

Cambridge University Press
978-1-107-00490-0 - Research Methods in Language Variation and Change
Edited by Manfred Krug and Julia Schlüter
Frontmatter
[More information](#)

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.

Cambridge University Press

978-1-107-00490-0 - Research Methods in Language Variation and Change

Edited by Manfred Krug and Julia Schlüter

Frontmatter

[More information](#)

Cambridge University Press

978-1-107-00490-0 - Research Methods in Language Variation and Change

Edited by Manfred Krug and Julia Schlüter

Frontmatter

[More information](#)

Research Methods in Language Variation and Change

Edited by

MANFRED KRUG

and

JULIA SCHLÜTER



Cambridge University Press
 978-1-107-00490-0 - Research Methods in Language Variation and Change
 Edited by Manfred Krug and Julia Schlüter
 Frontmatter
[More information](#)

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

© Cambridge University Press 2013

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

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

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

| | |
|-----------------------------|-------|
| <i>List of figures</i> | viii |
| <i>List of tables</i> | xiv |
| <i>List of contributors</i> | xviii |
| <i>Preface</i> | xxi |
| <i>Acknowledgements</i> | xxvi |

| | |
|--|---|
| Introduction: Investigating language variation and change | 1 |
| MANFRED KRUG, JULIA SCHLÜTER AND ANETTE ROSENBACH | |

| | |
|---|----|
| PART 1 Collecting empirical data | 15 |
|---|----|

| | |
|--|----|
| Part 1.1 Fieldwork and linguistic mapping | 17 |
|--|----|

| | |
|---|----|
| 1 Collecting ethnographic and sociolinguistic data | 17 |
| DANIEL SCHREIER | |

| | |
|--|----|
| 2 Using participant observation and social network analysis | 36 |
| LYNN CLARK AND GRAEME TROUSDALE | |

| | |
|--|----|
| 3 Computer mapping of language data | 53 |
| WILLIAM A. KRETZSCHMAR, JR. | |

| | |
|---|----|
| Part 1.2 Eliciting linguistic data | 69 |
|---|----|

| | |
|---|----|
| 4 Designing and conducting interviews and questionnaires | 69 |
| MANFRED KRUG AND KATRIN SELL | |

| | |
|---|----|
| 5 Obtaining introspective acceptability judgements | 99 |
| THOMAS HOFFMANN | |

| | |
|--|-----|
| Part 1.3 Alternatives to standard reference corpora | 119 |
|--|-----|

| | |
|---|-----|
| 6 Using historical literature databases as corpora | 119 |
| JULIA SCHLÜTER | |

| | | |
|----|---|------------|
| vi | Contents | |
| | 7 Using the OED quotations database as a diachronic corpus | 136 |
| | GÜNTER ROHDENBURG | |
| | 8 Using web-based data for the study of global English | 158 |
| | MARIANNE HUNDT | |
| | PART 2 Analysing empirical data | 179 |
| | Part 2.1 Corpus analysis | 181 |
| | 9 Using ‘small’ corpora to document ongoing grammatical change | 181 |
| | CHRISTIAN MAIR | |
| | 10 Using tag sequences to retrieve grammatical structures | 195 |
| | SEBASTIAN HOFFMANN | |
| | 11 Categorizing syntactic constructions in a corpus | 212 |
| | NICHOLAS SMITH AND ELENA SEOANE | |
| | Part 2.2 Phonetic and phonological analysis | 228 |
| | 12 Analysing phonetic and phonological variation on the segmental level | 228 |
| | ULRIKE GUT | |
| | 13 Analysing phonetic and phonological variation on the suprasegmental level | 244 |
| | ULRIKE GUT | |
| | 14 Reconstructing stress in Old and Middle English | 260 |
| | DONKA MINKOVA | |
| | Part 2.3 Combinations of multiple types of data | 278 |
| | 15 Combining elicitation data with corpus data | 278 |
| | ANETTE ROSENBACH | |
| | 16 Using convergent evidence from psycholinguistics and usage | 295 |
| | MARILYN FORD AND JOAN BRESNAN | |
| | 17 Applying typological methods in dialectology | 313 |
| | LIESELOTTE ANDERWALD AND BERND KORTMANN | |

