The Handbook of Dialectology
This outstanding multi-volume series covers all the major subdisciplines within linguistics today to offer a comprehensive survey of linguistics as a whole. To see the full list of titles available in the series please visit www.wiley.com/go/linguistics-handbooks

The Handbook of Child Language
Edited by Paul Fletcher & Brian MacWhinney

The Handbook of Phonological Theory, Second Edition
Edited by John A. Goldsmith, Jason Riggle, & Alan C. L. Yu

The Handbook of Contemporary Semantic Theory
Edited by Shalom Lappin

The Handbook of Sociolinguistics
Edited by Florian Coulmas

The Handbook of Phonetic Sciences, Second Edition
Edited by William J. Hardcastle & John Laver

The Handbook of Morphology
Edited by Andrew Spencer & Arnold Zwicky

The Handbook of Japanese Linguistics
Edited by Natsuko Tsujimura

The Handbook of Linguistics
Edited by Mark Aronoff & Janie Rees-Miller

The Handbook of Contemporary Syntactic Theory
Edited by Mark Baltin & Chris Collins

The Handbook of Discourse Analysis
Edited by Deborah Schiffrin, Deborah Tannen, & Heidi E. Hamilton

The Handbook of Language Variation and Change, Second Edition
Edited by J. K. Chambers & Natalie Schilling

The Handbook of Historical Linguistics
Edited by Brian D. Joseph & Richard D. Janda

The Handbook of Language, Gender, and Sexuality, Second Edition
Edited by Susan Ehrlich, Miriam Meyerhoff, & Janet Holmes

The Handbook of Second Language Acquisition
Edited by Catherine J. Doughty & Michael H. Long

The Handbook of Bilingualism and Multilingualism, Second Edition
Edited by Tej K. Bhatia & William C. Ritchie

The Handbook of Pragmatics
Edited by Laurence R. Horn & Gregory Ward

The Handbook of Applied Linguistics
Edited by Alan Davies & Catherine Elder

The Handbook of Speech Perception
Edited by David B. Pisoni & Robert E. Remez

The Handbook of the History of English
Edited by Ans van Kemenade & Bettelou Los

The Handbook of English Linguistics
Edited by Bas Aarts & April McMahon

The Handbook of World Englishes
Edited by Braj B. Kachru; Yamuna Kachru, & Cecil L. Nelson

The Handbook of Educational Linguistics
Edited by Bernard Spolsky & Francis M. Hult

The Handbook of Clinical Linguistics
Edited by Martin J. Ball, Michael R. Perkins, Nicole Müller, & Sara Howard

The Handbook of Pidgin and Creole Studies
Edited by Silvia Kouwenberg & John Victor Singler

The Handbook of Language Teaching
Edited by Michael H. Long & Catherine J. Doughty

The Handbook of Language Contact
Edited by Raymond Hickey

The Handbook of Language and Speech Disorders
Edited by Jack S. Damico, Nicole Müller, & Martin J. Ball

The Handbook of Computational Linguistics and Natural Language Processing
Edited by Alexander Clark, Chris Fox, & Shalom Lappin

The Handbook of Language and Globalization
Edited by Nikolas Coupland

The Handbook of Hispanic Sociolinguistics
Edited by Manuel Díaz-Campos

The Handbook of Language Socialization
Edited by Alessandro Duranti, Elinor Ochs, & Bambi B. Schieffelin

The Handbook of Intercultural Discourse and Communication
Edited by Christina Britt Paulston, Scott F. Kiesling, & Elizabeth S. Rangel

The Handbook of Historical Sociolinguistics
Edited by Juan Manuel Hernández-Camayo & Juan Camilo Conde-Silvestre

The Handbook of Asian Linguistics
Edited by José Ignacio Hualde, Antxon Olarrea, & Erin O’Rourke

The Handbook of Asian Languages
Edited by T. A. James Huang, Y.-H. Audrey Li, & Andrew Simpson

The Handbook of Spanish Second Language Acquisition
Edited by Brian MacWhinney & William O’Grady

The Handbook of Korean Linguistics
Edited by Lucien Brown & Jae Ho Yoon

The Handbook of Korean Languages
Edited by C.-T. James Huang, Y.-H. Audrey Li, & Andrew Simpson

The Handbook of European Languages
Edited by Brian Paltridge & Sue Starfield

The Handbook of Spanish Second Language Acquisition
Edited by Brian MacWhinney & William O’Grady

The Handbook of Korean Linguistics
Edited by Lucien Brown & Jae Ho Yoon

The Handbook of Speech Production
Edited by Marnie Reed & John M. Levis

The Handbook of Contemporary Semantic Theory, Second Edition
Edited by Shalom Lappin & Chris Fox

The Handbook of Classroom Discourse and Interaction
Edited by Numa Markee

The Handbook of Narrative Analysis
Edited by Anna De Fina & Alexandra Georgakopoulou

The Handbook of English Pronunciation
Edited by Marina Reed & John M. Levis

The Handbook of Discourse Analysis, Second Edition
Edited by Deborah Tannen, Heidi E. Hamilton, & Deborah Schiffrin

The Handbook of Bilingual and Multilingual Education
Edited by Wayne E. Wright, Sovicheth Boun, & Ofelia García

The Handbook of Portuguese Linguistics
Edited by W. Leo Wetzel, João Costa, & Sergio Menuzzi

The Handbook of Portuguese Languages
Edited by W. Leo Wetzel, João Costa, & Sergio Menuzzi

The Handbook of Italian Linguistics
Edited by Mark Aronoff & Janie Rees-Miller

The Handbook of Translation and Cognition
Edited by John W. Schwieter & Aline Ferreira

The Handbook of Dialectology
Edited by Charles Boberg, John Nerbonne, & Dominic Watt
# Contents

