

The Written Questionnaire in Social Dialectology

IMPACT: Studies in Language and Society

ISSN 1385-7908

IMPACT publishes monographs, collective volumes, and text books on topics in sociolinguistics. The scope of the series is broad, with special emphasis on areas such as language planning and language policies; language conflict and language death; language standards and language change; dialectology; diglossia; discourse studies; language and social identity (gender, ethnicity, class, ideology); and history and methods of sociolinguistics.

For an overview of all books published in this series, please see <http://benjamins.com/catalog/impact>

General Editor

Ana Deumert
University of Cape Town

Advisory Board

Peter Auer
University of Freiburg

Jan Blommaert
Ghent University

Annick De Houwer
University of Erfurt

J. Joseph Errington
Yale University

Anna Maria Escobar
University of Illinois at Urbana

Guus Extra
Tilburg University

Marlis Hellinger
University of Frankfurt am Main

Elizabeth Lanza
University of Oslo

William Labov
University of Pennsylvania

Peter L. Patrick
University of Essex

Jeanine Treffers-Daller
University of the West of England

Victor Webb
University of Pretoria

Volume 40

The Written Questionnaire in Social Dialectology. History, theory, practice
by Stefan Dollinger

The Written Questionnaire in Social Dialectology

History, theory, practice

Stefan Dollinger

University of Gothenburg & University of British Columbia

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/impact.40

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2015037176 (PRINT) / 2015039102 (E-BOOK)

ISBN 978 90 272 5831 1 (HB)

ISBN 978 90 272 5832 8 (PB)

ISBN 978 90 272 6777 1 (E-BOOK)

© 2015 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Company · <https://benjamins.com>

*for Nina-Greta
& Benjamin Maximilian*

Table of contents

List of common abbreviations	XV
List of tables	XVII
List of figures	XIX
List of illustrations	XXI
Acknowledgements	XXIII
Abstract	XXV
Companion website	XXVI
Author's preface	XXVII

CHAPTER 1

Written Questionnaires in the wider linguistic context	1
1.1 Three basic types of language data and WQs	4
1.2 Data in traditional dialect geography	6
1.2.1 The Fieldworker Interview (FI) method	7
1.2.2 Wenker's Written Questionnaire (WWQ) method	10
1.3 Today's Written Questionnaire (WQ) methods	11
1.4 The organization of this Book	13

Part I. History & theory

CHAPTER 2

A history of written questionnaires in social dialectology	19
2.1 German-language pioneers	21
2.2 From Wenker's <i>Deutscher Sprachatlas</i> to Mitzka's <i>Wortatlas</i>	23
<i>Advantages</i>	25
2.3 Dutch and Flemish WQs	27
2.4 Early English language WQs in the US	28
2.4.1 A new beginning: Alva L. Davis' (1948) WQ Survey	30
2.4.2 Cassidy's and Allen's WQ studies	34
2.5 Scotland and <i>The Linguistic Atlas of Scotland</i>	36
2.6 WQs in Canada: A special case	38
2.6.1 Canadian beginnings	39
2.6.2 <i>Survey of Canadian English</i> (1972)	42
2.6.3 Other Canadian WQs	44

- 2.6.4 Dialect Topography of Canada (1991–2004) 45
- 2.6.5 *North American Regional Vocabulary Survey* 46
- 2.7 Other, more recent applications 48
- 2.8 Chapter conclusion 51

CHAPTER 3

A comparison of data collection methodologies 53

- 3.1 Corpus linguistics and WQs: A methodological comparison 54
 - 3.1.1 Limited linguistic contexts: Problem #1 55
 - 3.1.2 Low-frequency items: Problem #2 57
 - 3.1.3 (Positive) Evidence and negative evidence: Problem #3 59
 - 3.1.4 Documentation of social backgrounds: Problem #4 60
 - 3.1.5 Corpora and WQs: A comparison 61
 - Linguistic examples: Attested and reported 64
- 3.2 Comparison of elicitation techniques: WQ and FI 65
 - 3.2.1 *The Linguistic Atlas of the Upper Midwest* (1947–1953; 1973–6) 65
 - 3.2.2 Allen's WQ and FI data: Chambers' selection 67
 - 3.2.3 Selecting the best test 68
 - Testing the data for equivalence 70
 - Allen's original assessment 71
- 3.3 Comparison of elicitation techniques: Sociolinguistic interview and WQ 73
 - 3.3.1 The Observer's Paradox and WQs 74
 - 3.3.2 McDavid's test 76
 - 3.3.3 WQs and sociolinguistic interviews in Vancouver, Canada 77
 - Low-back vowels 78
 - Yod-dropping 82
- 3.4 Chapter conclusion 86

CHAPTER 4

Types of traditional WQ variables 87

- 4.1 Lexis (vocabulary) 87
 - 4.1.1 A Canadianism is dying out: *chesterfield* 88
 - The rise and fall of *chesterfield* 92
 - 4.1.2 A Canadianism is staying put: The case of *tap* 95
 - 4.1.3 A Canadianism is entering the scene: TAKE UP #9 99
 - Interpreting the data on TAKE UP #9 102
 - Tracing TAKE UP #9 in the Canadian Oxford 104

4.2	Morphology	105
	<i>Snuck</i> as a global form?	107
4.3	Syntax and usage	109
4.3.1	<i>Different</i> from/than/to?	109
4.3.2	<i>Between</i> you and me or I?	113
	The source of the confusion	114
4.3.3	Telling time: 11:40 or twenty-to-twelve?	117
4.4	Pronunciation: Phonemic variables	121
4.4.1	Yod-dropping	122
	Yod-retention in <i>avenue</i> : An urban vs. rural split?	124
4.4.2	Variation in lexical item <i>vase</i>	125
4.5	Outlook	129

CHAPTER 5

World Englishes, multilingualism and written questionnaires		131
5.1	Canadian English and the multilingual speaker	133
5.1.1	From monolingual to multilingual perspectives	134
5.1.2	From the national to the transnational: The sociolinguistics of globalization	138
	A new kind of sociolinguistics?	139
	Local literary practices and WQs	141
5.2	World Englishes, Global Englishes: Concepts	142
5.3	WQ Elicitation in World English contexts	144
5.3.1	Some problems of WQs in contact scenarios	150
5.3.2	Conceptualizing space in dialect geography	152
5.3.3	Select morphosyntactic features of World Englishes	153
5.4	English as a Lingua Franca (ELF)	156
5.4.1	Concepts	156
5.4.2	Polling language teacher attitudes towards ELF	158
	Group identity, mutual intelligibility and ELF	161
5.4.3	Linguistic error or innovation? The case for WQs	162
5.4.4	Discovering variables and variants	164
	Lexical innovation	164
	Pragmatic innovation: Idiomatcity in ELF	166
	Some principles for variable detection	168
5.5	Addendum: Global Englishes and expert WQs	171
5.6	Chapter summary	173

CHAPTER 6

WQ data and linguistic theory

175

- 6.1 Real time and Apparent time 176
 - 6.1.1 Age-grading 178
- 6.2 The S-curve of linguistic change 180
 - 6.2.1 The case of N/V+ing + N compounds 181
- 6.3 Change from above, change from below: Social class 184
- 6.4 Gender (sex) 185
 - 6.4.1 Principle 1: Stable situations: Women use the standard more than men 185
 - 6.4.2 Principle 2: Women use more standard forms in changes from above 187
 - 6.4.3 Principle 3: Women use more of the incoming variant in changes from below 188
 - 6.4.4 Indexing social meaning: Gender 189
- 6.5 Border effects: Autonomy vs. heteronomy 191
 - 6.5.1 Insights from Dialect Topography 193
 - A cross-border continuum: St. Stephen (New Brunswick) and Calais (Maine) 193
 - Political borders as linguistic divides: *Shone* in Ontario and New York 196
 - Heightened differences in immediate border regions 197
 - 6.5.2 Insights from NARVS: The North American lexical perspective 198
- 6.6 Sociohistorical framework and explanations 200
 - 6.6.1 Canada's five major immigration waves 201
 - 6.6.2 Trudgill and Schneider: Two complementary approaches? 202
 - Trudgill's (2004) New-dialect Formation Theory 203
 - Schneider's (2007) Dynamic Model 205
- 6.7 Indexing social meaning 208
 - 6.7.1 Three "waves" in sociolinguistics 209
 - 6.7.2 Yod-dropping in CanE as an indexical field 211
- 6.8 Homogeneity and heterogeneity 214
 - 6.8.1 Homogenization on the national level 215
 - Homogeneity & Standard Canadian English 216
 - 6.8.2 Homogenization on the continental level 216
 - 6.8.3 Heterogenization (diversification) 217
- 6.9 Chapter summary 221