| | Contents | vii |
|--|----------|-----|
| PART 3 Evaluating empirical data | 335 | |
| Part 3.1 Basic statistical analysis | 337 | |
| 18 Quantifying variation and estimating the effects of sample size on the frequencies of linguistic variables | 337 | |
| HEIKKI MANNILA, TERTTU NEVALAINEN AND HELENA RAUMOLIN-BRUNBERG | | |
| 19 Elementary statistical testing with R | 361 | |
| STEFAN TH. GRIES | | |
| Part 3.2 Multifactorial analysis | 382 | |
| 20 Analysing and interpreting variation in the sociolinguistic tradition | 382 | |
| SALI A. TAGLIAMONTE | | |
| 21 Identifying multi-dimensional patterns of variation across registers | 402 | |
| DOUGLAS BIBER AND BETHANY GRAY | | |
| 22 Computing linguistic distances between varieties | 421 | |
| APRIL MCMAHON AND WARREN MAGUIRE | | |
| 23 Analysing aggregated linguistic data | 433 | |
| BENEDIKT SZMRECSANYI | | |
| <i>Bibliography</i> | 456 | |
| <i>Index</i> | 501 | |

Figures

| | | |
|-----|---|----|
| 1 | The deductive approach typically used in quantitative research and the inductive logic of research in a qualitative study | 4 |
| 2 | The empirical cycle | 5 |
| 3 | Types of linguistic data, arranged according to their degree of naturalness/monitoring | 6 |
| 2.1 | Aggregate network sociogram for WFHPB | 42 |
| 3.1 | Map of American dialects (Kurath 1949) | 54 |
| 3.2 | <i>Dragonfly</i> (Kurath 1949) | 55 |
| 3.3 | Graphic plotter grid for 'kindling' (1986) | 60 |
| 3.4 | Layers in a GIS display | 62 |
| 3.5 | GIS display with highlighted respondents | 63 |
| 3.6 | Density estimation for the response <i>pail</i> | 64 |
| 3.7 | Line segments indicating Delaunay Triangulation for LAMSAS | 66 |
| 3.8 | Levenshtein Distance displayed with MDS from www.let.rug.nl/~kleiweg/lamsas-old/results | 67 |
| 4.1 | Types of interviews and degrees of structuredness | 72 |
| 4.2 | The different levels of measurement and typical categories of answers | 77 |
| 4.3 | Likert scale | 77 |
| 4.4 | Histograms (spoken data) for <i>Do you want I get us some ice cream?</i> and <i>You want I buy a drink for you?</i> | 87 |
| 4.5 | Histograms (spoken vs. written data) for <i>Do you want I get us some ice cream?</i> | 88 |
| 4.6 | Histograms (spoken vs. written data) for <i>I like this painting, I prefer the other one, but.</i> | 89 |
| 4.7 | Bamberg questionnaire for lexical and morphosyntactic variation in English (Maltese English version, informant information sheet) | 94 |
| 4.8 | Bamberg questionnaire for lexical and morphosyntactic variation in English (Maltese English version, excerpt from auditory-based grammar part) | 95 |
| 4.9 | Bamberg questionnaire for lexical and morphosyntactic variation in English (Maltese English version, excerpt from lexical part) | 96 |

| | | |
|------|--|-----|
| 4.10 | Bamberg questionnaire for lexical and morphosyntactic variation in English (Maltese English version, excerpt from written grammar part) | 97 |
| 5.1 | Pied Piping across relativizers and PP types compared with fillers (British English speakers) | 113 |
| 5.2 | Pied Piping across relativizers and PP types compared with fillers (German L2 English learners) | 113 |
| 5.3 | Stranding means for all relativizers across PP types compared with fillers (British English speakers) | 115 |
| 5.4 | Stranding means for all relativizers across PP types compared with fillers (German L2 learners) | 115 |
| 6.1 | Standard Search window and Keyword Browse window in the EEPF database | 124 |
| 6.2 | Brief Summary of Matches window and Context of Matches window in the EEPF database | 125 |
| 6.3 | The distribution of <i>min(e)</i> and <i>my</i> in prenominal position as a function of the initial sound of the following word in PPCME2 and PPCEME | 127 |
| 6.4 | The distribution of <i>min(e)</i> and <i>my</i> in prenominal position as a function of the initial sound of the following word in (PPCME2 and) EEPF | 128 |
| 6.5 | The distribution of regular and irregular past tense and past participle forms of the verbs <i>burn</i> , <i>dream</i> , <i> dwell</i> , <i> kneel</i> , <i> leap</i> and <i> learn</i> in ARCHER (BrE only) | 129 |
| 6.6 | The distribution of regular and irregular past tense and past participle forms of the verbs <i>burn</i> , <i>dream</i> , <i> dwell</i> , <i> kneel</i> , <i> leap</i> and <i> learn</i> in EEPF, ECF and NCF | 130 |
| 6.7 | The occurrence of adverbs intervening between the negators <i>not</i> and <i>never</i> and attributive adjectives in ECF, NCF and EPD (1701–1903) | 132 |
| 8.1 | Search for potential combinations of the progressive passive with the present or past perfect in the Advanced Search mode of <i>Google</i> | 164 |
| 8.2 | Present progressive passives in stratified reference corpora and web-derived monitor corpora | 176 |
| 9.1 | Five matching corpora of twentieth-century written standard English | 183 |
| 9.2 | Specificational clefts (all types) in five corpora of twentieth-century written English | 189 |
| 9.3 | Specificational clefts – towards a constructional split | 189 |
| 9.4 | Four types of specificational clefts in two corpora of twentieth-century spoken English | 190 |
| 10.1 | Extract from a BNCweb query result for the query <code>_AT0 _{N}</code> | 200 |

