Research Methods in Language Variation and Change

Methodological know-how has become one of the key qualifications in contemporary linguistics, which has a strong empirical focus. Containing twenty-three chapters, each devoted to a different research method, this volume brings together the expertise and insight of a range of established practitioners. The chapters are arranged in three parts, devoted to three different stages of empirical research: data collection, analysis, and evaluation. In addition to detailed step-by-step introductions and illustrative case studies focusing on variation and change in English, each chapter addresses the strengths and weaknesses of the methodology and concludes with suggestions for further reading. This systematic, state-of-the-art survey is ideal for both novice researchers and professionals interested in extending their methodological repertoires. The book also has a companion website which provides readers with further information, links, resources, demonstrations, exercises and case studies related to each chapter.

MANFRED KRUG is Chair of English and Historical Linguistics in the Department of English and American Studies at the University of Bamberg.

JULIA SCHLÜTER is Associate Professor of English and Historical Linguistics in the Department of English and American Studies at the University of Bamberg.

Research Methods in Language Variation and Change

Edited by
MANFRED KRUG
and
JULIA SCHLÜTER

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Published in the United States of America by Cambridge University Press, New York

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9780521181860

First published 2013

Printed in the United Kingdom by TJ International Ltd. Padstow Cornwall

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Research methods in language variation and change / Edited by Manfred Krug, Julia Schlüter.
pages cm
Includes bibliographical references and index.
ISBN 978-1-107-00490-0 (Hardback) – ISBN 978-0-521-18186-0 (Paperback)
1. Language and languages–Variation. 2. Language and languages–Research. 3. Linguistic change. 4. Linguistics–Methodology. I. Krug, Manfred G., 1966– editor of compilation.
II. Julia Schlüter, 1973– editor of compilation.
P120.V37R47 2013
417′.0721–dc23 2013010575

ISBN 978-1-107-00490-0 Hardback
ISBN 978-0-521-18186-0 Paperback

Contents

Figures

Tables

Contributors

Lieselotte Anderwald,
English Department, University of Kiel, Germany

Douglas Biber,
Department of English, Northern Arizona University, USA

Joan Bresnan,
Department of Linguistics, Stanford University, USA

Lynn Clark,
Linguistics Department, University of Canterbury, Christchurch, New Zealand

Marilyn Ford,
School of Information and Communication Technology, Griffith University, Nathan, Queensland, Australia

Bethany Gray,
Department of English, Iowa State University, Ames, USA

Stefan Th. Gries,
Department of Linguistics, University of California, Santa Barbara, USA

Ulrike Gut,
Department of English, University of Münster, Germany

Sebastian Hoffmann,
Department of English Studies, University of Trier, Germany

Thomas Hoffmann,
Department of English and American Studies, University of Osnabrück, Germany

Marianne Hundt,
English Department, University of Zurich, Switzerland

Bernd Kortmann,
Department of English, University of Freiburg, Germany

William A. Kretzschmar, Jr.,
Department of English, University of Georgia, Athens, USA

Manfred Krug,
Department of English and American Studies, University of Bamberg, Germany

Warren Maguire,
Linguistics and English Language, University of Edinburgh, UK

Christian Mair,
English Department, University of Freiburg, Germany

Heikki Mannila,
Department of Information and Computer Science, Aalto University, Finland

April McMahon,
Vice-Chancellor, Aberystwyth University, UK

Donka Minkova,
Department of English, University of California, Los Angeles, USA

Terttu Nevalainen,
Department of Modern Languages, University of Helsinki, Finland

Helena Raumolin-Brunberg,
Department of Modern Languages, University of Helsinki, Finland

Günter Rohdenburg,
Department of English and American Studies, University of Paderborn, Germany

Anette Rosenbach,
Department of English and American Studies, University of Paderborn, Germany

Julia Schlüter,
Department of English and American Studies, University of Bamberg, Germany

Daniel Schreier,
English Department, University of Zurich, Switzerland

Katrin Sell,
Department of English and American Studies, University of Bamberg, Germany

Elena Seoane,
Department of English, French and German, University of Vigo, Spain

Nicholas Smith,
School of Education, University of Leicester, UK

Benedikt Szmrecsanyi,
Department of Linguistics, University of Leuven, Belgium

Sali A. Tagliamonte,
Department of Linguistics, University of Toronto, Canada

Graeme Trousdale,
Linguistics and English Language, University of Edinburgh, UK

Preface

MANFRED KRUG AND JULIA SCHLÜTER

Synopsis

For several decades, linguistic research has seen an increasing trend towards empirical methodologies. On the one hand, this has led to a shift in linguistic interest away from the study of single example sentences as manifestations of a monolithic grammar and towards an investigation of (synchronic) variation and (diachronic) change on all levels of linguistic organization. On the other hand, this evolution has transformed many strands of linguistics into branches of an objective science and increased the need for falsifiable and, in many cases, quantifiable data. Consequently, the spectrum of methodologies used in contemporary linguistics has considerably broadened and diversified; different strands of variationist linguistics have developed a wide range of useful techniques for data collection, analysis and evaluation. As a result, methodological know-how has become one of the key qualifications for linguists, both newcomers to the discipline and professional practitioners.

