

# Introduction

The purpose of this manual is to introduce the concepts, principles, and procedures of a unique field of linguistic study, that of language acquisition. Our objective is to provide an overview of scientific methods for the study of language acquisition and to present a systematic, scientifically sound approach to this study. We hope to lead the reader to a greater understanding of the subject matter while providing him or her with the foundations to build a new body of knowledge through the scientific generation and analysis of new data.

Specifically, we intend to provide the background for one to be able to answer the following questions:

- How can one study language and language acquisition scientifically?
- What empirical methods are used in the scientific study of language acquisition? We concentrate on the simplest of methods—those that do not require complex equipment (other than recording equipment) and those on which other more complex methods are built, those which are useful with very young children but also potentially useful with a full developmental range. In particular, we favor those that can be used crosslinguistically. The principles of the scientific method we introduce for simple tasks carry over to more specialized tasks.
- How is the concept of *data* defined in the study of language?
- How does one generate such data?
- How does one prepare language data so one can process and analyze it once it has been collected?
- What methods and processes ensure that one's inductions or inferences are best grounded scientifically when designing or participating in language studies?
- How can principles of data management and insurance of data replicability best be developed?

Our emphasis herein is on methods for directly assessing a learner's language and on primary research designed to test specific hypotheses regarding this language knowledge. Thus, we do not provide a review of various standardized tests in the field or of various caretaker assessment tools. We deal with the most fundamental methods for primary language assessment to introduce basic principles and procedures. We do not attempt to provide a comprehensive review of the more specialized methods for language assessment currently available (e.g., "on-line" processing methods) or of numerous variations on tests of language comprehension and production. Nor do

we attempt to provide what a “statistics” manual would provide. Rather, we aim to articulate the basic principles that underlie various statistical models for the conversion of complex collected raw data to the quantitative analyses of those data. We introduce the reader to principles of the “data pipeline” that must underlie successful and scientifically sound research in this area (e.g., Leek & Peng, 2015). Armed with these basic principles, the user of this manual can be prepared to pursue the numerous variations in statistical modeling that are available to the field.

We begin by providing general principles of research in the area of the language sciences and scientific methods. We introduce basic methods and principles for data collection. We then move on to provide examples of specific approaches to data capture, such as elicitation of language production or language comprehension, and we present basic data analyses. This manual concludes by providing general approaches to data management and collaborative data sharing.

Although most of the tasks and procedures we introduce are aimed at children who have begun to produce and comprehend speech, we include a chapter that briefly reviews methodology for the study of infants, often prior to the production or comprehension of speech. We do this because of the many important results being obtained today through the study of early infant discrimination of language and because we believe these forms of infant studies, requiring specialized methods, can provide critical developmental continuity in the study of language development.

This manual differs from previously published ones in several ways. First, it aims to explicate fundamental principles and practices of applying the scientific method to study of linguistic data. Although the research methods presented here apply to many areas in language acquisition, the examples come mainly from studies of the acquisition of syntax and morphology and mainly from first language acquisition. However, we use these examples to explicate more general principles of the scientific method applied to linguistic data. The principles and methods we introduce have been easily extended to the study of second- (or more) language acquisition in children and adults and to other areas of language knowledge.

Second, this volume is aimed at new researchers. This may include students or teachers of students in this area or advanced scholars who are new to primary research in language acquisition. It does not assume any knowledge of what constitutes research data or proper experimental methodology and explains in detail the basic principles underlying scientific research. It does not assume more than a general familiarity with linguistics and linguistic terms, although for anyone who wants to become truly involved in the study of language acquisition, a basic course in linguistics, at minimum, is advisable. An attempt has been made here to provide introductory definitions of critical linguistic concepts. Similarly, we recommend to anyone wishing to apply experimental methods at least an introductory course in statistics.

Third, this volume explicitly presents practical aspects of lab work and data management. These practical aspects are usually absent from research manuals, but they constitute fundamental, basic knowledge for anyone intending to work in this field.

Last but not least, the importance of collaborative research in our field is stressed. The study of language acquisition must reach beyond the study of English alone, and it must include comparative cross-linguistic studies. For this, collaboration among researchers across languages is critical. In addition, the study of language acquisition is inherently interdisciplinary; researchers from across fields (e.g., psychology, linguistics, neuroscience) must develop an infrastructure for shared research designs, shared data, and shared collaborative data analyses. Linguists are usually not trained in experimental methods, and scholars trained in experimental methods are often not trained in linguistics. Thus, interdisciplinary team cooperation is essential, and an infrastructure allowing this collaborative sharing of the research endeavor is essential to facilitate this.

## Overview of the Book's Contents

A brief chapter-by-chapter outline is presented next, highlighting the main concepts we wish to convey.