| | | |
|------|---|-----|
| x | List of figures | |
| 10.2 | Extract from a BNCweb query result for the query <code>_DPS _{N}</code> | 201 |
| 10.3 | Extract from a BNCweb query result for the query <code>(_AT0 _DT0 _DPS)* _{A} _{N}</code> | 201 |
| 10.4 | Extract from a BNCweb query result for the query <code>(_AT0 _DT0 _DPS)* ((_AV0)* _{A} (_CJCN,_PUN))+ _{N}</code> | 204 |
| 10.5 | Extract from a BNCweb query result for the query <code>(_AT0 _DT0 _DPS)* ((_AV0)* _{A} (_CJCN,_PUN))* (_{N})+</code> | 205 |
| 10.6 | Extract from a BNCweb query result for the query <code>((longest)) (_AT0 _DT0 _DPS)* ((_AV0)* _{A} (_CJCN,_PUN))* (_{N})+</code> | 206 |
| 10.7 | Extract from a BNCweb query result for the query <code>((longest)) (_AT0 _DT0 _DPS)* ((_AV0)* _{A} (_CJCN,_PUN))* (_{N})+ (of (_AT0 _DT0 _DPS)* ((_AV0)* _{A} (_CJCN,_PUN))* (_{N})+)+</code> | 207 |
| 11.1 | Prominence hierarchies | 214 |
| 11.2 | Categorizing information status of long passives within corpus software (WordSmith) | 218 |
| 11.3 | An annotated database of the long passive, using Excel | 220 |
| 12.1 | The IPA transcription symbols for the cardinal vowels Reprinted with permission from the International Phonetic Association. Copyright 2005 by International Phonetic Association | 231 |
| 12.2 | Spectrogram of the vowels [i], [ɪ], [e], [æ], [ɐ], [ɔ], [o] and [u] with indication of the formants | 233 |
| 12.3 | Acoustic map for some American English vowels. Reprinted with permission from Gut (2009: 154) | 235 |
| 12.4 | A participant in an EMA experiment. Reprinted with permission from Draxler (2008: 73) | 237 |
| 12.5 | A speaker on a nasometer | 237 |
| 12.6 | The IPA transcription symbols for the pulmonic consonants Reprinted with permission from the International Phonetic Association. Copyright 2005 by International Phonetic Association | 238 |
| 12.7 | The IPA transcription symbols for phonetic details: the diacritics. Reprinted with permission from the International Phonetic Association. Copyright 2005 by International Phonetic Association | 240 |
| 12.8 | Measurement of the VOT of the /k/ in <i>came</i> | 240 |
| 12.9 | Palatograms of [t] and [d] for two speakers. Reprinted with permission from Gut (2009: 47) | 242 |
| 13.1 | Measurement of the pitch height and intensity at mid-point of the first vowel in <i>cassette</i> | 246 |