However, it is increasingly difficult even for professionals to keep track of the methodological advances in neighbouring fields of linguistic study: most of the discussion in publications and conference meetings revolves around the findings that have resulted from the successful application of research methods. Too little space and time, at least in our view, is devoted to making the methods explicit and to communicating them in a way that would allow others to replicate them. This lack of methodological transparency results in a situation in which empirical studies run the risk of failing to meet two fundamental principles of objective science: reproducibility and falsifiability. Advanced undergraduate and graduate students are faced with a similar problem: BA, MA and Ph.D. theses are expected to involve original research projects demonstrating their authors' ability to do empirical research, but many students receive little explicit guidance on questions of methodology – at least beyond the immediate field of their supervisor(s).

The present book aims to fill this niche by providing an overview of empirical research methods used in the field of language variation and change. It brings together chapters written by leading scholars and aims at a balanced and representative survey of many of the established and some more innovative methodologies in the field of empirical linguistics. The focus of the chapters is on the methodological issues involved, which are illustrated with exemplary

case studies of specific phenomena in the domains of language variation and change. Though the case studies are drawn from the English language, the methodologies discussed are not restricted to English linguistics, but are similarly used in the investigation of variation and change in different natural languages. Further resources, exercises, sample material for case studies, web links and downloads corresponding to each of the chapters can be found on the companion website at www.cambridge.org/krug_schluter.

Readership

The book is intended for readers from a wide range of levels and backgrounds. It is, we believe, of fundamental interest to advanced students of language and linguistics who are engaging in empirical work for the first time, as is generally the case when it comes to preparing a BA or MA thesis. It is highly relevant for Ph.D. students, who in our view should represent the primary readership of the book, because for doctoral theses, an informed choice of approach is as essential as an in-depth methodological awareness. Doctoral students may in addition wish to draw on more than one of the approaches outlined in the following chapters, for instance corpus analysis + experiment; standard reference corpus + worldwide web; OED + historical text databases; phonological analysis + multifactorial statistical testing.

The methodological steps involved in each of the analyses in the book chapters are made explicit and are thus reproducible for readers who have some basic knowledge of linguistics, but no prior experience with the methodologies outlined. Each chapter concludes with a juxtaposition of pros/potentials vs. cons/caveats characterizing each methodology, and is followed by suggestions for further reading for those seeking more detailed information on a certain approach.

This volume seeks to encourage a methodological discussion among experienced practitioners of linguistic research, since the field of linguistics has recently seen an unprecedented increase in the diversity and complexity of the methods employed. In this respect, the focus on methodological issues pursues a twofold aim: On the one hand, there is a real need for insights into the approaches used by other linguists in the field, to which other forms of publication (conference papers and journal articles) devote little space or time. Such insights are indispensable for a critical assessment of the findings obtained by colleagues and will eventually enhance the transparency within the field. On the other hand, we strongly believe that linguists can profit from a look beyond their horizons and can enlarge their own methodological repertoires by adopting or adapting the approaches chosen by others.

The chapters of this handbook have been specially commissioned from leading experts and practitioners in their fields, who share their experience with beginning researchers as well as colleagues. This multi-authored design has the

advantage of offering a more balanced, objective survey of methodologies, written by people with extensive first-hand experience with the approaches they describe. The design of this handbook, finally, reflects our firm conviction that the methodological pluralism prevailing in modern linguistics cannot be adequately represented by a single author or a small number of collaborators.

Structure

The book is introduced by a stage-setting chapter on linguistics as an empirical discipline and the importance of studying variation and change in present-day linguistic research. The main body of the book is subdivided into three parts, mirroring the major stages of an empirical research project: collecting, analysing and evaluating data. Each part comprises two or three sections, containing between two and four chapters each. Each of the sections is devoted to a fundamental approach to variation and change:

Part 1: Collecting empirical data

Unless a researcher decides to use a ready-made database, the first stage of a project involves the compilation of an appropriate dataset. Three major approaches are highlighted in this respect.

Part 1.1 Fieldwork and linguistic mapping: This part deals with ways of investigating unexplored terrain. Unknown language communities can be entered for purposes of linguistic research if the researcher takes certain conditions into consideration; he/she can observe linguistic features in a language community he/she is not part of; or he/she can become a member of such a community and monitor language usage for certain features he/she is interested in. Dialectal differences, isolated by questioning informants, can be represented in linguistic atlases.