Part II. Practice

CHAPTER 7

Questionnaire design and data collection	225
7.1 Planning the questionnaire: Purpose & research question	226
7.2 Structure of the questionnaire	227
7.2.1 Questionnaire length	231
7.2.2 Choice of medium: Paper or online?	233
7.3 Question design	234
Self-reporting and community-reporting	235
7.3.1 Types of questions	236
Closed-response items	237
Checklists	237
Multiple-Choice items: Binary (nominal) & categorical	238
Rating scales	238
Multiple-items scales in social dialectology	240
Inter-related items	241
Open-response items	243
Mixed-response types	244
7.3.2 From raw questions to questionnaire items	245
7.3.3 Self-reporting linguistic behaviour	249
Inventories vs. social correlations	249
Socio-syntactic reformulations	250
Magnitude Estimation Tasks and grammaticality judgements	251
7.3.4 Self-reporting language attitudes and perceptions	252
Language attitudes	252
Perceptions	254
7.3.5 Community-reporting of linguistic behaviour	258
7.3.6 Mitigating prescriptive influence: Framing the questions	259
The role of instructions	259
Using informal language in the questionnaire	261
Reliability	261
Some tentative insights: How to ask and how better not	262
Defining evaluative categories	265
Harnessing a pedagogical phonetic alphabet for social dialectology?	266
Order of stimuli and trial items	267
Intuitive formatting & item ordering	267

7.3.7	Piloting and revising the questionnaire	268
7.3.8	Social background questions	269
7.4	Population sampling	270
7.4.1	Random or judgement sampling?	270
7.4.2	A combined sampling method	273
7.5	Chapter summary	274
CHAPTER 8		
	Working with WQ data	275
8.1	The Dialect Topography portal	275
8.1.1	Dialect Topography questionnaire	277
8.1.2	“View Results”: One variable in one location	277
8.1.3	View Results: Comparing two locations	280
8.1.4	Tutorials	283
8.2	Calculating social indices	284
8.2.1	The Regionality Index (RI)	285
	Complex cases	290
	Applying the RI	291
	Adjusting the RI for multilingual respondents	291
8.2.2	Language Use Index (LUI)	292
8.2.3	Ethnic Orientation Index (EOI)	293
8.2.4	Occupational Mobility Index (OMI) and Social Class (SC)	294
8.3	Data-readying in Excel	296
8.3.1	Importing DT data into Excel	296
	PC Users	298
	Text Import Wizard	299
	Mac Users	300
8.3.2	Three basic Excel commands	302
	Principles of Excel	302
	COUNTIF and SUM	304
	Exercises on the file “GH (All) – q1-different”	307
	Multiple conditions: COUNTIFS	308
	Excel graphing tool	310
	More exercises on the file “GH (All) – q1-different”	311
	COUNTIFS with indices: RI	311
	Exercise with RIs	313
8.3.3	Beyond manual commands: Pivot tables	313
8.4	Chapter summary	

CHAPTER 9**Statistical testing with R 319**

Downloading and installing R 320

Descriptive & Analytical statistics 321

9.1 Why use statistical tools? 321

9.2 Preliminaries: Types of variables and forming a hypothesis 323

9.2.1 Types of variables in traditional WQs 324

9.2.2 Formulating a hypothesis 326

9.3 Univariate analytical statistics 328

9.3.1 One dependent variable 329

9.3.2 One dependent variable, one independent variable 333

9.3.3 Converting a categorical variable into an ordinal variable 335

9.4 Multivariate Analysis with R 339

9.4.1 Linear models: Hierarchical Configural Frequency Analysis 340

9.4.2 (Non-linear) logistic regression modelling 346

Importing “data frames” into R 348

The case for data import checking: Two errors 349

A first logistic regression with interactions 351

When you get a “Warning message” 354

Logistic regression modelling without interactions 355

9.5 Chapter summary 359

CHAPTER 10**Epilogue 361**

10.1 The revival of WQs in social dialectology 362

10.2 WQs and linguistic variables 366

10.3 Desiderata 369

10.3.1 Guidelines for WQ design in social dialectology 369

10.3.2 WQs, geographical space, and potential risks 370

10.3.3 WQs and WQ-internal checks and controls 372

10.4 WQs: The delayed method 373

References 375**Index 395**

List of common abbreviations

AAE	African American English
AfrE	African English
ALF	<i>Atlas linguistique de la France</i>
AmE	American (US) English
AusE	Australian English
AutE	Austrian English
AutG	Austrian German
BlSAE	Black South African English
BNC	<i>British National Corpus</i>
BrE	British (UK) English
CanE	Canadian English
ChinE	Chinese English
CL	corpus linguistics
CONTE	<i>Corpus of Early Ontario English</i>
DARE	<i>Dictionary of American Regional English</i>
DCHP-1	<i>Dictionary of Canadianisms on Historical Principles, First Edition</i>
DCHP-2	<i>Dictionary of Canadianisms on Historical Principles, Second Edition</i>
DT	<i>Dialect Topography of Canada</i>
EAFrE	East African English
EFL	English as a Foreign Language
ELF	English as a Lingua Franca
ELT	English Language Teaching
EngEng	English English
EOI	Ethnic Orientation Index
ESL	English as a Second Language
FI	Field-worker interview
GerE	German English
GerG	German German
GhaE	Ghanaian English
IndE	Indian English (East Indian)
IndSAE	Indian South African English
L1	first language
L2	second language
LakE	Sri Lankan English

LAP	Linguistic Atlas Project
LAS	<i>Linguistic Atlas of Scotland</i>
LAUM	<i>Linguistic Atlas of the Upper-Midwest</i>
LAUSC	<i>Linguistic Atlas of the United States and Canada</i>
LUI	Language Use Index
MalE	Malayan English
NARVS	<i>North American Regional Vocabulary Survey</i>
NigE	Nigerian English
NORM	Non-mobile, older, rural male interviewees
OED	Oxford English Dictionary, Third Edition
OMI	Occupational Mobility Index
PhilE	Pilippine English
RI	Regionality Index
SC	Social class
SCE	<i>Survey of Canadian English</i>
SED	<i>Survey of English Dialects</i>
SingE	Singapore English
StCanE	Standard Canadian English
WE	World English
WEs	World Englishes
WQ	written questionnaire
WWQ	Wenker's written questionnaire

List of tables

Table 2.1	<i>Survey of Canadian English</i> , overall results	43
Table 3.1	Lexical item <i>cool</i> in the <i>Corpus of early Ontario English</i> (CONTE), 1776–1849	54
Table 3.2	<i>Cool</i> in the Strathy Corpus of Canadian English (1985–2011)	55
Table 3.3	50 most frequent words in BNC	57
Table 3.4	20 most frequent verbs and noun lemmas in BNC	58
Table 3.5	Comparison of features of three linguistic data collection methods	62
Table 3.6	Data from FI and WQ by state in the <i>Linguistic Atlas of the Upper Midwest</i>	66
Table 3.7	Comparison of 35 lexical variables from LAUM (Allen 1973–76)	69
Table 3.8	Interviewees in the sociolinguistic interviews	78
Table 3.9	Matches between self-reports and acoustic data	80
Table 3.10	Glide deletion in Vancouver: acoustic measurements of yod-ratio	83
Table 4.1	SCE data for <i>chesterfield</i> (question #29) in percent	89
Table 4.2	National SCE data for <i>tap</i> (question #31) in percent	96
Table 4.3	Comparison of OUP dictionary entries for TAKE UP #9	104
Table 4.4	<i>Different from</i> by gender and social class	112
Table 4.5	First person pronoun singular in Old English and Present-Day English	113
Table 4.6	Old English a-stem nouns declension for <i>cyning</i> ‘kings’ (plural)	114
Table 4.7	Cross-tabulation of answers of 2440 responses (Pi 2000)	119
Table 4.8	Trajectory of change for telling the time with approximate time line	120
Table 5.1	Calculating the Regionality Index: Two fictional examples from Toronto	135
Table 5.2	Excerpt from the lexical part of the Bamberg Questionnaire (Krug & Sell 2013)	146
Table 5.3	Extract from grammatical section of Bamberg Questionnaire (Krug and Sell 2013)	146
Table 5.4	Lexical innovation in ELF	164
Table 6.1	Vancouver yod-retention	190
Table 6.2	Replies for <i>sneakers</i> in 14–19 year-old, RI 1–5 (%)	194
Table 6.3	<i>Pop</i> (vs. <i>soda</i>) in New Brunswick and Maine among 14–19-year-olds	197
Table 6.4	Yod-ful <i>avenue</i> , east to west in %	197
Table 6.5	Strength of lexical boundaries between Canada and the U.S.	198
Table 6.6	Strength of lexical boundaries between Canadian regions	199
Table 6.7	Canada’s five major immigration waves	201
Table 6.8	Three stages in Trudgill’s New-Dialect Formation Theory	203