List of Contributors viii

Introduction 1

**CHARLES BOBERG, JOHN NERBONNE, AND DOMINIC WATT**

**Section 1: Theory (section editor: Dominic Watt) 17**

Section Introduction

**DOMINIC WATT**

1 Dialectology, Philology, and Historical Linguistics 23

**RAYMOND HICKEY**

2 The Dialect Dictionary 39

**JACQUES VAN KEYMEULEN**

3 Linguistic Atlases 57

**WILLIAM A. KRETZSCHMAR, JR.**

4 Structural Dialectology 73

**MATTHEW J. GORDON**

5 Dialectology and Formal Linguistic Theory: The Blind Man and the Lame 88

**FRANS HINSKENS**

6 Sociodialectology 106

**TORE KRISTIANSEN**

7 Dialectometry 123

**HANS GOEBL**

8 Dialect Contact and New Dialect Formation 143

**DAVID BRITAIN**

9 Dialect Change in Europe—Leveling and Convergence 159

**PETER AUER**

10 Perceptual Dialectology 177

**DENNIS R. PRESTON**

11 Dialect Intelligibility 204

**CHARLOTTE GOOSKENS**

12 Applied Dialectology: Dialect Coaching, Dialect Reduction, and Forensic Phonetics 219

**DOMINIC WATT**
Section 2: Methods (section editor: John Nerbonne) 233

Section Introduction

JOHN NERBONNE

13 Dialect Sampling Methods
   RONALD MACAULAY 241

14 The Dialect Questionnaire
   CARMEN LLAMAS 253

15 Written Dialect Surveys
   J.K. CHAMBERS 268

16 Field Interviews in Dialectology
   GUY BAILEY 284

17 Corpus-Based Approaches to Dialect Study
   BENEDIKT SZMRECZANYI AND LISELOTTE ANDERWALD 300

18 Acoustic Phonetic Dialectology
   ERIK R. THOMAS 314

19 Computational Dialectology
   WILBERT HEERINGA AND JELENA PROKIĆ 330

20 Dialect Maps
   STEFAN RABANUS 348

21 Identifying Regional Dialects in On-Line Social Media
   JACOB EISENSTEIN 368

22 Logistic Regression Analysis of Linguistic Data
   JOHN C. PAOLILLO 384

23 Statistics for Aggregate Variationist Analyses
   JOHN NERBONNE AND MARTIJN WIELING 400

24 Spatial Statistics for Dialectology
   JACK GRIEVE 415

Section 3: Data (section editor: Charles Boberg) 435

Section Introduction

CHARLES BOBERG

25 Dialects of British and Southern Hemisphere English
   KEVIN WATSON 439

26 Dialects of North American English
   CHARLES BOBERG 450

27 Dialects of German, Dutch, and the Scandinavian Languages
   SEBASTIAN KÜRSCHENER 462

28 Dialects of French
   DAMIEN HALL 474

29 Dialects of Italy
   TULLIO TELMON 486
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Author(s)</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>Dialects of Spanish and Portuguese</td>
<td>John M. Lipski</td>
<td>498</td>
</tr>
<tr>
<td>31</td>
<td>Dialects of the Slavic Languages</td>
<td>Vladimir Zhobov and Ronelle Alexander</td>
<td>510</td>
</tr>
<tr>
<td>32</td>
<td>Dialects of Arabic</td>
<td>Enam Al-Wer and Rudolf de Jong</td>
<td>523</td>
</tr>
<tr>
<td>33</td>
<td>Dialects in the Indo-Aryan Landscape</td>
<td>Ashwini Deo</td>
<td>535</td>
</tr>
<tr>
<td>34</td>
<td>Dialects of Chinese</td>
<td>Chaoju Tang</td>
<td>547</td>
</tr>
<tr>
<td>35</td>
<td>Dialects of Japanese</td>
<td>Takuichiro Onishi</td>
<td>559</td>
</tr>
<tr>
<td>36</td>
<td>Dialects of Malay/Indonesian</td>
<td>Alexander Adelaar</td>
<td>571</td>
</tr>
<tr>
<td></td>
<td>Index</td>
<td></td>
<td>582</td>
</tr>
</tbody>
</table>
List of Contributors

Alexander Adelaar
The University of Melbourne, Australia

Ronelle Alexander
University of California, Berkeley, CA, USA

Enam Al-Wer
University of Essex, Colchester, UK

Lieselotte Anderwald
University of Kiel, Germany

Peter Auer
Department of German Linguistics, University of Freiburg, Germany

Guy Bailey
University of Texas Rio Grande Valley, Texas, USA

Charles Boberg
McGill University, Montreal, Canada

David Britain
Department of English, University of Bern, Switzerland

J.K. Chambers
University of Toronto, Canada

Ashwini Deo
The Ohio State University, USA

Jacob Eisenstein
Georgia Institute of Technology, USA

Hans Goebel
Department of Romance Studies, University of Salzburg, Austria

Charlotte Gooskens
Center for Language and Cognition, University of Groningen, Netherlands

Matthew J. Gordon
Department of English, University of Missouri, Columbia, Missouri, USA

Jack Grieve
Aston University, UK

Damien Hall
Newcastle University, Newcastle, UK

Wilbert Heeringa
Fryske Akademy, Netherlands

Raymond Hickey
Institute for Anglophone Studies, University of Duisburg and Essen, Germany

Frans Hinskens
Meertens Instituut (KNAW) & Vrije Universiteit Amsterdam, Netherlands

Rudolf de Jong
University of Leiden, Netherlands-Flemish Institute in Cairo

Jacques Van Keymeulen
Department of Linguistics – Dutch, Ghent University, Belgium
List of Contributors