| | List of figures | xi |
|-------|---|-----|
| 13.2 | Power spectrum of the vowel /ɒ/ in <i>object</i> | 247 |
| 13.3 | Division of the utterance <i>Give her the post</i> into vocalic and consonantal intervals in Praat | 250 |
| 13.4 | Interlinear transcription of the intonation of the utterance <i>She didn't want to leave</i> | 254 |
| 13.5 | Example of a ToBI transcription of the utterance <i>a tiger and a mouse were walking in a field</i> with the waveform (top), spectrogram, in which the pitch movement (line) is plotted (middle) and three transcription tiers (bottom) | 256 |
| 13.6 | Pitch movement and points of measurement for wide and small pitch range according to Patterson (2000) | 258 |
| 14.1 | Pre-tonic syllable loss in English | 264 |
| 15.1 | Animacy versus weight in the experimental study: relative frequency of the <i>s</i> -genitive and the <i>of</i> -genitive | 287 |
| 15.2a | Weight of the possessor with human possessors in the ICE-GB: relative frequency of the <i>s</i> -genitive and the <i>of</i> -genitive | 290 |
| 15.2b | Weight of the possessor with inanimate possessors in the ICE-GB: relative frequency of the <i>s</i> -genitive and the <i>of</i> -genitive | 290 |
| 16.1 | The model formula for datives | 298 |
| 17.1 | eWAVE map for feature 155 (<i>ain't</i> as the negated form of <i>be</i>) | 320 |
| 17.2 | Accounting for gender diffusion in terms of Sasse's Individuation Hierarchy | 321 |
| 17.3 | Analyticity by syntheticity | 325 |
| 17.4 | Decrease in analyticity and increase in syntheticity in written vs. spoken varieties of English | 327 |
| 18.1 | Results given by the four different estimation methods for the frequency of (ing) | 345 |
| 18.2 | Results given by the four different estimation methods for the frequency of (you) | 346 |
| 18.3 | Scatterplots of the frequency of the (ing) variable at 40-year intervals, comparing the four estimation methods pooling, average of averages, Bayesian, and bootstrap | 348 |
| 18.4 | The Bayesian estimate and posterior interval as a function of the bootstrap estimate (a) (ing); (b) (you) | 349 |
| 18.5 | The Bayesian and bootstrap estimates and the width of the confidence interval for (ing) for samples ranging in size from 15 to 130 persons | 351 |
| 18.6 | (a) The standard deviation in the Bayesian estimate for samples of size 15, 30, and 50, as a function of the Bayesian estimate. (b) The standard deviation in the bootstrap estimate | |

| | | |
|------|--|-----|
| xii | List of figures | |
| | for samples of size 15, 30, and 50, as a function of the bootstrap estimate | 352 |
| 18.7 | (a) The standard deviation in the Bayesian estimate for samples of size 15, 30, and 50, as a function of the Bayesian estimate. (b) The standard deviation in the bootstrap estimate for samples of size 15, 30, and 50, as a function of the bootstrap estimate | 353 |
| 18.8 | (a, c) Bayesian estimates for the full data set vs. Bayesian estimates based on persons with at least 10 (a) or 5 (c) attestations. (b, d) Bootstrap estimates for the full data set vs. bootstrap estimates based on persons with at least 10 (b) or 5 (d) attestations | 355 |
| 18.9 | (a, c) Bayesian estimates for the full data set vs. Bayesian estimates based on persons with at least 10 (a) or 5 (c) attestations. (b, d) Bootstrap estimates for the full data set vs. bootstrap estimates based on persons with at least 10 (b) or 5 (d) attestations | 356 |
| 19.1 | Association plot for the relation between question and answer syntax | 370 |
| 19.2 | Boxplot for the relation between subordinate clause length and mode | 373 |
| 19.3 | Line plots and smoothers for the normalized frequencies of <i>in</i> and <i>just because</i> over time | 377 |
| 19.4 | Choosing a <i>U</i> test / two sample Wilcoxon test in the R commander | 380 |
| 19.5 | Performing a <i>U</i> test / two sample Wilcoxon test in the R commander | 381 |
| 20.1 | Overall frequency of consonant cluster simplification in York and Toronto | 390 |
| 20.2 | Frequency of consonant cluster simplification by individual in York | 391 |
| 20.3 | Frequency of consonant cluster simplification by individual in Toronto | 391 |
| 20.4 | Distribution of consonant cluster simplification by following phonological context by individual in York | 392 |
| 20.5 | Distribution of consonant cluster simplification by following phonological context by individual in Toronto | 392 |
| 20.6 | Distribution of consonant cluster simplification by following phonological context in York and Toronto | 393 |
| 21.1 | Mean scores of university registers along Dimension 1 – Oral vs. literate discourse | 412 |
| 21.2 | Mean scores of registers along Dimension 2 – Procedural vs. content-focused discourse | 416 |