Table 6.9	Phases in Dynamic Model applied to Canadian English (Schneider 2007)	205
Table 6.10	Regional dialects in Canadian English based on Boberg (2005, 2008)	218
Table 7.1	Perceived pleasantness of dialects (McKinnie & Dailey-O'Cain 2002)	256
Table 7.2	Large-scale sociolinguistic studies that include audio-recordings	272
Table 7.3	Sample sizes in two Vancouver WQ studies	273
Table 8.1	Overview of Dialect Topography data	276
Table 8.2	Independent (social) Variables in drop-down menu of DT web portal	278
Table 8.3	Calculation schema for the Regionality Index	286
Table 8.4	<i>Has drank</i> (not <i>has drunk</i>) in percent by OMI	295
Table 9.1	Fictitious example for yod-less and yod-fulness in <i>student</i> , low n's	322
Table 9.2	Fictitious example for yod-less and yod-fulness in <i>student</i> , high n's	322
Table 9.3	Types of variables in traditional WQ studies	325
Table 9.4	Overview of statistical procedures and sketch of application criteria	328
Table 9.5	Carbonated drink in Quebec City	329
Table 9.6	Answers to <i>roof</i> rhyming with <i>hoof</i> in absolute frequencies (<i>ns</i>) to Question 11	333
Table 9.7	Four major variant types in q74 (<i>wedgie</i>)	337

List of figures

Figure 2.1	LAMSAS data for variable (chest of drawer) (Kretzschmar 2009)	31
Figure 2.2	Dialect map (display and interpretive lines) for the pronunciation of route	41
Figure 3.1	Allen's data for PQ and FI (grey) and normal distribution (black)	70
Figure 3.2	Vowel plots for <i>caught/cot</i> (non-normalized)	79
Figure 3.3	Vowel plots for <i>Don/Dawn</i> and <i>sorry/sari</i> (non-normalized)	81
Figure 4.1	<i>Chesterfield</i> in the Greater Toronto Region and environs, 1991/2	90
Figure 4.2	Data from Figure 4.1 interpreted with the apparent-time hypothesis	91
Figure 4.3	<i>Tap</i> and <i>faucet</i> in the Dialect Topography of Canada database	97
Figure 4.4	Greater Toronto (1991/2) <i>tap</i> vs. <i>faucet</i>	99
Figure 4.5	Correct answers TAKE UP #9, by the place of formative years (8–18)	101
Figure 4.6	Canada Census data 1971–2012, In- and Out-migration of ON and AB	103
Figure 4.7	<i>dove</i> (not <i>dived</i>) and <i>snuck</i> (not <i>sneaked</i>) in the Golden Horseshoe by age (1991/2 data)	106
Figure 4.8	<i>Snuck</i> (not <i>sneaked</i>) by internet domains in % (31 May 2010 & 4 March 2015)	107
Figure 4.9	<i>Different from/than/to</i> in the Golden Horseshoe	110
Figure 4.10	<i>Different from</i> in seven Canadian regions	111
Figure 4.11	Non-standard <i>between you and I</i> by education	115
Figure 4.12	Analog answers in the greater Toronto region	118
Figure 4.13	<i>Student</i> pronounced with yod-dropping, BC (SCE), Vancouver and Toronto (DT)	123
Figure 4.14	<i>Avenue</i> with yod (you) in seven Canadian and three American locations (DT)	124
Figure 4.15	<i>Vase</i> rhyming with <i>days</i> , i.e. /veɪz/	126
Figure 4.16	<i>Vase</i> in Toronto and Vancouver (%) by strength of local ties (Regionality Index)	129
Figure 5.1	<i>Soother</i> and <i>dummy</i> in six varieties (14 May 2014)	148
Figure 5.2	Variant responses in traditional Inner Circle & hypothetical super-diversity settings	151
Figure 5.3	First-ranked varieties for “best” English accents among L2 ELT teachers	161
Figure 6.1	Deontic Obligation Markers in Canadian English	177
Figure 6.2	<i>Zed</i> in the Golden Horseshoe, 1991/2 and 2000	179

Figure 6.3	Depiction of S-curve of linguistic change	181
Figure 6.4	N/V+ing + N, 1980–2007	182
Figure 6.5	(ing) in Norwich, England	186
Figure 6.6	<i>Between you and me</i> (vs. I) in Ottawa	187
Figure 6.7	Subjunctive <i>was/were</i> in Ottawa	187
Figure 6.8	Quotatives in Toronto English (ages 9 to >80)	189
Figure 6.9	Canadian English & Canadian identity, Vancouver 2009 (%)	192
Figure 6.10	‘Athletic shoe’ in seven Canadian and four American regions (%)	194
Figure 6.11	<i>Shone</i> pronounced /ʃan/ in the Golden Horseshoe	196
Figure 6.12	Possible indexes of yod-ful <i>student</i> in Canadian English	212
Figure 6.13	Vowels before [r] in <i>guarantee</i> for respondents under 40	220
Figure 7.1	Position of questions and error rate by age of respondent	231
Figure 7.2	Answers to “Is there a Canadian way of speaking?” by education	253
Figure 8.1	<i>Soft drink</i> and <i>pop</i> by RI in Quebec City	291
Figure 8.2	<i>Sofa</i> in Ottawa Valley according to LUI and mother tongue	293

List of illustrations

Illustration 1.1	Types of linguistic data by degree of monitoring and naturalness	5
Illustration 2.1	Social background questions in Hempl's WQ	29
Illustration 2.2	Part of the questionnaire of the 1972 <i>Survey of Canadian English</i>	42
Illustration 4.1	<i>Chesterfield</i> (left) vs. <i>chesterfield</i> (right)	88
Illustration 5.1	Visualization of Kachru's Three Circle Model	143
Illustration 6.1	Youngest example of <i>dumping truck</i> in <i>The Globe (and Mail)</i>	183
Illustration 6.2	Canada-U.S. border along "0 Avenue" in Surrey, BC (left), and Blaine, WA (right)	193
Illustration 7.1	Informed Consent Form (UBC 2013)	228
Illustration 7.2	Beginning of Washington State English Survey (2011)	229
Illustration 7.3	Instructions for the linguistic part (Washington Survey)	230
Illustration 7.4	Raw material from Preston's 1984 map task	254
Illustration 7.5	Perceived dialect regions by 100 Ontarians and 100 Albertans	257
Illustration 7.6	Instructions from Gregg's BC Linguistic Survey	260
Illustration 8.1	DT of Canada Introductory Screen ("ABOUT")	276
Illustration 8.2	View Results screen	278
Illustration 8.3	Basic output for View Results function	279
Illustration 8.4	Automatic Graphing feature (full circles) with manual additions (semi-circles)	280
Illustration 8.5	Regional comparison feature	281
Illustration 8.6	Output of Regional comparison function	282
Illustration 8.7	Tutorial View with Split Windows	283
Illustration 8.9	Raw download format (Full Data Length)	297
Illustration 8.8	Data Request home screen (Dialect Topography)	297
Illustration 8.10	Importing the text into Excel 2010	298
Illustration 8.11	Copy-and-pasting the delimiting character into the Excel	299
Illustration 8.12	Imported data in Excel	300
Illustration 8.13	Mac procedure for inserting the Pipe symbol via Copy-and-Paste	301
Illustration 8.14	Downloaded DT file, Golden Horseshoe (All), q1-different, in Excel	303
Illustration 8.15	COUNTIF example	304
Illustration 8.16	Basic variant count for q1-different	305
Illustration 8.17a	SUM command	305
Illustration 8.17b	SUM command & percent calculation	306

Illustration 8.18	Creating a percent formula that is automatically “shiftable”	306
Illustration 8.19	Shift box (top) and full percentages with “shifted” function	307
Illustration 8.20	COUNTIFS in age groups	309
Illustration 8.21a	Formula dragged to the right	309
Illustration 8.21b	First correction of dragged formula	309
Illustration 8.22	q1-different by age group & three major variants	310
Illustration 8.23	Marking data for graphing tool and chart	310
Illustration 8.24	COUNTIFS and RI	312
Illustration 8.25	Calling up the PivotTable function in Excel	314
Illustration 8.26a	Submenu for command “PivotTable...”	314
Illustration 8.26b	Responses for q1 per age cohort	315
Illustration 8.27a	Manipulate-able Pivot Table for q1 and all other variables (header in top row)	316
Illustration 8.27b	Answer for <i>different from</i> only in display view in PivotTables	317
Illustration 9.1	Normal distribution with 95% area shaded (95% likelihood or 0.05 cut-off)	330
Illustration 9.2	Histogram for <i>wedgie</i> and variants in Vancouver	339
Illustration 9.3	Input routine for HCFA testing	343
Illustration 9.4	Select all tables from 0000 to 0015 (in our case)	344
Illustration 9.5	HCFA_output_sum.txt for differentvan.txt	345
Illustration 9.6a	Montreal.news.txt: original output in “q42”	349
Illustration 9.6b	Montreal.news2.txt: original q42 replaced with R-compatible format	349
Illustration 9.7a	Montreal.news2: missing value for dependent variable in line 234, id 5234	350
Illustration 9.7b	Montreal.news3: missing values removed	351
Illustration 9.8	Logistic Regression modelling, model 1, with interactions. AIC not shown	353
Illustration 9.9	Model 2, model.glm2, without interactions	356
Illustration 9.10	Model 3, model.glm3, without interactions	357
Illustration 9.11	Model 4, model.glm4, without interactions	358

Acknowledgements

A debt or gratitude is owed to many people.