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

Tore Kristiansen
Nordic Research Institute, University of Copenhagen, Denmark

Sebastian Kürschner
Catholic University of Eichstätt-Ingolstadt, Germany

John M. Lipski
Pennsylvania State University, State College, PA, USA

Carmen Llamas
University of York, UK

Ronald Macaulay
Pitzer College, USA

John Nerbonne
University of Groningen, Netherlands & University of Freiburg, Germany

Takuichiro Onishi
National Institute for Japanese Language and Linguistics, Tokyo, Japan

John C. Paolillo
Indiana University Bloomington, USA

Dennis R. Preston
Department of English, Oklahoma State University, Stillwater, Oklahoma, USA

Jelena Prokić
University of Marburg, Germany

Stefan Rabanus
University of Verona, Italy

Benedikt Szmrecsányi
University of Leuven, Belgium

Chaoju Tang
University of Electronic Science and Technology of China, Chengdu, Sichuan, P.R.China

Tullio Telmon
Università degli Studi di Torino, Torino, Italy

Erik R. Thomas
North Carolina State University, USA

Kevin Watson
University of Canterbury, Christchurch, New Zealand

Dominic Watt
Department of Language and Linguistic Science, University of York, UK

Martijn Wieling
University of Groningen, Netherlands

Vladimir Zhobov
Sofia University, Sofia, Bulgaria
Introduction

CHARLES BOBERG, JOHN NERBONNE, AND DOMINIC WATT

Dialectology is the study of dialect, or regional variation in language, a subfield of linguistics. This handbook presents a comprehensive survey of that subfield, including the theory of dialect variation; the methods of collecting, analyzing, and interpreting dialect data; and the facts of dialect variation in many of the world’s most widely spoken languages. Before proceeding with our survey, we offer by way of introduction the following reflections on some of the most basic issues in the field, as well as an explanation of the approach we have taken in planning this book and an outline of what is to follow.

1 The Origins of Dialect Variation and the Status of Dialectology

Dialect differences are caused by two forces operating in tandem: language change and the expansion of speech communities. Language change is of course a constant, on-going process in all speech communities: one of the axioms of historical linguistics is that all languages change all the time. As long as communities remain small, language changes are adopted or rejected by the community as a whole, or show only social differentiation. When a speech community expands sufficiently across a territory, however, the network of interpersonal communication that diffuses changes among its members is disrupted: sheer distance, or physical barriers like mountains and bodies of water—and sometimes also cultural, economic, or social divisions—make it impossible for change to diffuse evenly across the entire community. Eventually, an accumulation of undiffused or partially diffused changes causes community members in one region to recognize that people in other regions speak a different version of their language: what we would call a dialect.

Given enough time, this process of differentiation can cause dialects to diverge to the point where they are no longer wholly mutually intelligible, in which case we begin calling them separate but historically related languages. Such divergence lies at the heart of how historical linguists conceive of the development of families of related languages, like the Indo-European languages spoken across most of Europe and the Americas today, which hypothetically began their individual existence as dialects of a common ancestral language. In other cases, dialect differences can persist in a stable relationship for centuries, without leading to language divergence, or can decline and disappear, as the communication barriers that produced them are overcome by social or technological change. All normal languages, except those spoken in single, restricted locations, display regional variation and have
always done so: accounts of dialect differences are as old as written language itself, appearing two millennia ago in Ancient Greece and China. Given its universality, dialect variation should be seen as a fundamental aspect of human language and dialectology an important branch of linguistics, the scientific study of language. A linguistics that did not include dialectology would be incomplete.

Languages vary in many ways: across time and space, as just discussed, as well as across social categories. Today, dialectology is often seen as part of a larger sub-discipline of linguistics dealing with all of these types of variation, collectively called language variation and change (see, e.g., Chambers and Schilling 2013, another handbook in this series). This integrative approach reflects the many ways in which these types of variation have been shown to interact, first brought into clear focus in the work of William Labov (see below). Much of the variation we observe in speech communities is in fact the synchronic manifestation of diachronic processes, or changes in progress: newer forms, before being uniformly adopted, compete for dominance with older forms, in patterns that reflect an intersection of regional and social influences. Nevertheless, as difficult as it can be to isolate regional from other types of variation, the primary focus of this book will be on regional variation.