| | | |
|------|---|-----|
| 21.3 | Mean scores of registers along Dimension 4 – Teacher-centered stance | 418 |
| 22.1 | Length comparisons for English vowels | 424 |
| 22.2 | NeighborNet of all modern English Typical varieties | 426 |
| 22.3 | Words contributing to the difference between Glasgow and Sheffield | 427 |
| 22.4 | All modern English Typical varieties; items with postvocalic /r/ excluded | 428 |
| 22.5 | NeighborNet for Newcastle and Edinburgh – Typical and Traditional varieties plus individual speakers | 430 |
| 23.1 | Visualization of the 34 x 2 index matrix: BNC macro registers – analyticity by syntheticity (in index points, <i>ptw</i>) | 438 |
| 23.2 | Link map – traditional British English dialects | 443 |
| 23.3 | Scatterplots – morphosyntactic distance versus as-the-crow-flies distance, least-cost travel time and Trudgill’s linguistic gravity index (<i>log</i> scale) | 444 |
| 23.4 | Two-dimensional multidimensional scaling plot – varieties of English worldwide | 448 |
| 23.5 | Dendrogram deriving from hierarchical agglomerative cluster analysis – varieties of English worldwide (clustering algorithm: Ward) | 449 |

Tables

| | | |
|-----|---|-----|
| 1 | Survey of methods used in different linguistic subdisciplines, according to the information given in article abstracts in recent issues of relevant journals | 8 |
| 2.1 | Example of part of a binary matrix | 41 |
| 2.2 | Extract from the aggregate network | 41 |
| 2.3 | Social groups in WFHPB (summary of multiple clique analyses) | 43 |
| 2.4 | Linguistic factor groups for Varbrul analysis of (th) | 46 |
| 2.5 | Social factor groups for Varbrul analysis of (th) | 46 |
| 2.6 | Cognitive factor groups for Varbrul analysis of (th) | 47 |
| 2.7 | Multivariate analysis of the contribution of factors selected as significant to the probability of (th): [f] | 48 |
| 2.8 | Social groups which favour th-fronting in WFHPB | 50 |
| 2.9 | Social groups which disfavour th-fronting in WFHPB | 50 |
| 4.1 | Median, mean and standard deviation for <i>Do you want I get us some ice-cream?</i> | 85 |
| 4.2 | Median, mean and standard deviation for <i>You want I buy a drink for you?</i> | 85 |
| 4.3 | Median, mean and standard deviation for <i>Do you want me to get us some ice-cream?</i> | 86 |
| 4.4 | Median, mean and standard deviation for <i>You want me to buy a drink for you?</i> | 86 |
| 4.5 | Median, mean and standard deviation for <i>I like this painting, I prefer the other one, but.</i> | 89 |
| 5.1 | Token set-example <i>laugh at: You wouldn't believe the things . . .</i> | 107 |
| 5.2 | Counterbalanced material set sentence list | 108 |
| 5.3 | Repeated measures ANOVA for English L1 and German L2 informants | 112 |
| 6.1 | Details of three exemplary historical reference corpora | 120 |
| 6.2 | Details of six historical literature databases | 122 |
| 6.3 | Survey of the three case studies sketched in Sections 4.1–4.3 | 127 |
| 7.1 | Intransitive and transitive particle verbs like <i>come in/let in</i> associated with and immediately followed by prepositional phrases containing the noun <i>door</i> and its hyponymic | |

