1 Introduction

Devyani Sharma and Robert J. Podesva

> I would like to discuss an approach to the mind that considers language and similar phenomena to be elements of the natural world, to be studied by ordinary methods of empirical inquiry.
>
> Noam Chomsky 1995

> Linguists have forgotten, Mathesius argued, that the homogeneity of language is not an 'actual quality of the examined phenomena,' but 'a consequence of the employed method'.
>
> Uriel Weinreich, William Labov, and Marvin I. Herzog 1968

> Some have seen in modern linguistic methodology a model or harbinger of a general methodology for studying the structure of human behavior.
>
> Dell Hymes 1962

1 Overview

The three views expressed above remind us of the peculiar status of linguistics as a field. It represents a single discipline to the extent that it broadly shares a single object of analysis, but little else can be said to be uniform in terms of epistemology and method. Some linguists affiliate most closely with the social sciences, others with the natural sciences, and others with the humanities. Perhaps surprisingly, this diverse group has not (yet) splintered off into separate fields. Rather, the deep heterogeneity of the field has come to be seen by many as a strength, not a weakness. Recent years have witnessed a rise in creative synergies, with scholars drawing inspiration from the methods and data used by "neighboring" linguists in order to enrich and expand the scope of their own investigations.

This has occurred in part due to constant refinement of methodologies over time, leading to more clearly specified methodological norms within all subfields of linguistics, which in turn facilitate more targeted cross-fertilization and exchange. Sharing of methodological practices may take the form of bridge-building across subfields of linguistics, or exchanges of methods between linguistics and related fields. Bridge-building has taken many forms in recent work; a few examples include the adoption of corpora, experiments, and statistical measures in formal analysis (e.g., Bresnan and Ford 2010), the adoption of experimental methods in sociolinguistics (e.g., Campbell-Kibler 2009) and

1

pragmatics (e.g., Breheny 2011), and the use of sociolinguistic sampling in laboratory phonology (Scobbie and Stuart-Smith 2012). Similarly, though less central to the present collection, methods have been exchanged profitably with neighboring fields as well, such as the borrowing into linguistics of genetic modeling (McMahon and McMahon 2005), clinical imaging (e.g., Gick 2002; Skipper and Small 2005; Martins et al. 2008), and sociological sampling principles (e.g., Milroy 1980; Eckert and McConnell-Ginet 1992).

In this climate of creative collaboration and innovation, particularly in methodology, the dearth of general reference texts on methods used in linguistics is striking. Almost all current overviews of methodology are specific to particular subdisciplines (one exception is Litosseliti 2010, a shorter volume than the present one). Specialization naturally permits greater depth and detail, and these texts are indispensable for research in specific fields. But at present, few of these texts are able to foreground insights and principles that should ideally be adopted across the field, nor foster interdisciplinary methods. It is still common in core areas of theoretical linguistics for courses to omit training in research methods at all, and for courses in other areas to only review methods in that field. New researchers in linguistics often complete their training without any exposure to entire methodological paradigms – for example, experimental methods, methods for elicitation, statistics, or ethnography – many of which could strengthen their research contributions. If students do expand their training, this often occurs much later and haphazardly. This collection aims to offer a wider overview and a more diverse toolbox at the outset of methods training.

Given the reality of greater cross-fertilization in linguistics today, a general reference text can also better reflect the exciting state of the field today, and can help promote recent trends and best practices in contemporary methods. The present collection has been developed with this goal in mind. It is intended for use in the training of advanced undergraduate and graduate students, while also serving as a reference text for experienced linguists embarking on research involving new domains, or researchers who simply wish to expand their repertoire or familiarity with methods used to solve general problems in research, such as eliciting language forms, sampling subjects, designing experimental tasks, or processing raw data.

2   Structure of the book

Given the extraordinary complexity of the field of linguistics, it is naturally impossible to provide a comprehensive introduction to all methodologies in a single volume. This collection is designed to be comprehensive in breadth, but not necessarily in depth. Each chapter incorporates suggestions for further reading, to enable readers with specific interests or questions to selectively extend their knowledge.