2 Defining Dialects

We shall begin our discussion of regional variation with just this problem, by exploring the meaning of the word dialect, which cannot be properly understood without reference to social variation as well. As linguistic variation arises in speech communities, it usually reflects social differences: different ways of speaking, like different ways of dressing or eating or having fun, come to be associated with groups arrayed on a socio-economic hierarchy involving wealth, power, education, ethnic or social identity, and other factors. Varieties of speech associated primarily with social groups are properly called sociolects rather than dialects and are the main focus of the allied subfield of sociolinguistics, but this type of variation also has an important place in dialectology, since regional varieties of a language—the definition of dialects given above—often develop social attributes. In particular, one variety, usually that spoken by the social, economic, and political élite in a nation’s capital city or other great metropolis, normally comes to be seen as the “correct” form of the language. In many cases, this evaluation is shared not only by its own speakers, who use it as a symbol and even a justification of their higher social position, but also by others in the community, who accept that their own speech is by comparison inferior, or “incorrect.” Because of its perceived social superiority, the élite variety is promoted to the status of a regional or national “standard” variety, which is preferred or even required in domains like broadcasting, education, government, journalism, the law, literature, liturgy, and science. It often serves these functions not only in its city or region of origin but across the entire linguistic territory, at higher social levels. This establishes a nationwide diglossia between the pan-regional “standard” variety, which comes to be seen not as just another dialect but as the unmarked form of the language itself (for instance, the form taught to foreigners who want to learn the language), and the regionally restricted and socially inferior “dialects,” which continue to be the language of everyday life for peasants or farmers in the countryside and for factory workers and trades people in the towns and cities. Rural and urban dialects often receive distinct social evaluations. Rural dialects are frequently seen as quaint and musical, if also unsophisticated and somewhat comic, and are associated with idyllic notions of traditional country life. Urban dialects are more often seen as lazy, ignorant, and linguistically and morally degenerate, since they are associated (at least in many middle-class minds) with the social problems of the lower-class sections of large cities.
An amusing instantiation of this ideology can be seen in the animated adaptation of Kenneth Grahame’s children’s story *The Wind in the Willows* that was made in the 1980s for Thames Television in the U.K. Though all of the characters are animated figures of animals, the heroes of the story, Rat, Mole, and Badger, speak with subtly different versions of standard British English, or “Received Pronunciation”; the sympathetic minor characters, like a plainspoken otter and a benign cow, have rural, West Country dialects, but the local gang of criminals, the weasels, are given working-class dialects from London (“Cockney”) and the urban industrial North. That said, the great fool of the piece, Mr. Toad, the lord of the local manor and a sort of upper-class twit, has the poshest accent of all, reminding us that the correspondence between high-class speech and positive social attributes is not always simple or direct (indeed, not only fools but cads and villains often have upper-class accents in popular entertainment). Nevertheless, the fact that this is a children’s program—and a delightful and brilliantly produced one at that, it should be admitted—emphasizes the extent to which dialect ideologies are inculcated in children at a young age by schools, media, and other institutions.

Even more problematic than negative attitudes about dialects is the transfer of such attitudes to the speakers themselves: people who speak what some think of as “lazy” or “ignorant” dialects are thought of as lazy or ignorant themselves, a stereotype that can be used to justify denying them educational, occupational, or social opportunities. Conversely, speakers of standard varieties may be given unfair advantages in the same contexts, a fact that has encouraged many ambitious people from working-class social backgrounds to try to “improve” their speech, often with measurable benefits. This, indeed, is the main justification for teaching standard varieties in schools, whose main purpose is to maximize the socioeconomic opportunities of their students. Defenders of the exalted status of standard varieties might argue that they are, in fact, democratizing (or at least meritocratizing) instruments, since they can be learned in school or by other means, thereby conferring socioeconomic benefits on the ambitious and becoming a symbol of individual achievement rather than of inherited privilege. Sociolinguists have argued passionately—and correctly—that these notions of superior and inferior dialects are based purely on social prejudice rather than linguistic fact, but they have proven very difficult to dislodge from popular culture, persisting at both ends of the social spectrum (for a critical look at the concepts of “standard” versus “dialect” in English, see Milroy and Milroy (1999) and the contributors to Bex and Watts (1999)).

Not all “dialects,” of course, are socially stigmatized, at least not by general consensus. Many non-standard dialects, if they lose points on the “status” dimension that governs access to the most prestigious schools and jobs, gain them on the “solidarity” dimension: their speakers are perceived as friendlier, more attractive, more relaxed, funnier, or more honest than speakers of the standard variety, if not more suitable as surgeons or bank presidents. Other non-standard dialects may be generally disparaged by people outside their own region or social group but are the focus of intense local pride within it. Speakers of these dialects often have a correspondingly negative view of the standard variety and its speakers: as Fischer (1958: 56) observed half a century ago in the pioneer of sociolinguistic studies, “A variant which one man uses because he wants to seem dignified another man would reject because he did not want to seem stiff.” Still other non-standard dialects are valued even by speakers of the standard variety as genuinely beautiful or cultured, even if inappropriate for some of the domains reserved to the standard variety.

Moreover, not all regional differences are socially marked. It is easy to think of variables in North American English, for example, that appear to be purely regional, with no common perception that one variant is more correct than the other. This is often true of lexical variation, which juxtaposes forms like *see-saw* and *teeter-totter*, both meaning a tilting board that children play on, or *cottage* and *cabin*, terms for a rural summer vacation home, or *pop* and
soda, generic terms for non-alcoholic carbonated beverages, without social prejudice. Many regional phonological variables, too, seem to lack social symbolism: Americans as a whole have no opinion on whether it is correct to pronounce pairs of words like cot and caught, or stock and stalk, differently, as in large sections of the eastern half of the country, or the same, as in most of the western half (where opinions exist, they relate to the phonetic qualities of the vowels involved, not the presence of phonemic contrast). Grammatical variation, by contrast, is more frequently aligned with social factors: everyone in the United States, as well as in other English-speaking countries, knows that “double negatives” and lack of “standard” subject-verb agreement are “wrong” and that those who use them mark themselves as lower-class, a message continually reinforced by schools and other institutions. Potential interactions with social factors, then, are an important aspect of dialect study.

If national standard varieties of languages coexist in a diglossic relationship with dialects of those languages, they also, in the case of multinational languages like English, French and Spanish, coexist with other national standards. In this context, such “standard varieties” are themselves “dialects.” In some cases, as between fellow ex-colonies like the United States and Canada or many Latin American countries, these relationships are fairly egalitarian, with national differences viewed as purely regional rather than social. In other cases, as between ex-colonies and their former colonizers, unequal sociolinguistic status can persist long after political independence. A general equality between the standard varieties of British and American English has now, after two centuries, come to be accepted by many English-speakers, including those in second-language education. Few people in the United States today would consider shifting toward British standards when reading the news on television or teaching English to foreigners. This equality, however, reflects the enormous size, power, and prestige of the United States, which has clearly surpassed that of the mother country. By contrast, relations between standard European French and the ex-colonial varieties of French spoken in Canada or other parts of the former French empire are still more hierarchical, with varieties closer to the European standard preferred in broadcasting and second-language teaching, for example. In some cases, opinion about such matters is regionally divided: while many in Spain consider castellano, the standard variety of Iberian Spanish based on the dialect of Castile, a global standard, Latin Americans are less likely to accept this notion and the form of Spanish taught in the United States most commonly follows a Mexican rather than Castilian standard, for instance in failing to preserve the Castilian distinction between s and z (casa, “house,” versus caza, “hunts”).