Commissioning editor **Kees Vaes** is thanked for realizing the relevance of this topic and suggesting the proper format. **Ana Deumert** for seeing the potential of this book and for soliciting the perfect pair of referees. I am indebted to one anonymous proposal reviewer and to two anonymous manuscript reviewers, who offered most valuable critique.

Vielen Dank to **Erik Schleef** for most incisive, measured and detailed feedback on a draft version of the entire book. **Manfred Krug** for reading Section 5.3 in the right light and for volunteering unpublished data from his Malta study. **Stefan Th. Gries** for lightning-fast and good-spirited help with R and his permission to include the HCFA code on the companion website. **Sandra Clarke** and **Suzanne Power** for help with questions around WQs in Newfoundland, **J. K. Chambers** for piquing my interest in questionnaires in the first place. All shortcomings remain, as always, my sole responsibility.

Ruth Dollinger for reading the entire first draft of a hastily composed manuscript and for tolerating my chronic tardiness. Thanks must go to my Joseph-Conradian editor, **Ania Basiukiewicz**, for straightening out the more opaque ELF constructions I failed to notice in the heat of the writing battle. Since I need to live by what I preach, I occasionally opted for my own innovations regardless.

The STUDENTS in my UBC ENGLISH LANGUAGE MAJORS SEMINAR (ENGL 489) in the fall of 2014, who, working hard, pushed the limits of the method, and the course's 2012 and 2013 installments for teaching me what does not work. My lab assistants and junior colleagues **Sasha (Alexandra) Gaylie** and **Jessica Tam** for taking on, besides their work on DCHP-2, odd jobs and ends, and always happily so.

My Viennese teachers I offer thanks: **Henry Widdowson** and **Barbara Seidlhofer** for introducing me first to *englische Sprachwissenschaft*. **Nikolaus Ritt** for showing me that English linguistics does not stop with current English linguistics literature and, especially, **Herbert Schendl**, who is always there with the right *doktorfatherly* words of advice. Finally, **Richard Schrodtt** for relating the Austrian German vernacular to the bigger linguistic picture and allowing me to connect my *Anglistik* with my *Germanistik* existence.

My UBC colleagues, **Elizabeth Hodgson** for her patience and good counsel, **Margery Fee** for her encouragement in 2002, 2015 and in between, **Laurel Brinton** for her high-principled input. **Ira Nadel**, **Mary Chapman**, **Barbara Dancygier** and **Tiffany Potter** for straight talk and support.

Vielen Dank, tack så mycket and thank you.

Abstract

The methods and procedures used to collect linguistic data comprise some of the most central aspects in social dialectology, the study of regional and social variation in language. Since the early 20th century, interview methods have been preferred over the “indirect method” of written questionnaires. While written questionnaires hitherto played only a minor role in the field at large, the last decade or so has seen some sort of revival in a number of subfields and various innovations have pushed the limits of this method. It therefore stands to reason that there is more to written questionnaires than usually meets the (linguistic) eye.

This book is the first monograph-length account on the theory, history and administration of written questionnaires in the study of regional and social linguistic variation. Reconnecting to a questionnaire tradition that was last given serious treatment in the 1950s, the present book combines the older practice with more recent instantiations and reincarnations and offers an up-to-date, near-comprehensive treatment for the newcomer to the method and the beginner in empirical linguistics and sociolinguistics alike.

The text explores the advantages and limitations of written questionnaires in social dialectology in two distinct, yet connected parts: a historical-theoretical and a practical part. The scene is set with a re-evaluative history of the use and avoidance of written questionnaires in traditional dialect geography and sociolinguistics since the late 19th century, with a special focus on English. Methodological comparisons of interview and corpus data with written questionnaires throw into sharp relief the written questionnaire’s strengths and weaknesses, which are illustrated with detailed linguistic variables from traditional (dialect geography) and novel contexts (Global Englishes and English as a Lingua Franca). The most pervasive sociolinguistic theories are explained and contextualized with examples and case studies from Canadian English, a variety that has by historical and geographical accident greatly benefitted from written questionnaires.

The practical section is a guide for the newcomer to the field. It caters to the needs of advanced undergraduate and graduate students, was written with special consideration for students in the Arts and Humanities and assumes no knowledge of quantitative linguistics. This part leads readers through a step-by-step process from start to finish, from formulating a research question to the interpretation of (statistically enhanced) data analyses. In the second part, readers should acquire the necessary

skills for conducting their own written questionnaire studies, from question design and data administration to the tabulation and statistical testing of the most typical variable types of written questionnaire data. The book is addressed to anyone wishing to use written questionnaires for the study of language variation and change and will be of relevance to linguistic geographers, social dialectologists, variationists and sociolinguists of many stripes.

Companion website

All data files for Excel (Sections 8.3.2 and 8.3.3) and R commands (Chapter 9) can be downloaded from the book's companion website:

<<http://dchp.ca/stefan/WQ.html>>

or

<<http://dx.doi.org/10.1075/impact.40.website>>

Author's preface

When somebody studies a method without the direct assistance of an experienced practitioner, one is forced to glean insights from the existing literature on the one hand, and to learn by trial-and-error on the other hand. In linguistics, one will quickly find that articles reporting results based on the method in question do not always offer, or if then only in a very limited way, practical instructions on how to proceed. This book has its origins in such attempt. When I first used written questionnaires in the summer of 2008, I was trying to make sense of the methodological sections of existing questionnaire studies, but quickly realized that the method appeared to be only loosely defined and that practices would sometimes stand in outright contradiction.

The problem of an apparent lack of universal guidelines became especially obvious when the method was presented to students, whom I required to collect data in a UBC course on Varieties of English (ENGL 323A) in the fall of 2008. While the results from this class-based survey were as good as those published (some discussed in Dollinger 2012a, 2012b), it became obvious that a lot of methodological potential remained untapped. As attempts to gather information on principles and best practices from those who had been actively engaged with written linguistic questionnaires was not particularly helpful either, the idea of composing some sort of “guide” for the design of linguistic questionnaires was first conceived in the spring of 2009. Over the years, the blueprint of the book was extended more and more, including theories, some of long-standing and some newer ones, as three draft manuscripts were tested in advanced upper-level undergraduate courses.

Originally, this book was intended as a combination of previously published articles and some newly commissioned papers and section introductions. It is thanks to Kees Vaes of John Benjamins, who, while seeing the potential in the idea, suggested that the text should offer maximum coherence. At a time when Praat and sociophonetics were already buzzwords and were being used in more and more contexts by more and more people, a proposal by a junior scholar on an apparently “old-fashioned” method of yesteryear might have seemed strange on many an editor’s desk. Not so with John Benjamins and Ana Deumert’s *IMPACT* series, whose assistance resulted in a much improved book. I hope that the outcome will at least in part meet with approval.

Experience with multiple methods improves any field. Should this book facilitate this overall goal somewhat, it will have served its purpose.

Vancouver, Canada, 1 May 2015

Chapter 1

Written Questionnaires in the wider linguistic context

The discipline of linguistics and language study has long been straddling the demarcation line between the humanities and the social sciences, which has, traditionally but simplistically, been defined as qualitative in the former and quantitative in the latter case. A particularly vexing issue is the nature of evidence deemed admissible in linguistic inquiry and, consequently, the definition of what precisely constitutes that discipline. In more than one way, one's view of linguistics is shaped by evidence that is permitted, or, in other words, by the methodologies that are accepted. Most disciplinary disputes seem to be anchored in disagreements about the nature of linguistic data. This book will address one aspect of that debate.