The volume follows a structure designed to encourage users – teachers, students, researchers – to take a wider view of questions of methodology and to aim for a more comprehensive set of methodological skills than simply those typically adopted for a particular subfield or question. While best practices tend to arise independently within each of the subfields of linguistics, many of the same issues surface from one subdiscipline to the next, and projects on seemingly unrelated topics often require attention to the same methodological concerns. For example, transcription is a concern for nearly all linguists, from those who focus on the details of speech articulation to those who analyze turn-taking in conversation. Similarly, any researcher who collects data from human subjects needs to devote attention to the composition of their sample. And all linguists who record speech, whether the recordings are intended for acoustic analysis, for survey materials, or simply as records of data elicitation sessions, need to grapple with many of the same issues. Finally, regardless of which statistical approaches dominate any given subdiscipline, the underlying theoretical assumptions are the same. For this reason, the volume is not strictly organized by subdiscipline, but rather follows the trajectory of any research project, highlighting general issues that arise in three core areas: data collection, data processing, and data analysis.

Part I (Data collection) focuses on types of data used in linguistics and best practices in the collection of each data type. The section starts, in Chapter 2, with Eckert's discussion of ethical issues that must be considered at the start (and for the duration) of any research project. The next two chapters cover methods of data collection that require working with relatively small numbers of informants. In Chapter 3, Schütze and Sprouse review recent methods used in the collection of grammaticality judgment data, and Chelliah guides the reader through the process of eliciting data for language description and documentation in Chapter 4. Researchers who work with larger numbers of participants must devote special consideration to sampling, discussed by Buchstaller and Khattab in Chapter 5. One special population that linguists frequently work with is children; we do not devote a specific chapter to children, but rather discuss issues pertaining to children as subjects where relevant (in Chapter 2, Chapter 5, and in the discussion of longitudinal data in Chapter 22, discussed below). Researchers might collect data from selected population samples using a wide array of instruments. While Chapters 3 and 4, described above, deal with close elicitation from fewer numbers of participants, other methods often use larger groups. These include surveys and interviews, discussed by Schilling in Chapter 6, experiments, as reviewed by Abbuhl, Gass, and Mackey in Chapter 7 and by Kaiser in Chapter 8, and ethnography, as discussed by Levon in Chapter 10. Levon's discussion of how to work with community members is an important concern not only for ethnographers, but for anyone who might need to make ties with community members for research purposes. The final three chapters of the section turn their attention to how to collect some of the most common media that linguists work with. Podesva and Zsiga cover methods for making sound and articulatory recordings in Chapter 9; Levon discusses concerns with working with video data in Chapter 10,

and van Kemenade and Los discuss many of the challenges of working with textual data in Chapter 11.

Data often need considerable preliminary processing before analysis and interpretation. Part II (Data processing and statistical analysis) deals with some of the common challenges of processing data once they have been collected. Since transcription is involved in many subfields of linguistics, Nagy and Sharma provide common practices and recommendations in Chapter 12. Corpora are also increasingly used for a range of formal and variationist analyses, and Gries and Newman review how to construct and extract data from corpora in Chapter 13. The final three chapters of Part II focus on statistics, as linguists are now more than ever adopting a quantitative approach to linguistic analysis. Johnson describes the fundamentals of characterizing the basic distributional properties in a set of data in Chapter 14. Gries moves from descriptive statistics to the selection of appropriate statistical tests and how to execute them in Chapter 15. And Part II concludes with Chapter 16, in which Baayen covers the most widely used multivariate statistics, including multiple regression. Chapters 14–16 are cumulative; each is written under the assumption that readers will be familiar with concepts discussed in earlier statistics chapters.