While dialects can differ at every level of structure—phonetic, phonological, morphological, lexical, syntactic, semantic, and so on—the term dialect is often used in a complementary relation with another term, accent, whereby dialect means differences in grammar and lexicon, while accent is restricted to phonological and especially phonetic differences, such as the quality of vowel sounds (as in the exhaustive survey of “accents of English” compiled by Wells 1982). This distinction takes on an important social dimension in Britain, for instance, where a three-level structure of language variation was traditionally observed: the national élite, particularly those educated at Oxford and Cambridge Universities, spoke “standard” British English with a “standard” or non-regional accent known as “Received Pronunciation,” regardless of where they lived (at least within England—the Celtic “nations” were to some extent exempt from this standard and had their own regional standards); the urban middle class spoke “standard English” with a regional “accent,” differing from the élite “standard” only in pronunciation, especially of vowel sounds; and the working class, urban and rural, spoke regional “dialect,” with non-standard grammar and lexicon, which also implied a marked regional “accent.” These social distinctions have recently been waning, with a decline in élite use of some traditionally prestigious features now seen as unattractively snobbish and a deliberate promotion of regional accents in domains like national broadcasting (e.g., on the BBC) where they were not previously accepted. Nonetheless, to a large
extent this differentiation can still be heard today and might also be argued to apply increasingly to the United States, where a non-regional “General American English” is pushing out local speech patterns in many regions (see Chapter 26 for examples).

A particularly problematic issue in defining dialect has been its taxonomic relation to the term language, the latter supposedly comprising a set of mutually intelligible dialects: if two people speak differently but can understand one another, they are speaking dialects of the same language; if they cannot understand one another, they are speaking different languages. It has often been pointed out that popular and even academic ideas about classifying varieties as languages or dialects reflect non-linguistic factors, like political boundaries and cultural history, as much as strictly linguistic criteria of mutual intelligibility. The stock examples in this discussion include, on the one hand, Mandarin and Cantonese, which many people think of as dialects of a single Chinese language but which are not mutually intelligible in speech (see Tang, this volume); and on the other hand, Hindi and Urdu, spoken in India and Pakistan respectively, which many people think of as separate languages but which are in fact largely mutually intelligible, separated more by an international boundary and by the cultural and religious affiliations of their speakers than by any marked linguistic divergence (see Deo, this volume). Europe, too, includes many instances of political boundaries creating and reinforcing “language” differences across what were once gradually shifting continua of local dialects, such as those between Germany and the Netherlands (see Kürschner, this volume), Spain and Portugal (see Lipski, this volume) or parts of the former Soviet Union (see Zhobov and Alexander, this volume); Italy presents a particularly complex blend of regional “dialects” and “languages” that are all offshoots of Latin (see Telmon, this volume).

Mutual intelligibility is itself a hazy concept, of course, involving not the binary distinction implied by the terms language and dialect but a cline or scale of linguistic similarity. At one end of the scale, we find cases of minimal regional difference with unrestricted mutual intelligibility, as between the major national standard varieties of English or Spanish: middle-class people in London and Los Angeles, or Madrid and Mexico City, recognize clear differences in each other’s speech but have very little difficulty understanding one another, if any at all. At the other end of the scale, we find complete unintelligibility, as between English and Arabic or Mandarin. In the middle, however, are many degrees of partial intelligibility. Some of these involve varieties that differ markedly from the most widely recognized international standards, such as the types of English spoken in Glasgow, Belfast, Appalachia, Jamaica, Singapore, or Nigeria. Others involve closely related “languages,” such as the Scandinavian or Romance languages, which began their histories as dialects of a common ancestral language and still retain a large common grammar and vocabulary, but have since drifted far enough apart to make mutual comprehension difficult, especially in speech. In many of these cases, moreover, the partial intelligibility that does exist is not symmetrical: Danes understand Swedes better than Swedes understand Danes and Portuguese speakers can generally make out more Spanish than vice versa. Intelligibility can be affected by non-linguistic factors like education, exposure, and the comparative social status and population sizes of the languages and cultures involved, as much as by purely linguistic matters like sound change or vocabulary replacement. These problematic issues will be reprised in several chapters of this book (see, i.a., Gooskens on dialect intelligibility).

3 The Origins and Development of Dialectology

From issues surrounding the nature and definition of dialects, let us now turn to a brief review of the history of dialectology, setting the stage for the chapters that follow. While it certainly has precedents in other places and earlier times, the modern “western” tradition of dialectology began in Europe in the nineteenth century. For many of its earliest practitioners,
Handbook of Dialectology

dialect study was a hobby: an entertaining pastime for self-taught philologists with an
interest in cultural history and folkways and a romantic conception of rural life, then very
much in fashion (seen also in the literature, music, and painting of the period). Some early
dialect collectors, for instance, were parish priests or schoolteachers, who had both a mea-
sure of formal education and a strong connection to the local communities they served.
Henry Higgins, the fictitious dialect phonetician parodied by George Bernard Shaw in his
1913 play Pygmalion, though based partly on real-life phonetician and philologist Henry
Sweet (1845–1912), comes off more as a gentleman of leisure with eccentric interests than as
the sort of person we would recognize today as a professional academic or serious scholar.
Early interest in dialect was given extra urgency by a genuine concern, not altogether unjus-
tified, that rural culture would soon be irretrievably altered or lost as industrialization and
urbanization increased. This “curatorial” approach to dialect study sought to record as much
of traditional rural speech as possible before it was too late, not unlike the efforts of modern
linguists to record and study the thousands of indigenous and minority languages whose
vitality is now threatened by digital technology and cultural globalization. Dialects also
came to be seen as entertaining by the growing urban bourgeoisie of the nineteenth century,
who enjoyed tittering at rustic stereotypes presented in theatres and music halls or in novels.
Many of the greatest writers of the time, like the Brontë sisters, Dickens, and Hardy, filled
their novels with passages of dialect, not only as a creative device, adding realism to their
rural or working-class characters, but as comic entertainment for their largely urban, middle-
class readers (a tradition that continues today in film and television).