Linguistics has seen more than its fair share of competing approaches over the course of the 20th century. In hindsight, it seems as though an earlier period's universal agreement about the goals of the discipline had to be seriously challenged. Until about World War I, linguistics was a discipline unified by the commonly accepted methodology of comparative historical linguistics, often called, or considered a vital part of, *philology*. The goal of comparative linguistics was to reconstruct earlier stages of language development, within or across languages, and to establish the lineages between them. The quest for the Indo-European language, ancestor to most European and some Asian languages and spoken till about 4,000 BC, was a particularly successful venture for much of the 19th and the early 20th centuries. Philological greats like Jacob and Wilhelm Grimm (the Grimm Brothers of fairy tale fame), Karl Verner, Julius Pokorny, August Schleicher, the Junggrammatiker (Neogrammarians) Karl Brugmann and Hermann Paul, or the Anglicist Karl Luick (1964 [1914–40]), to name but a few, propelled the comparative historical method to unprecedented heights.

Beginning with Ferdinand de Saussure (1916), an array of different perspectives and foci on language study developed and the available methodological approaches multiplied, if not to say dispersed. Saussure, among other things, introduced two principled distinctions that had profound consequences on the conception of the field. First, Saussure distinguished between a speaker's actual instances of speech or language (*parole*) and language as a collectively shared abstract system (*langue*). Most significantly, he declared *langue* the focus of study. Second, he separated the diachronic (historical) study of language, which was de facto the overwhelmingly dominant mode of activity, from the synchronic (contemporaneous) study of language. Importance was

placed on synchronic perspectives, making historical perspectives no longer the only, or even the most important, kind of perspective.

Since Saussure's work, linguistic approaches have multiplied and several schools have developed. Noam Chomsky's generative linguistics is one such school, a school that is based on a dichotomy between *competence* and *performance*. These are not to be confused with Saussure's conceptual pair. *Competence* represents the speakers "tacit knowledge" of linguistic structures, the abstract idea of language, while *performance* includes all practical "limitations", such as fatigue, memory restrictions, and, crucially, concrete uses of language in given situations. While with *competence* Chomsky declared the object of (his kind of) linguistics, *performance* was relegated to the status of an epiphenomenon, to be merely controlled for and abstracted away from. Chomsky's categorical dismissal of usage as *performance*, and thus beyond the scope of the discipline, and his elevation of *competence*, the tacit knowledge of the infamous "ideal speaker-listener" caused great concern among linguists of many persuasions. With one strike, many linguists' objects of study were ruled out as beyond the scope of linguistics by Chomsky, whose school concerned himself exclusively with

an ideal speaker-listener, in a completely homogeneous speech-community, who knows its [the speech community's, SD] language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations [and the like, SD].

(Chomsky 1965:3)

Many protested against this narrow definition of linguistics and the field was, quite understandably but regrettably, rife with dispute. In the early 1970s, Derwing (1973: 25) characterized the time since 1957 (Chomsky's first monograph) as a period when the discipline gave "the appearance of being racked with disputes, lack of communication, even downright hostility – almost as though it were organized into armed camps". Philologists and experimental linguists alike were outraged over a field definition that excluded their objects of interest as performance, which was supposed to be sidelined.

Shortly after Chomsky's proposal, a new school of thought gestated and would be established by William Labov, Peter Trudgill and others, who (re)discovered in a quantitative framework that patterns in precisely those aspects ruled out by Chomsky were systematic (e.g. Labov 1972). The new school's focus on the social dimension of language use was, while building on earlier dialect geographical methods, in that consequence a novelty, giving rise to the field of variationist sociolinguistics. Quantitative methods are at the centre of this discipline that focuses, in contrast to philologists and historical linguists, on the spoken language.

Labov's quantitative approach was, of course, not utterly new. At least since the 1920s, corpus linguists had approached data from a bottom-up perspective and aimed

to synthesize linguistic principles from linguistic behaviour, which was accessible via text collections (“corpora”). Early corpus linguists such as Charles C. Fries (e.g. 1925), in Michigan in the American context, or, in the European tradition, Alvar Ellegård (1953) in Gothenburg, Sweden, showed that a principled approach to quantification – then carried out with pen on paper – would produce important insights into language structure. This kind of empirical approach returned on a grand scale only in the 1980s with the advent of affordable home computers, which greatly facilitated the time-consuming and tedious but necessary tasks of searching for and counting linguistic forms (see Kretzschmar 2009: 6–63 for a summary). Today, a basic quantitative method is part and parcel of almost all schools of thought and there seems to be agreement, even in most generativist circles, that some form of data collection is required beyond immediate native speaker introspection.

So much shall suffice on the wider disciplinary context of language variation, and the regional and social study of language. The present book explores one method of data collection, or rather a group of methods, that employs written questionnaires (WQs) for sociolinguistic, dialectological and variation studies. These approaches combined shall be collectively referred to as “social dialectology”. The term has a somewhat older ring to it, which is fully intended as it harks back to a period in which philological approaches were still part of the linguistic discourse. The term’s simultaneous coverage of social and regional variation and its implied link to historical approaches make it the term of choice. This book, with its focus on regional and social linguistic variation, expressly includes the study of language attitudes and issues of linguistic perception. The examples in this book come from European languages, with a good deal from English. Canadian English will be given a prominent place in the case studies, because of the method’s enduring legacy in that field and the prevalence of WQs in that variety’s scholarship.

This introduction is setting the stage for WQ studies in the subdisciplines that I call social dialectology. Immediately following, however, a further step back will be taken on the question of data and evidence in the field with the goal to isolate some of the characteristic, high-level similarities and differences between the most important data collection methods. Thereafter, a brief account of traditional dialect geographical projects, starting with Gilliéron’s and Wenker’s paradigm-setting approaches, will be given. The introduction will show, as will several examples within the book, that Wenker, while a pioneer in many important ways, is quite incorrectly considered the archetypal proponent of the WQ method. This distinction will be reflected in the terminology used, in which WWQ is used for Wenker’s WQ and WQs for modern-day WQs. The final part of this introduction then characterizes this book’s two parts and nine following chapters.

1.1 Three basic types of language data and WQs

Three basic types of data collection have traditionally been distinguished in language study. These are:

- Introspection
- Elicitation
- Observation

Introspection is the method used in standard generativist theory: centred on the idea of the native speaker as the ultimate and best judge of the “grammaticality” of sentences, inferences are made about language structure. Termed “armchair linguistics” (Fillmore 1992), this method in its most extreme form requires only one native speaker of the language – often the linguists themselves – to produce “data”. Elicitation requires more effort and is today one of the most widespread techniques across various subdisciplines. Generative linguists working on endangered languages (see, e.g., Rau 2013) habitually elicit linguistic structures by asking informants “how do you say X or Y in your language” or the like. The way elicitation is usually practiced means that a linguist with no or with limited knowledge of the target language is gleaned insights into the language. Another elicitation technique entails a more informal polling that is often carried out by linguists of all persuasions in the form of “Do you know the term/meaning/construction ...?”. As preliminary hypothesis-building attempts, these techniques are entirely reasonable. Generally, though, such anecdotal reports are not taken as evidence and need to be substantiated with more solid data.

Observation is considered the most superior form of data collection by many linguists, the “gold standard”, so to speak, especially pertaining to naturally occurring speech (i.e. not in formal settings, performance contexts etc.). Rather than relying on a speaker’s introspection or on the responses elicited from someone conversant in the language, utterances – whether in writing or speech – are collected after the fact and then, in a second instance, systematically analyzed. Corpus linguistic data is the only kind of data that can strictly be classified as observation,¹ as will be discussed in Chapter 3. Corpora are collections of text (either spoken, written or both) that are constructed post-hoc, either by sampling texts, or (transcriptions of) recordings, which were made for different purposes. One limitation is that most corpus material today still consists of written texts, since this medium is much more easily accessible than spoken language. The goal of variationist sociolinguists is to elicit data that is as

1. Even sophisticated ethnographic observation is subject to the observer’s paradox, in like manner as the sociolinguistic interview. What is available are mitigation procedures, but no real solution to the problem outside of corpus linguistics, as is addressed later in this chapter and in Chapter 3.

minimally monitored and as natural and informal as possible. Labov developed a form of elicitation interview called the sociolinguistic interview (see Becker 2013), designed to elicit a number of speech styles, from more formal to less formal ones, the latter of which were considered the most-prized and most natural forms of human language, i.e. the “vernacular”.