Part 1.2 Eliciting linguistic data: Some very efficient ways of collecting linguistic data involve elicitation, i.e. getting informants to produce relevant utterances or judgements on given utterances. Linguistic questionnaires and interviews as well as highly controlled experimental settings are appropriate instruments for eliciting the forms and structures of interest or for investigating issues such as grammaticality/acceptability judgements. This method is especially useful for rarer phenomena. Rather than relying on subjective intuition (which may be biased towards the expected effect), the authors in this section introduce elicitation techniques from unstructured sociolinguistic interviews via purpose-built questionnaires to elaborate experimental settings that are designed to minimize the distorting effects of the participants' awareness of the research situation.

Part 1.3 Alternatives to standard reference corpora: In modern linguistics, corpus analysis is the most widely used methodology for studying

synchronic variation and practically the only one for quantitative analyses of diachronic change. If a researcher is interested in language use beyond what is documented in widely available standard corpora, he/she may consider resorting to other large-scale sources of computer-readable texts. In this way, he/she can access historical dialect data, employ the quotations included in the electronic version of the *Oxford English Dictionary* as a database, or even tap the internet as a source for lesser-known national varieties. This section discusses the methodological pitfalls inherent in these approaches and ways to sidestep them.

Part 2: Analysing empirical data

The second stage of a research project, subsequent to data collection, involves the analysis of the accumulated data. For semantic, syntactic and morphological purposes, (semi-)automated searches (usually in connection with manual post-editing) in standard or purpose-built databases are the most common choice, while phonetic and phonological projects require more specialized methods. In addition, it is possible, though still far from common, to fruitfully combine two or more approaches.

Part 2.1 Corpus analysis: Obvious sources of data on variation and change are the ever more numerous and increasingly comprehensive reference corpora. Although these are tailored to variationists' needs, their use poses a number of methodological problems, the most important of which are discussed in this section. Corpus size is a critical issue when exploring infrequent phenomena in a language. How far can one get using relatively small (one-million-word) corpora? Which corpus analysis tools are there and what are their assets and weaknesses? How does one go about retrieving and annotating concordance entries? Which amenities does a grammatically tagged corpus offer?

Part 2.2 Phonetic and phonological analysis: For the analysis of phonetic distinctions and phonological systems, as well as phonological features above the segmental level, analysts have a variety of auditory, acoustic and articulatory methods at their disposal. These are introduced in two chapters of this section, one focusing on segmental and the other on suprasegmental features. In the reconstruction of historical sound systems, more indirect evidence has to be adduced. An example of such an approach is provided in the third chapter.

Part 2.3 Combinations of multiple types of data: The division of the present volume into subsections may appear to suggest that the methodological approaches described should be used in isolation. However, it is possible and often desirable to combine different approaches so as to make up for the weaknesses of each and to enhance their reliability. The chapters in this pivotal section illustrate this with regard to the cross-over between a variety of elicitation techniques (such as are typically used in psycholinguistics and typology) and corpus-based data (the current pet method of variationist linguistics and dialectology).

Part 3: Evaluating empirical data

In many cases, it may not be sufficient in variationist research to simply count occurrences and compare frequencies or percentages. The chapters in this part explain some basic statistical techniques and provide an outlook on more advanced computational procedures for the evaluation of empirical data.

Part 3.1 Basic statistical analysis: Most linguists come from a background in which they have received linguistic training, but are less comfortable with basic statistical techniques for handling data. This section introduces simple procedures that every empirical linguist should master. Rather than pooling and averaging one's data, bootstrapping and Bayesian statistics allow the researcher to make better use of small datasets. In addition, it is nowadays common practice in empirical linguistics to subject one's data to basic statistical tests to ensure that observed effects are not simply due to chance. Readers are provided with both the know-how and (on the companion website) the software for doing this, and the need for further statistical elaboration is discussed.

Part 3.2 Multifactorial analysis: The final section of the volume concerns the possibilities opened up by complex statistical procedures for the evaluation of empirical data. There has recently been a profusion of rather advanced multidimensional approaches to variation and change, which entail certain advantages and disadvantages for the analysis. Some influential examples, their methodological and theoretical backgrounds as well as their applications to linguistic data are made explicit. Though not all readers will find themselves in a position to replicate them, most will be interested in learning about the rationale behind such techniques. The chapters in this section also serve to illustrate the almost unlimited possibilities of data analysis, to provide clues to identifying and combining appropriate methods for a given project, and to encourage researchers to develop their own approaches.

Acknowledgements

Our sincere thanks go to our contributors, not only for their chapters, but also for their patience and (in most cases) for their participation and lively discussion at a Research Methods Symposium and other invited talks at the University of Bamberg; the University of Bamberg and the Bavarian State Ministry of Sciences, Research and the Arts for financial support for said symposium; two anonymous readers for Cambridge University Press; our proofreader Shane Walshe; our student assistants David Stewart and Fabian Vetter; and the participants in several advanced linguistics courses at the universities of Bamberg and Regensburg for their contributions to the companion website, which can be accessed at www.cambridge.org/krug_schluter.