Over time, however, the subject also developed a more serious, academic side. As will be
recounted in several subsequent chapters (see also the general accounts of the development
of dialectology in Petyt 1980, Francis 1983, or Chambers and Trudgill 1998), serious academic
study of dialects began in Germany, where Georg Wenker and his colleagues carried out a
postal survey of dialect variation across the German-speaking territory of Europe, starting in
the 1870s (Wenker et al. 1927–1956). Wenker asked school teachers to translate a set of 40
sentences into local dialect, as they observed it in their communities; he then collected these
records and compiled them in a Deutscher Sprachatlas, or “German language atlas,” showing
where each form was found and how one region differed from another. This effort was
closely followed by the Atlas linguistique de la France, published by Jules Gilliéron in several
volumes over the first decade of the twentieth century. The French study took a different
approach to data collection: it was based instead on face-to-face interviews with dialect
speakers carried out in the field, using a standard questionnaire administered by a trained
fieldworker (Gilliéron and Edmont 1902–1920).

The ultimate goal of these projects, like many that came after them, was to produce a
dialect atlas (see Kretzschmar, this volume): a collection of maps showing the regional distri-
bution of linguistic variants—sets of alternate words, pronunciations, or grammatical
forms—over a given territory (usually the territory covered by speakers of a single language,
or a subdivision of that territory). On these maps, symbols or transcriptions indicated the
variants occurring in each location and lines called isoglosses could be drawn to divide spatial
distributions of variants or mark the outer limit of a distribution; bundles of these isoglosses
were taken to indicate major dialect boundaries. This aspect of dialectology is also known by
the term dialect geography. As the name implies, dialect geographers used their maps to
develop geographic interpretations of the spatial distribution of dialect forms, such as the
role of barriers to communication, like mountain ranges, in preventing the diffusion of vari-
ants and thereby creating dialect divisions, or of channels of communication, like rivers and
roads, in encouraging diffusion over wider areas; they also turned to information on cultural
and settlement history in their efforts to explain the location of dialect boundaries.

Alongside dialect geography, an allied tradition of dialect lexicography also emerged in the
nineteenth century, which involved the production of dictionaries of dialect words and
Introduction

phrases, with definitions, examples, and usage notes, recorded in list form rather than on maps (see Van Keymeulen, this volume). At the turn of the twentieth century in England, for instance, Alexander Ellis published his records of dialect pronunciation, collected two decades earlier (Ellis 1890), and the English Dialect Society produced an English Dialect Dictionary, compiled by Joseph Wright (1898–1905).

As the field evolved, the interests of many dialectologists expanded beyond dialect variation itself to include connections with other aspects of language study. For instance, some dialectologists became involved in a debate with linguistic historians over the nature of language change. A group of nineteenth-century historical linguists known as the Neogrammarians had proposed that sound change—gradual shifts in the pronunciation of sounds found in sets of words—was a regular and exceptionless process that operated rather like the physical laws of natural science (Osthoff and Brugmann 1878). Systematic sound changes, gradually transforming all of the instances of a given sound simultaneously, were held to be responsible for the linguistic diversification of speech communities. Over thousands of years, this process had given rise to families of “genetically” related languages. These families could be modeled as a tree, whose trunk represented the original “proto-language” and whose branches represented the innovations that distinguished each sub-family and, ultimately, each individual language. The primary nineteenth-century example of this was the Indo-European language family and its Italic, Celtic, Germanic, Slavic, and other branches, which were readily observable in contemporary Europe, with a fascinating extension to Iran and northern India. For instance, the initial /p/ sound of the hypothetical ancestral Indo-European language, evident in Latin words like pater, pe(di)s, and piscis, had become an /f/ in the Germanic languages, producing German Vater, Fuß, and Fisch, or English father, foot, and fish, against French père, pied, and poisson or Italian padre, piede, and pesce, among dozens of other examples. Such regular similarities, called systematic correspondences, served as the basis for reconstructing the sound system and vocabulary of the now-extinct proto-language (the three correspondences just mentioned have been reconstructed respectively as *pH₂tér, *pōds/ped-, and *pisk-). Notwithstanding the apparently solid evidence of the Neogrammarians, other linguists saw their theory as an extreme view, which idealized the process of sound change and ignored a great deal of contradictory evidence. Some apparent exceptions to sound change patterns could be resolved by refinements to already formulated rules, but others were more tenacious, perhaps reflecting factors like borrowing, dialect mixture, or social pressures, which were seen as external to the mechanism of sound change but nevertheless posed problems for the theory.