Part I describes some fundamental concepts in language acquisition research. Chapter 1 outlines the goals of the manual, describes the scientific method, contrasts experimental and observational methods, and explains what constitutes linguistic data. Because this manual is aimed at the new researcher, Chapter 2 provides a brief introduction to research with human subjects, reviewing the specific requirements for working with children. In it, we discuss the researcher's preparation for working with participants and schools. In Chapter 3, we introduce the concept of *metadata*—mainly with respect to subject and session metadata—and explain its importance in research. We also explain general principles for using recording equipment and present common lab practices before and after a recording session to allow for effective collaboration and data management.

Part II builds on the foundations outlined in Part I by describing various observational and experimental methods for gathering speech data. In Chapter 4, we discuss the issue of using speech to provide primary data in the study of language acquisition. This chapter addresses the “natural speech” data collection method as one approach to gathering such primary data, along with its advantages and disadvantages when contrasted with experimental methods. Chapter 5 covers basic concepts of experimental research design at a beginner level: What is a good research question? What is a hypothesis? What is good experimental design? Basic principles of statistical analysis and recommendations for more advance readings are introduced. In Chapter 6, we discuss tasks used in conjunction with experimental methods for gathering language production data, focusing mainly on elicited imitation and elicited production. In Chapter 7, we discuss experimental tasks for gathering language comprehension data. We focus mainly on the act-out task and the truth-value judgment task. In Chapter 8, we review the relationship of grammaticality judgment data to

competence and performance. Its advantages and disadvantages and the challenges of using it for testing children are discussed as well.

The tasks described in Part II result in a large collection of data, so Part III explores how to work with those data in more detail. In Chapter 9, we discuss how to create scientifically valid data from language, and we detail the complex process of data creation for both natural speech and experimental research. In this chapter, we explain basics for transcription and reliability checking and introduce some of the complexities of cross-linguistic research. In Chapter 10, we explain principles of preparation of data for analysis, whether using natural speech or experimental data. We introduce principles for basic linguistic analysis in a way that can help train new students with little linguistic background, laying the foundation for researchers to develop more precise data analyses directed to their specific research question. In Chapter 11, principles for interpreting results from different types of research tasks and methods are discussed. We help students become aware of complexities of interpretation of results at the end of a data pipeline.

Part IV concludes the book with some considerations about working with different populations, as well as some final thoughts about the volume as a whole. In Chapter 12, María Blume presents special challenges inherent in assessing multilingual acquisition and surveys developments in methods for multilingual research. Chapter 13 provides a brief introductory review of the principles underlying research on infant language. It overviews some of the main methodologies while referring students to more specialized bibliographical sources. Finally, in Chapter 14, we draw conclusions regarding what we hope to have conveyed in this manual, and we sketch directions for the future of the field of language acquisition to strengthen its scientific foundations. We also introduce the vision for a new resource available to students and researchers working in the field. We describe an example of a web-based portal (Virtual Linguistics Lab) of materials for learning and practicing methods introduced in previous chapters. The portal provides an overview of a cybertool: the Data Transcription and Analysis Tool, which allows web-based shared data management, data transcription, and data analyses supporting collaborative research.

We also provide a number of supporting materials in our chapter and end-of-book appendices and online at <http://pubs.apa.org/books/supp/blume/>. These include, for example, sample scoring criteria, scoring sheets for specific experiments using specific tasks, explanations of some transcription symbols (Appendix A), the International Phonetic Alphabet commonly used in speech transcription (Appendix B), and a template for writing a research proposal (Appendix C).

Although the principles and procedures we introduce in this manual have been developed over more than 30 years in the Cornell University Language Acquisition Lab as well as in several related labs, we hope that the principles that underlie them will be helpful to many labs approaching the scientific study of language acquisition. We hope they help to advance the scientific foundations of our field.

## The Virtual Linguistic Laboratory: An Introduction

Before continuing, it is relevant for all educators, students, and researchers to understand the manual's context and the complementary materials that are available to be used in conjunction with it.

The Cornell Language Acquisition Laboratory has created an Internet portal, the Virtual Linguistic Laboratory (VLL), introduced in Chapter 14, where the user of this manual can find a multitude of resources to complement his or her learning experience. These resources include pedagogical materials for setting up and teaching a class on methods for studying language acquisition, learning modules with actual child data in audio and video formats, a discussion board for collaborating with other students and institutions, and a Data Transcription and Analysis Tool for storing and processing acquisition data.

The VLL portal can be found at <http://www.clal.cornell.edu/vll>. Full access to all its materials is made available to participating members.<sup>1</sup> Membership is offered to educational institutions seeking to offer a class on language acquisition research methodology using the VLL materials or to researchers interested in using the tools developed at the Cornell Language Acquisition Laboratory and the Virtual Center for Language Acquisition. The VLL is offered as an example of the type of Internet-based portal that could be designed and established at any institution with the resources to support server access and dissemination.

---

<sup>1</sup> For membership details contact the authors: Barbara C. Lust ([bcl4@cornell.edu](mailto:bcl4@cornell.edu)) and María Blume ([mblume@pucp.pe](mailto:mblume@pucp.pe)).