Finally, Part III (Foundations for data analysis) is included because we believe that in many subfields of linguistics, such as theoretical argumentation or discourse analysis, the analytic process is itself a method that should be taught systematically. Furthermore, the intended analytic method directly informs data collection and processing stages, so these should be planned together. This third part is not intended as a comprehensive overview of theoretical approaches in linguistics, but rather a practical guide to analytic methods in major areas. In Chapter 17, Boersma reviews the fundamentals of speech acoustics and its most useful forms of representation for linguists. In Chapter 18, Beavers and Sells outline the incremental process of building a reasoned argument in theoretical linguistics. Computational models have increasingly become a key method for testing linguistic theories, and desiderata for a robust and reliable approach to developing such models are set out by de Boer and Zuidema in Chapter 19. The next two chapters cover analytical approaches in sociolinguistics, from the mostly quantitative approaches that dominate variation, reviewed by Walker in Chapter 20, to the mostly qualitative approaches that are prevalent in the study of discourse, discussed by Ehrlich and Romaniuk in Chapter 21. Part III concludes in Chapter 22 with Blondeau's comparison of synchronic and diachronic methods for analyzing language over time in different subfields of linguistics.

The volume will be supplemented with a companion website, so that links to more time-sensitive online resources can be made available to users of the book, and information on technological advances in software and equipment can be updated regularly.

As noted, the present collection is designed to enhance research in all subdisciplines of linguistics by sharing best practices relevant to shared challenges. We have aimed to facilitate this style of use with highlighted keywords within

chapters, detailed cross-referencing across chapters, as well as a detailed index, and a few illustrations, described next, of how researchers might use the book.

3   Sample projects

In order to exemplify a few of the many ways this volume might be utilized, we briefly describe a few sample projects here, involving typical research questions in various subfields of linguistics. These examples highlight a few reasons why thinking about methodology strictly in terms of subdisciplines can be limiting, and demonstrate that a research project from any subdiscipline might make use of several different chapters spanning the three sections of the book.

3.1   Phonological analysis of an understudied language

Research on an understudied language requires finding language consultants (Chapter 5), attending to the ethical issues associated with working with human subjects (Chapter 2) and observing the community (Chapter 10). Upon entering the community, the student will focus primarily on language documentation (Chapter 4), which will likely draw on various forms of speaker introspection (Chapter 3). To facilitate note-taking, students may elect to record elicitation sessions (Chapters 9, 10), and resultant recordings can subsequently be used in acoustic analysis (Chapter 17). Depending on the phonological and phonetic phenomena under investigation, students may find it useful to collect articulatory data, such as static palatography, in order to ascertain the place of articulation of a sound (Chapter 9). Upon returning from the field, students could transcribe and construct a searchable corpus (Chapters 12, 13). Finally, students can refer to best practices in constructing a phonological analysis (Chapter 18), and phonological claims could be supported by identifying statistical trends in acoustic data or distributional facts gleaned from the corpus (Chapters 14–16).

3.2   Social analysis of variation or code-switching

An investigation of the social meaning of a linguistic feature or practice may take a qualitative and/or quantitative approach, any of which requires attention to ethical considerations (Chapter 2). Students focusing on patterns of production would likely need to establish contacts in a community (Chapter 10), decide which speakers to interview (Chapter 5), and audio-record interviews (Chapters 6, 9). In the context of ethnographic fieldwork (Chapter 10), students might find it helpful to ask speakers to introspect about their own and others' language use (Chapter 3) or to video-record naturally occurring interactions (Chapter 10). Students embarking on more quantitative investigations will likely need to sample a population appropriately (Chapter 5), transcribe data (Chapter 12), use basic scripting to extract and properly format data

(Chapter 13), and conduct statistical analyses (Chapters 14–16). Those who examine the strategic use of linguistic features may pay special attention to interactional factors – for example, considering how variables occur over the course of conversations or structure narratives (Chapter 21). Students examining social meaning from the perspective of perception may employ surveys and questionnaires (Chapter 6) and/or experiments (Chapters 7, 8), both of which require careful sampling (Chapter 5) and statistical analysis (Chapters 14–16). Finally, if the researcher is studying a phonological feature, the researcher will usually rely on basic principles of acoustic analysis (Chapter 17) and/or formal argumentation (Chapter 18) to establish the linguistic constraints on variation.