As dialectologists began their survey work, the Neogrammarians initially hoped that the collection of data on traditional rural dialects, which were thought to be free of the complicating impurities of urban speech communities and standardized literary languages, would prove their theory right, by showing systematic and regular application of sound changes. When these data began to be analyzed, however, dialectologists found that they often revealed glaring exceptions to the hoped-for patterns of regular change. In some villages, a mixture of changed and unchanged forms was found, suggesting that some changes, at least, were irregular, affecting some instances of a sound but not others, and that the basic unit of phonological change was the word, not the phoneme, or sound. A classic illustration comes from Dutch: Kloke (1927) found that local forms of house and mouse, which both had long /u:/ in Proto-Germanic and should have followed parallel developments according to Neogrammian theory (as they did in English), displayed different sounds in some Dutch towns, a direct contradiction of the regularity of sound change (see Bloomfield (1933: 328–331) for an influential discussion of these data).

In response to the Neogrammian dictum that sound change is regular and suffers no exceptions, dialectologists therefore advanced their own, opposite slogan, that “every word has its own history,” apparently denying any sort of regularity to sound change. In its French
form, *chaque mot a son histoire*, this view is usually attributed to Gilliéron (see Gilliéron and Roques (1912)), but it goes further back to Hugo Schuchardt in the nineteenth century and perhaps as far as Grimm (1819: XIV), who says, “*…jedes Wort hat seine Geschichte und lebt sein eigenes Leben*” (“every word has its history and lives its own life”). The dispute over the regularity of sound change produced a deep cleft between what would become two separate traditions of linguistic thought. The dialectologists accused the Neogrammarians of ignoring the complexity of actual data in their efforts to attain higher levels of generalization and theoretical abstraction, while the Neogrammarians accused the dialectologists of obsessing over minutiae and variability for their own sake, like stamp collectors, without addressing questions of broader scientific significance. This rift is still observable today, in the division between formal theoretical linguistics, which is in some ways the heir of Neogrammarian philology (with other important influences, like the work of Saussure), and the field of language variation and change, including much of modern historical linguistics, sociolinguistics, and dialectology, which carries on the more skeptical or at least more empirical and data-oriented viewpoint of the nineteenth-century dialectologists.

While formal theoretical approaches to the study of language, such as the structuralism of the mid-twentieth century and the generative school of the late twentieth century, came to dominate modern academic linguistics, especially in eastern North America, a robust tradition of work on language variation and change, including dialectology, also continued to thrive, even if it was increasingly sidelined in many prestigious linguistics programs at major universities. By the 1930s, the French method of interviewing dialect speakers in the field and making meticulous records of their speech that could later be transformed into maps was extended to North America by Hans Kurath, who produced a *Linguistic Atlas of New England*, intended to be the first of several regional dialect atlas projects that would eventually cover the entire continent (Kurath et al. 1939–1943). This project was sadly never completed but has nevertheless produced a great deal of data and a tradition of work that continues today. In addition to the original New England atlas, the major published atlases of American English now cover the Middle and South Atlantic states (McDavid et al. 1980), the Gulf states (Pederson, McDaniel, and Adams 1986–1993) and the Upper Midwest (Allen 1973–1976). Following World War II, the *Survey of English Dialects* published maps of dialect variation across England (Orton and Dieth 1962–1971; see also Upton and Widdowson 1996), and in the 1960s a second major American dialectology project, the *Dictionary of American Regional English*, was undertaken, which is now complete (Cassidy and Hall 1985–2012; see also Carver 1987). Traditional dialect survey work has also continued in many other countries across the globe, as attested in the chapters of Section 3 of this volume.

Dialectology received a new stimulus in the 1960s from the work of William Labov in the closely allied field of sociolinguistics, which investigates relationships between linguistic variation and social structure and identity. One of the main concerns of early dialectology, as mentioned above, had been the effects of urbanization, mass education, and other forms of social change on traditional rural dialects, which were feared to be disappearing. A priority of many dialectologists was therefore to collect and study records of these dialects before they were lost. The best exemplars of traditional dialects were thought to be older rural men with minimal formal education and long family histories in the region, who were consequently favored as informants. Comparatively little interest was taken in other types of speakers, who were seen as less representative of “pure dialect,” or in cities, which were seen to offer nothing more than chaotic mixtures of modified regional dialects brought in by migrants from the surrounding countryside, or working-class urban varieties that were seen as linguistically and morally corrupt. By the 1960s, these assumptions no longer seemed justified. A new generation of sociolinguists sought to base their descriptions and theories of linguistic variation on the speech of the majority of the population. In the United States, Britain, and other western nations, this majority now lived in cities, where it
correlated not only old men of local stock but women, young people, recent migrants, and a wide range of social classes and ethnic groups, including those who spoke varieties stigmatized as debased, indolent, and ugly.

When Labov began to study urban speech communities, starting with New York City, he found that they displayed not the chaotic dialect mixture dismissed by some dialectologists as uninformative but *orderly heterogeneity*, a pattern in which the probability of occurrence of competing linguistic variants, such as standard and non-standard pronunciations or grammatical forms, depends on a complex yet systematic interplay of many different factors (Labov 1972). These factors included social attributes of speakers, like age, sex, and social class, as well as speech style, or the social context of speech (the identities of interlocutors, the setting and topic of conversation, etc.). Labov’s focus on correlations between linguistic and social variables and on shifting frequencies of variants has earned this type of sociolinguistics the names *correlational* or *quantitative* or *variationist sociolinguistics*. Some have called it *urban dialectology* (e.g., Chambers and Trudgill 1998: 54), though of course the sort of variation that Labov and others have studied in major cities can also be found in small towns and rural communities (as Labov himself did on the island of Martha’s Vineyard, off the Massachusetts coast), if on a smaller scale reflective of their narrower range of social diversity. Once the focus of interest shifted from a curatorial mission to preserve obsolescent traditional speech varieties to a more objective interest in how language reflects social identity, systematic variability could be found in any speech community; subsequent studies in British cities like Norwich (Trudgill 1974a) and Glasgow (Macaulay 1977) clearly demonstrated that this property of speech communities was not unique to New York City or to the United States. Moreover, communities can be compared simultaneously in the regional and social dimensions: they differ both one from another and within themselves, so that regional comparisons have to take local, community‐internal variability into account. For instance, regional divergence may be greater at certain social levels, or among particular ethnic groups. This hybrid approach has been called *social* or *socio-dialectology* (e.g., by Rona 1976; see also Kristiansen, this volume).