Illustration 1.1 depicts the possible approaches to empirical data collection in linguistics, with the exception of introspection (Krug, Schulte & Rosenbach 2013: Figure 3). The methods are arranged on a scale from least natural (–natural), which allows speakers to actively “monitor” or manipulate their linguistic output, perhaps in accordance with socially desirable norms, to the most natural (+natural) and least monitored (–monitor) types of data. In this hierarchy, corpora are considered as more desirable forms than data from the observation category, as only in corpus data with its reliance on “authentic” materials from natural situational contexts, no observing or participant-observing researcher is present who might interfere with or influence linguistic performance (“observer’s paradox”, see Chapter 3). Granted, there are good workaround and mitigation procedures to reduce this effect, but for the sake of the present argument unless one works with texts created for different scenarios than the research purpose, and with no researcher, whether insider or outsider participant, an observer’s paradox is interfering to one degree or another.

WQs are elicitation techniques and are to be classified as “metalinguistic” in Illustration 1.1 since they generally present direct linguistic questions (not all do, see Chapter 7). As addressed above, there has been a bias against the use of WQs in social dialectology. As will be shown in Chapter 2, from the historical perspective, and from a different angle in Chapter 3, WQs are generally not employed in many variationist studies (see Boberg 2013 for the few exceptions), so that their use demands a special case be made in each instance. In corpus linguistics, somewhat differently yet related, a focus on authentic examples from real-life contexts also creates a bias against WQs, which do not offer such attributes.

+natural –monitor		Corpora		(Surreptitiously recorded) spontaneous speech (Various genres of) written texts
		Observation		(Surreptitious) participant observation Unconcealed observation with subject consent
		Elicitation		Sociolinguistic interviews Metalinguistic interviews and questionnaires
–natural +monitor		Experimentation		Minimally invasive experiments Invasive experiments

Illustration 1.1 Types of linguistic data by degree of monitoring and naturalness (Krug, Schlüter & Rosenbach 2013: Figure 3)

Despite the differences, sociolinguistic interviews and WQs share a common characteristic as elicitation techniques, which becomes clear from Illustration 1.1. Sociolinguistic interviews, however, have the advantage of the spoken medium and audio recordings, a trait that WQs cannot offer. Instead, WQs have the advantage of ease of use, and ease of dissemination, collection and analysis, as will be shown in detail in later chapters. The dividing line between Corpora and Observation is of a more categorical nature as shown in Illustration 1.1, since the ethical treatment of participants usually forbids surreptitious recordings in most, if not all, contexts today. Corpora, therefore, need to be set aside as offering a level of authenticity that neither WQs nor sociolinguistic interviews can provide.² WQs and interviews are elicitation methods, working with different media, but otherwise sharing a number of features. Interviews, however, have been the method of choice in dialect geography for more than a century, a fact that will be explored next.

1.2 Data in traditional dialect geography

Traditional dialect geography focuses on the description of linguistic variation across geographical space. It charts this variation by location, e.g. in location A, speaker X uses linguistic form Y, and the like. The major projects of dialect geography are linguistic atlases, which are thick and large volumes, ideally with maps. They are, unfortunately and inconveniently, completed over longer periods of time, which are generally measured in decades not in years. Whether one looks at the *Deutscher Sprachatlas*, or the *Linguistic Atlas of the United States and Canada* (LAUSC) or the *Survey of English Dialects* (SED), one common denominator binding them together is that it took several decades from the start of the data collection to their actual publication or (partial) completion. With such lengthy projects, it seems clear that methodological innovations often “overtake” works-in-progress. And that is a good thing: if over the course of two or three decades no new methods were invented, something would seriously be wrong in any field.

In order to understand the disciplinary contexts, it is necessary to briefly recount the discipline’s historical development. Standard textbooks in dialect geography and dialectology (e.g. Nelson 1983; Chambers and Trudgill 1998) generally give prominence to two pioneering, and methodologically very different projects: Georg Wenker’s postal questionnaire method from the 1870s, which would eventually epitomize in the *German Linguistic Atlas* (*Deutscher Sprachatlas*) on the one hand, and Jules Gilliéron’s

2. The bottom of the issue is whether corpora are considered, like the other three forms, a data collection method. This is, in my mind, clearly the case, though there are different opinions on this precise issue.

fieldworker interview method, which was developed for the *Linguistic Atlas of France* (*Atlas linguistique de la France*, Gilliéron & Edmont 1902–1910), on the other hand.

1.2.1 The Fieldworker Interview (FI) method

For the longest part of the 20th century, the fieldworker interview (FI) method was the undisputed method of choice. The FI method was developed by Jules Gilliéron, the Swiss director of the *Atlas linguistique de la France* (ALF), who produced the work based on his sole fieldworker's transcriptions. Gilliéron pioneered the method for his first dialect atlas on the southern Rhône dialects, published in 1880. He surveyed these while hiking the area (Lamelli 2010: 576–7). Gilliéron's fieldworker Edmond Edmont, on his part, is known to legions of dialectology students as the 'bicycling fieldworker' (e.g. Chambers & Trudgill 1998: 17); while Edmont cycled around France, sending his transcriptions off as they became available, Gilliéron edited ALF in record time in faraway Paris. From 1896, when fieldwork began, to 1910, when the last volume of ALF appeared (Gilliéron & Edmont 1902–1910), only 14 years passed. Their method of the face-to-face interview is often referred to as the 'direct method' in dialectology, as opposed to the 'indirect' method of written questionnaires. The FI method, obviously, relies heavily on trained and skilled fieldworkers who transcribe – in the pre-audio-recording period – answers with pen on paper in narrow phonetic script.³

The FI method would become the preferred method for dialect geography in English and Romance linguistics. While originally fieldworkers asked questions and transcribed the interviewees' answers precisely and immediately, starting around the 1940s, FIs would be occasionally and partially recorded. It was not until the late 1960s, however, that a linguistic atlas project was taped in its entirety (Lee Pedersen's *Linguistic Atlas of the Gulf States* as part of the *Linguistic Atlas of the US and Canada*). After ALF, the FI was first applied in Italy and the Italian-speaking regions of Switzerland, which produced a much-celebrated linguistic atlas (Jaberg & Jud 1928–1940). Switzerland became a hub for important dialectological projects so that in the years 1940 to 1958, fieldwork for the model-defining *Sprachatlas der Deutschen Schweiz* (SDS) was undertaken, with volumes being published between 1962 and 2003. SDS was a methodological continuation of the Italian-Swiss Atlas and ALF alike and established a model for smaller regional atlases that are still the norm in German-speaking areas (Scheuringer 2010: 167). Romance dialectology as a whole continued to follow ALF's lead and its preference of FIs (see Pop 1950).

Without much delay, the fieldwork method also found its way to the United States where Hans Kurath, the Austrian-raised American dialectologist, began the *Linguistic*

3. ALF's first volume (which contains no maps), can be accessed in open access at <<https://archive.org/details/atlaslinguistiq00edmogooq>> [1 May 2015].

Atlas of the United States and Canada (LAUSC) in 1929. Kurath split the vast North American continent into subareas and introduced, as occurs in every new project, some methodological innovations. Methodologically, Kurath took over Gilliéron's method "as refined and modified for Italy by Karl Jaberg and Jakob Jud" (Atwood 1986 [1963]: 67) and studied the method in person with members of the Italian team. The FI became the method of choice of almost all⁴ aspects of the *Linguistic Atlas of the United States and Canada* projects, whose fieldwork continued into the 1990s.

The first and paradigm-setting atlas in the English context was the *Linguistic Atlas of New England* (LANE, Kurath et al. 1939, Kurath et al. 1972), followed by the *Linguistic Atlas of the Middle Atlantic States* (LAMSAS, Kurath 1949; Kurath & McDavid 1961), the *Linguistic Atlas of the North and Central States* (LANCS), for which fieldwork was completed, but no publications came forth. Two other important dialect atlases are the *Linguistic Atlas of the Upper Midwest* (LAUM, Allen 1973–6), to be discussed further in Chapter 3 for its special relevance for WQs and the ultimate *Linguistic Atlas of the Gulf States* (see Pederson, McDaniel & Adams 1986–93), which, directed by Lee Pederson, is perhaps the "best" LAUSC atlas. For an up-to-date list see the *Linguistic Atlas Project* website, see <<http://www.lap.uga.edu/>>, which, for the first time, offers an easily accessible, clearly formatted and organized overview of the many LAUSC projects. The website is run by the *Linguistic Atlas Project*, the umbrella organization for LAUSC and other projects, which is directed and curated by William Kretzschmar Jr. The website offers more LAUSC content than previously available, including the entire, hitherto unpublished LAMSAS data set, which is in terms of the number of interviews the most comprehensive LAUSC atlas.