3.3   Analysis of a set of syntactic constructions

Currently, a student wishing to pursue a syntactic study rarely gets detailed input regarding research methods. However, this book may open up important methodological questions for such a project. First, the student would have to consider what combination of data would be appropriate for the given research question, including researcher or consultant intuition (Chapters 3, 4), corpus data (Chapters 11, 13), and experimental data (Chapters 7, 8) from an appropriate sample (Chapter 5). In building an analysis (Chapter 18), relevant glossing conventions might be used (Chapter 12) and quantitative measures requiring statistical analysis (Chapters 14–16) may be appropriate if several speakers have been consulted or if experimental methods were chosen.

3.4   Longitudinal study of second language acquisition

A study of change in an individual learning a second language could take a number of forms. The first consideration for the researcher would be what type(s) of learner to study (Chapters 5, 10) and the associated ethical implications (Chapter 2), whether to combine naturalistic and elicited data from that individual (Chapters 6, 7, 8, 10), and whether to use written or recorded responses (Chapters 6, 9, 10). The student should be advised to consider all aspects of longitudinal analysis (Chapter 22) before collecting any data. For instance, they may decide to favor data that can be replicated in a repeat run after a given period of time has elapsed. The subsequent analysis could rely on relevant chapters from Part III, depending on the nature of the linguistic phenomenon under investigation, and the student could refer to the chapters on statistics (Chapters 14–16) if they need to select appropriate quantitative measures for their analysis.

4   Concluding remarks

Every linguist would probably design a volume on linguistic methodology differently. Our focus in this collection has been on certain stages

universally shared by any research project – data gathering, data processing, and data analysis. These shared concerns are prioritized over disciplinary divisions, but we have aimed to keep methodological norms within different fields of linguistics accessible and visible as well. We hope that the wide-ranging expertise shared by the contributing authors here will continue to support the creative new methodological practices we are already witnessing in the field.

References

Breheny, R. 2011. Experimentation-based pragmatics. In W. Bublitz and N. Norrick, eds. *Handbook of Pragmatics, Volume 1: Foundations of Pragmatics*. Berlin: Mouton de Gruyter, 561–86.

Bresnan, J. and M. Ford. 2010. Predicting syntax: processing dative constructions in American and Australian varieties of English. *Language* 86.1: 186–213.

Campbell-Kibler, K. 2009. The nature of sociolinguistic perception. *Language Variation and Change* 21.1: 135–56.

Chomsky, N. 1995. Language and nature. *Mind* 104.413: 1–61.

Eckert, P. and S. McConnell-Ginet. 1992. Think practically and look locally: language and gender as community-based practice. *Annual Review of Anthropology* 21: 461–90.

Gick, B. 2002. The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association* 32: 113–21.

Hymes, D. 1962. The ethnography of speaking. In T. Gladwin and W. C. Sturtevant, eds. *Anthropology and Human Behavior*. Washington, DC: Anthropology Society of Washington.

Litosseliti, L. 2010. *Research Methods in Linguistics*. London: Continuum.

Martins, P., I. Carbone, A. Pinto, and A. Teixeira. 2008. European Portuguese MRI based speech production studies. *Speech Communication* 50: 925–52.

McMahon, A. and R. McMahon. 2005. *Language Classification by Numbers*. Oxford University Press.

Milroy, L. 1980. *Language and Social Networks*. London: Basil Blackwell.

Scobbie, J. and J. Stuart-Smith. 2012. The utility of sociolinguistic sampling in laboratory-based phonological experimentation. In A. Cohn, C. Fougeron, and M. Huffman, eds. *The Oxford Handbook of Laboratory Phonology*. Oxford University Press, 607–21.

Skipper, J. I. and S. L. Small. 2005. fMRI studies of language. In K. Brown, ed. *The Encyclopedia of Language and Linguistics*, 2nd edn. Oxford: Elsevier Science.

Weinreich, U., W. Labov, and M. Herzog. 1968. Empirical foundations for a theory of language change. In W. Lehmann and Y. Malkiel, eds. *Directions for Historical Linguistics*. Austin: University of Texas Press, 95–188.

PART I

Data collection