| | |
|---|-----|
| compounds like <i>back door</i> in the OED quotations between 1490 and 1988 | 139 |
| 7.2 The evolution of nominal and sentential complements governed by the noun <i>risk</i> in the OED quotations | 140 |
| 7.3 Number of quotations and estimated words for the five centuries between 1500 and 1988 | 145 |
| 7.4 <i>in (the) event</i> governing NPs, gerunds and finite clauses in the quotations drawn from the OED | 146 |
| 7.5 The rivals <i>in (Adj) hopes</i> and <i>in the (Adj) hope</i> governing NPs and sentential complements in the OED quotations (pmw = per million words) | 147 |
| 7.6 The relativization of the standard of comparison by means of <i>than which/whom</i> in the OED quotations and a series of fiction databases | 148 |
| 7.7 Fictional text collections to be compared with the OED quotations database | 148 |
| 7.8 Selected constructions and expressions – regarded as relatively formal – in the OED quotations database and the narrative component of the BNC (=wridom1) | 149 |
| 7.9 Selected constructions and expressions – regarded as relatively informal – in the OED quotations database and the narrative component of the BNC (=wridom1) | 150 |
| 7.10 The rivalry between more or less explicit variants with selected constructions in the OED and the narrative component of the BNC (=wridom1) | 150 |
| 7.11 Selected constructions assumed to be associated with communicative needs specific to fictional or non-fictional text types | 151 |
| 7.12 The use of <i>thereby</i> (including <i>there by</i>) between 1600 and 1699 in the OED quotations and the EEPF | 152 |
| 7.13 The rivalry between infinitival clauses and prepositional gerunds governed by <i>with a view</i> in the OED quotations and fiction databases | 152 |
| 7.14 The type <i>that of NP/-ing</i> in the OED quotations drawn from the works of Defoe, Swift, Steele and Addison | 153 |
| 7.15 The rivalry between the past participial forms <i>gotten</i> and <i>got</i> in the OED quotations drawn from the works of Defoe, Swift, Steele and Addison | 154 |
| 7.16 Structures dependent on <i>in (the) event</i> in the OED quotations database | 155 |
| 7.17 Objects associated with the (active) verb <i>attest</i> in the OED quotations database between 1800 and 1988 | 155 |
| 8.1 Potential progressive passives co-occurring with present/past perfect and modal auxiliaries (27 January 2009) | 164 |

| xvi | List of tables | |
|-------|---|-----|
| 8.2a | Proportion of progressive passives combining with the perfect in sub-samples of 100 hits for ‘been being’ | 167 |
| 8.2b | Comparison of original results with re-run | 167 |
| 8.2c | Proportion of progressive passives combining with the perfect in sub-samples of hits for ‘been being’ | 168 |
| 8.3 | Number of hits for two exact phrase searches from different locations | 172 |
| 8.4 | Present and past tense progressive passives in Pakistani, Bangladeshi and Indian English – relative frequencies per million words | 175 |
| 9.1 | Specificational clefts (all types) in five corpora of twentieth-century written English | 189 |
| 9.2 | Four types of specificational clefts in two corpora of twentieth-century spoken English | 190 |
| 10.1 | Query strings matching various types of premodification | 202 |
| 12.1 | The standard lexical sets for the analysis of vowel inventories with keyword and example words | 230 |
| 14.1 | Stress doublets in <i>Troilus and Criseyde</i> | 273 |
| 15.1 | Factor correlation in English genitive variation | 279 |
| 15.2a | Weight as a relative notion: example with 1-word possessor | 284 |
| 15.2b | Weight as a relative notion: example with 3-word possessor | 284 |
| 15.3 | Genitive constructions: pre- and postmodification patterns | 285 |
| 15.4a | Testing for the status of animacy: critical conditions | 286 |
| 15.4b | Testing for the status of weight: critical conditions | 286 |
| 16.1 | Comparison of items 2 and 3 | 301 |
| 16.2 | Random effect adjustments of US participants, using (4) | 304 |
| 16.3 | Model parameters for the ratings experiment | 306 |
| 16.4 | Model parameters for the reaction time experiment | 308 |
| 16.5 | Comparison of Australian and US participants in tolerance to dative structures with long first arguments | 309 |
| 18.1 | Number of informants in the CEEC with at least 1 attestation of the linguistic variables, both for the whole data and for London | 343 |
| 18.2 | Values of (ing) for the 20-year period 1460–1479 for those individuals for whom there is at least one occurrence of the variable | 344 |
| 19.1 | Table 2 (Appendix 2) from Hundt and Smith (2009), cross-tabulating tenses and corpora | 363 |
| 19.2 | Reorganization of Table 2 (Appendix 2) from Hundt and Smith (2009) | 364 |
| 19.3 | Fictitious frequencies obtained in a question-answer experiment | 367 |
| 19.4 | Lengths of subordinate clauses in two samples (of spoken and written data) | 372 |

| | | |
|------|--|-----|
| 20.1 | Three logistic regression analyses of the occurrence of zero in subject relative clauses in three communities | 395 |
| 21.1 | Composition of the T2K-SWAL Corpus | 405 |
| 23.1 | Correlation coefficients and R^2 values – morphosyntactic distances versus as-the-crow-flies distance, least-cost travel time, and Trudgill’s linguistic gravity index | 444 |

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.