LAUSC reveals the long reach of Gilliéron's method with a direct lineage from Gilliéron's ALF to Kurath's LAUSC, and consequently to Raven I. McDavid and now William A. Kretzschmar, who succeeded Kurath in their positions as atlas directors. With the exception of preliminary work by Guy Lowman, the key fieldworker on Hans Kurath's team, in 1937/8,⁵ the FI method was *not* used in England until after World War II when data from 313 locations was gathered between 1950–1961 for the *Survey of English Dialects* (SED). Kloeke (1952: 134) speaks at the time of Scotland and England as having "no tradition in linguistic cartography". Paradoxically, after the original project was completed, SED did not offer maps and, at first, published only the raw data in table format and no cartographic representation (the so-called "Basic Materials", Orton

4. Some exceptions are discussed in Chapter 2, and include Atwood (1962) in the US South and Bright (1971) in the US West.

5. The fieldworker method, interestingly, seems to have been brought to England from the US and not from France. Guy Lowman, principal and legendary fieldworker of LANE, carried out fieldwork in Southern England just before the outbreak of World War II. This unpublished data is analyzed in Viereck (1975).

et al. 1962–1971). One had to wait until the late 1970s before the first (!) scientific dialect maps of England appeared (e.g. Viereck 1975; Orton, Sanderson & Widdowson 1978; Upton & Widdowson ¹1996, ²2006; Viereck & Ramisch 1997).

England is a comparatively late FI adopter, and continental European scholars, above all the Swiss Eugen Dieth, must be credited for bringing the method to the UK (Mather & Speitel 1975, I: 6). SED and LAUSC overlap in about a quarter of their variables (McDavid 1953a: 566), which allows for interesting inferences of historical input and linguistic change in the former North American colonies and the motherland. All of the mentioned projects are based on FIs and FIs depend, as one of their key characteristics, on very detailed and extensive interviews that generally take more than one day to complete and produce large amounts of data from each interviewee. As a consequence, all projects have the common limitation of the numbers of interviewees they can process, i.e. small samples.

To illustrate this point, France's ALF is based on 639 interviews for the entire country and the Italian and Southern Swiss Italian Atlas on 387 locations (available online, see Tisato 2009). The major atlases in the USA interviewed 208 people (LAUM), 416 (LANE), 564 (LANCS), 1118 (LAGS) and 1162 (LAMSAS) respectively, with usually two or three interviewees per location. LAMSAS and LAGS comprise by far the most extensive data sets. LAMSAS, which has never been published on paper, is now available on the Linguistic Atlas Project website in full and for the first time since the completion of fieldwork in 1974. LAGS, by contrast, was published in paper and conducted rather swiftly from 1968, the start of fieldwork, to 1993, when the last volume appeared.

In English dialectology, two schools of thought exist on the FI. The American dialect geography tradition used "work sheets" for data elicitation, which left some room for the individual fieldworkers to find adequate ways to elicit a variable in indirect ways, i.e. without mentioning the target pronunciation, word or construction (see p. 67, question #20.3b for an example). In England, by contrast, the SED was more restrictive by ensuring that every fieldworker asked precisely the same question of each informant. One can see advantages in the US method, such as more lively conversations that tend to reduce the level of monitoring, but also some disadvantages, such as interference from differently worded questions. In the end, though, no matter the approach, both schools share their unequivocal focus on the fieldworker, as expressed here for the SED:

No matter how well and ingeniously the questions are drawn up, a questionnaire [for use in fieldwork] will not work or produce the desired results unless it is handled by a competent fieldworker. Much depends upon his [or her, SD] conduct of the interview: there is an art in asking questions in a lively and sympathetic way. Naturally, the questions cannot be put to just anybody. The informant must be both knowledgeable and intelligent, and also quick to respond. (Dieth & Orton 1952: vii)

Dieth and Orton's term *questionnaire* refers to a guide for the fieldworker which is not to be confused with the WQ that is filled out by a respondent. The fieldworker questionnaire is merely a list of stimuli to be elicited and may include ways to elicit them. As much as traditional dialectology focuses on the FI, variationist sociolinguists would focus in like manner on the sociolinguistic interview, which can be considered a methodological advancement in the sense that it operates with a more explicit structure (speech styles) (see Becker 2013 for a summary).

Generally speaking, interview methods and protocols have received a great deal of attention, starting with the FI method. One measure exemplifies the FI's dominance very clearly: in the most substantial survey of dialectological projects until the early 1950s, which was arguably the heyday of dialectology, Sever Pop (1950: 1133–1175) places emphasis on the FI method by devoting 40 pages to its principles, while the WQ method is dealt with on merely two pages.

1.2.2 Wenker's Written Questionnaire (WWQ) method

Georg Wenker's method (WWQ) is usually presented as the prototypical WQ study. Wenker was a pioneer of the structured elicitation of dialect features and the first to gather linguistic data for linguistic theory building. One should stress, however, that Wenker's method, which is discussed in detail in Chapter 2, is *not* the method used in recent WQ studies. Wenker's method consisted of eliciting information from German schoolmasters, who, for instance, were rarely from the region at hand themselves and who were not phonetically trained or instructed to translate sample sentences from standard German into the local dialect. In other words, Wenker requested a type of community reporting (see Section 7.3.5) of a region's typical linguistic behaviour. WWQ is therefore very different from most present-day WQ studies, in which speakers themselves fill out the questionnaire without an intermediary. Wenker's survey eventually covered the entire German Reich, for which data from some 50,000 locations was collected. The *Deutscher Sprachatlas* was thereby confronted with a massive amount of data, much more so than in Gilliéron's case, where data from about 700 speakers was collected. The survey grid in Germany was therefore much tighter than the one in France. For each data point in France, more than 60 data points would be available in Germany. This high density in locations is one of the undisputed assets of the WQ method.

In a way, the success of Gilliéron's method was perhaps contingent on his being able to produce a complete national atlas in less than 15 years, still a feat by today's standards, while Wenker's method was mired in logistical problems from the beginning. These included problems in data processing characteristic of the pre-computer era and problems in cartographic representation. Technological innovations could only slightly alleviate these problems until very recently: for example, Wenker's successor in

the 1950s, Walther Mitzka, praised microfilm techniques as a godsend, while we know today of their limited practicability. Only the advent of the internet allowed Wenker's data to become fully available in DiWA, the *Digital Wenker Atlas*, which finally presents the data more than a century and a quarter after the start of the project to the interested public (see <<http://www.diwa.info/titel.aspx>>. Gilliéron must be credited in addition to his linguistic skill for his clear vision to see a large project through from start to finish. Had he not been so efficient, WQs might have had a different status in dialect geography.

1.3 Today's Written Questionnaire (WQ) methods

WQs have generally not figured prominently in dialectology, though at some point in time they played their part. In the early 1950s, McDavid (1953b: 568) characterized the use of a lexical WQ for the *Linguistic Atlas of Scotland* in opposition to Wenker's survey as "a new attempt to obtain by correspondence the materials for a linguistic atlas" and considered the method as suitable to giving "accurate information about the distribution of linguistic forms" (ibid: 570). Four decades later, the WQ had all but lost its momentum. As Chambers and Trudgill (1998: 16) put it so succinctly: the WQ "is no longer the primary method of data-gathering". Flash-forward 15 years and one can witness an increase in the use of and renewed interest in WQs in social dialectology and variation studies. Buchstaller et al. (2013: 97), for instance, believe that "questionnaire based approaches can be suitable for studying both morphosyntax and phonology" and thus take an approach towards WQs that was quite unthinkable only a decade or so ago. It seems that WQs have finally come to be seen for what they are: a highly interesting method that is often set aside too quickly.

As we have seen above, WQs clearly do not produce observation data. Rather, they elicit linguistic information *about* behaviour. They are tools for different kinds of reporting – either SELF-REPORTING – on one's own use, attitudes or perceptions – or COMMUNITY REPORTING – reporting on language use in a community. The basic WQ approach as such is of course not new. What is new, though, is that scholars are starting to exploit the WQ for its strengths in unprecedented ways. WQs, as a rather intuitive method, have been developed independently in a number of locations and contexts. At least partly as a consequence of William Labov's sociolinguistic revolution, WQs lost much of their attraction. They have at times been sidelined for the wrong reasons and have only recently been re-gaining some form of acceptance, yet to varying degrees (e.g. Schlee 2013; Buchstaller et al. 2013; Boberg 2013), outside of their well-established base in the speaker-evaluation tradition (e.g. Giles & Billings 2004 for an overview of that tradition since the 1960s).

It will be argued in the present book that WQs have a lot to offer for both the quantitative and qualitative study of the correlations between linguistic and social phenomena. The present focus will be primarily on a quantitative angle, as this seems to be the area that is most in need of attention. It will be shown that WQs should be considered as a viable method alongside other choices and should be part of the standard methodological toolkit. WQs are defined as questionnaire-based elicitation tools that are filled out by literate and semi-literate respondents without assistance. WQs are used in a number of linguistic disciplines, from applied linguistics and language pedagogy (see, e.g. Brown 2001; Dörnyei 2003) to speech act theory (e.g. Beebe & Cummings 1996) and the study of language use (e.g. Fuller 2005; Pi 2000). Their range is considerable and question types vary widely. The following overview focuses on the question types that are predominantly used in dialectology, dialect geography and sociolinguistics. Schleef (2013) presents five types of questionnaires used in the latter field. Building on these, I suggest a more general, three-tiered WQ question typology:

1. **Questions concerning regional language variation and social language variation:** from the use of linguistic varieties in given locales and settings (e.g. Extra & Yagmur 2004) to regional and social variation in language (as discussed in Chapter 4) to social variation of particular linguistics items (e.g. Fuller 2005; Lillian 1995 on the use of “Ms”)
2. **Questions concerning language perception & language attitudes** (e.g. Preston & Long 1999–2002; Watson & Clark 2014; and e.g. Lambert et al. 1960; Lambert 1967; Bourhis, Giles & Howard 1981; or Jenkins 2007)
3. **Questions using acceptability judgements of grammaticality:** originally a mainstay in generative linguistics on a binary scale, WQs have come to be used on gradient scales outside of the generative domain since Bard et al.’s (1996) Magnitude Estimation Method (e.g. Sorace & Keller 2005; Hoffmann 2006).

This typology by subject area is intended to facilitate the classification of the different approaches that are directly relevant to social dialectology and use of WQs. Questions should also be classified by type of reporting, distinguishing between SELF-REPORTING or COMMUNITY REPORTING, as well as by the type of information sought, with assessments of LINGUISTIC BEHAVIOUR on the one hand or reporting of LANGUAGE ATTITUDES AND PERCEPTIONS on the other hand. We will refer to studies in these areas throughout this book and the suggested typology will aid with their classification.

1.4 The organization of this Book

The content of this book is organized into two parts: a historical-theoretical part (Part I) and a practical part (Part II). Part I (Chapters 2–6) is comprised of a historically well-grounded, theoretical overview of the development of WQ methodology. It takes recourse to some predecessors, characterizes typical applications and results of traditional WQ variables, addresses recent adaptations of WQ methodology in the context of global Englishes and migratory studies and, last but not least important, probes into the reliability of WQs when compared to FI, corpus linguistic and sociolinguistic interview data. Part II (Chapters 7–9) was written as a practical aid or ‘handbook’. It aims to illustrate, in as detailed a way as possible, how written questionnaires can be devised, administered and analysed. The overall goal of this book is to offer an introduction to the WQ method to anyone interested in social dialectology and variation studies, from the (upper-level) undergraduate student to the language scholar wishing to explore another method. A brief synopsis of each part and chapter of the book is offered below.

The theoretical part begins with a history of the use of WQs in social dialectology and related fields in Chapter 2. The chapter’s overarching goal is to identify key moments in the development of the discipline. WQs were used frequently in the late 19th century in both Europe and, with some minor delay, in the US, and it was only thereafter that FIs replaced them as the primary data gathering tool. From the early 1970s onwards, the sociolinguistic interview more and more replaced the traditional FI method. It will be argued that the 1940s and 50s saw an interesting renaissance of interest in the WQ method in the US. While used in some projects, WQs failed to regain a status as a fully accepted method in English linguistics. However, WQs have been used continuously in non-English linguistics (e.g. Dutch) and the reasons for their lack of acceptance in English will be explored in that chapter. Anglophone Canada is the exception to the rule, as WQs have been in continuous use there since the late 1940s and have provided some of the major findings. It is not just for this reason that examples from Canadian English figure prominently in this book, but also because a focus on one variety offers avenues for theory-building that are otherwise difficult to establish.

After this fairly detailed historical sketch, Chapter 3 will probe into the reliability of WQ data. The major concern with WQs is that they do not provide observation data, which is why this chapter begins with a principled comparison of WQ data with corpus linguistic data. Following this, both FI and WQ data are stacked up with sociolinguistic interview data. It will be shown that in the comparison of FI and WQ, WQ data is no different from FI data. In the comparison of WQ and sociolinguistic data, some problems will be identified and earmarked for further exploration. Overall, however, and varying with the linguistic level and the precise linguistic variable and variable contexts, it will be found that WQ data delivers results that are largely equivalent and, generally, highly useful.

An examination of traditional WQ variables, defined as variables used by the mid-20th century, is offered in Chapter 4. This is intended as a kind of “smallest common denominator” and established practice in question design and WQ data analysis. The chapter focuses almost exclusively on variables in Canadian English, as it will set the empirical backdrop for an explication of a number of more general theoretical concepts in Chapters 5 and 6. The elicitation and basic analysis of lexical, morphological, syntactic and usage variables will be illustrated, before more recent approaches are addressed. Chapter 5 explores the application of WQs beyond a traditional scope in the contexts of World Englishes and Global Englishes, where special consideration will be afforded to the study of English as a *Lingua Franca*, i.e. communication among non-native speakers of English. In this area the method shows special potential to help address data gaps in the description of super-regional and global varieties of English.

Key concepts in sociolinguistic theory and historical linguistic theory are the focus of attention in Chapter 6. The idea is to introduce the beginning and intermediate student of linguistics to theoretical concepts and findings that will aid in the work with WQ data, which is the focus of Part II. A number of theoretical approaches from both synchronic and diachronic perspectives will be offered and illustrated using Canadian English. These include staple concepts such as real time and apparent time, the s-curve of linguistic change, change from above and change from below and some concepts involving gender. Among the newer approaches are linguistic border effects, which are of considerable relevance in Canada, sociohistorical frameworks of dialect development and new-dialect formation theory, the indexing of social meaning and thoughts on homogenizing and heterogenizing forces in today’s dialects. This concludes the theoretical part.

The practical part is intended to guide the novice in empirical methods in the design of a WQ study, from the conception of an idea (or shortly after that), to the statistical modelling of the data in the open source software suite R. This part was written with the advanced undergraduate student of the Arts in mind, who is generally familiar with qualitative methods of language study but not necessarily, or not at all, with quantitative methods. Part II begins with Chapter 7, which explores questionnaire design from a number of perspectives: deciding on which variables to focus on, finding a question style that works, determining questionnaire length and protocols for data collection, or determining which questions and variables can be polled with WQs and which cannot or not easily so. Chapter 7 also offers a typology of WQ questions in the context of social dialectology, from self-reporting and community reporting of linguistic behaviour to the assessment of attitudes and perceptions. In the context of a question typology, more recent WQ question types, those that have come into use in the 2000s, are discussed in this section.

Chapter 8 introduces the reader to practical work with WQ raw data. By using the online database of the *Dialect Topography of Canada* Project (Chambers 1994), a freely accessible and quite substantial data collection in what may be considered the standard framework, the reader will be shown in a step-by-step process how to download and manipulate the data. The only software tools needed throughout this book are a version of Excel 2007 or higher (or another spreadsheet software, though some commands may not function in the same way) and the freeware statistics and graphics utility suite R. The chapter assumes no prior Excel knowledge and is built around a step-by-step tutorial, with screenshots every step of the way. It limits itself to only a handful of Excel commands that will enable students to work with large data sets.

Until this point in the book, with the exception of Chapter 3 for evidential purposes, no statistical tests will be applied, which affords full focus on the WQ data. Chapter 9, finally, looks at the statistical testing of and hypothesis modelling with WQ data. This chapter is an introduction to linguistic computing with R, a suite that is in widespread use in statistics departments and increasingly used in linguistics. Limited to the set of variables found in traditional WQs, this chapter introduces a fully illustrated, step-by-step approach to four procedures that will assist greatly in the detection of patterns and in the identification of significant factors and predictors that (co-)determine the linguistic variables in question.

An Epilogue is offered in Chapter 10, which summarizes the most important points on WQs and aims to gauge their potential in social dialectology. In addition to highlighting some immediate desiderata, the attempt of an overall assessment of WQs, with special reference to methodological trends and their perceptions since the mid-20th century, completes this book. It is my hope that this treatment may highlight the great versatility of WQs, which is a feature that has usually been overlooked for some of their perceived shortcomings, while at the same time emphasising their drawbacks in an adequate, balanced and nuanced way. WQs are time-efficient, cost-effective, easy-to-administer tools that show, for a large number of variables, high reliability and validity. A review of the vexing history of WQs in the study of language in geographical and social space, which is a history of changing fortunes, shall be the start of the exploration.