Labov’s contributions to modern dialectology go beyond shifting the focus to cities. He also pioneered the use of acoustic phonetic analysis to make detailed and reliable measures of vowel quality (Labov, Yaeger, and Steiner 1972; see also Thomas, this volume), which could be used to track the progress and distribution of sound changes that were continually modifying the pronunciation of urban dialects in contemporary American English. This work produced a new hybrid subfield normally called *socio-phonetics*; despite this label, the variation measured and analyzed by these techniques was as much regional as social.

Finally and perhaps most importantly, following on the insights of several predecessors (Weinreich 1954; Moulton 1960, 1962), Labov sought to re-establish connections between dialect study and theoretical linguistics, particularly structural phonemics (see Gordon, this volume). By framing his investigations of linguistic variation in terms of major questions of linguistic theory, such as resolving the Neogrammarian controversy discussed above (Labov 1981), or explaining what kinds of linguistic elements typically get transferred between dialects in contact situations and what kinds need to be learned by children from their parents (Labov 2007), Labov hoped to end what he saw as the intellectual isolation of dialectology, thereby augmenting its scientific value and stature. At the most fundamental level, he argued persuasively that the development of linguistic theory should not be divorced from the close study of data on how language is actually used by real people in real communicative contexts, and must give a satisfactory account of the variability found in these data, rather than dismissing it as optional rule application. He further believed that the study of dialect, in turn, could profit from new insights provided by reference to concepts and questions in general theoretical linguistics, as shown by Chambers (1973) and discussed here by Hinskens (this volume). Though not all dialectologists today feel the need to engage with questions of general linguistic theory, Labov’s work has illuminated
many opportunities for those who do; it has transformed the modern study of language variation and change, producing a whole new tradition of dialect study best represented by the *Atlas of North American English* (Labov, Ash, and Boberg 2006), the journal *Language Variation and Change*, and the annual *New Ways of Analyzing Variation* (NWAV) conference, now a major venue for the latest research in dialectology as well as sociolinguistics.

Modern dialectology has seen other advances as well. Where traditional dialectologists had to draw their maps by hand, a laborious and necessarily selective process, today’s practitioners benefit from a growing array of computer cartography tools, which support new insights into the regional distribution of dialect variants (see Rabanus, this volume; also Wikle 1997; Lameli, Kehrein, and Rabanus 2010). Another group of scholars has been concerned with developing objective measures of dialect difference that rise above anecdotal accounts and avoid selective analysis, a subfield known as *dialectometry*, which has incorporated sophisticated and rigorous quantitative methods from other social sciences and from statistics and computer engineering (Goebel and Nerbonne and Wieling, this volume; also Goebel 1982; Kretzschmar 1996; Boberg 2005; Nerbonne and Kleiweg 2007; Nerbonne and Kretzschmar 2006; Nerbonne 2009; Grieve, Speelman, and Geeraerts 2011).

Still other dialectologists have focused on the nature of the borders between dialect regions (Watt, Llamas, and Johnson 2010; Watt and Llamas 2014). These have often been observed to be “fuzzy,” involving *transition zones* in which the features of neighboring dialects are commingled, with a gradual shift from one dialect to the next, rather than sharp, with a sudden and easily perceptible change in speech at a specific location. Related to these topics are studies of the rise of dialect continua across stretches of terrain (Heeringa and Nerbonne 2001), of contact between dialects (Trudgill 1986), of the transitional forms that arise from this contact (Britain 1997; Chambers and Trudgill 1998), of the spatial diffusion of linguistic elements from one dialect to another (Trudgill 1974b; Callary 1975; Bailey et al. 1993; Auer, Hinskens, and Kerswill 2005) and of the rise of new dialects created by migration and dialect mixture (Kerswill and Williams 2005).

Another recent trend in dialect study has been to turn from production to perception, by examining what ordinary people think about the dialect diversity that surrounds them (Preston, this volume; Preston and Long 1999); how dialect differences interfere with cross-dialectal intelligibility (Labov and Ash 1997); or how listeners categorize speakers by dialect region and which features these categorizations rely on (Clopper and Pisoni 2004).

Finally, dialectology, like all fields of study, has not been immune to the recent influence of the internet, which presents new opportunities (and challenges) for data collection and analysis, dramatically increasing both the quantity of data available to researchers and the speed at which these data can be collected and analyzed (e.g., Eisenstein, O’Connor, Smith, and Xing 2010; Grieve, Asnaghi, and Ruette 2013). Technological advances have also contributed to the creation and analysis of large, searchable corpora of data, as discussed by Szmrecsanyi and Anderwald (this volume), allowing conclusions about variation and change to be drawn from ever larger sets of data. In particular, internet searches make possible the rapid collection of vast quantities of data on regional variation in ordinary language—most commonly written language but also speech—as opposed to language deliberately collected for the purposes of study. This is a potentially transformative change that minimizes the gap between dialectology and the variation it attempts to study.

### 4 The Present and Future State of Dialectology

Studies of diffusion, which includes the spread of features not only from one region to another but from one social group to another, give rise to the question of whether the advent of mass education, personal mobility, and instant communication in modern, industrialized
nations threatens the very survival of dialects, echoing the original concerns of nineteenth-century dialectologists (Britain 2009; Kristiansen 1998). Insofar as many traditional rural and non-standard urban dialects are now declining or disappearing, this implies a gradual contraction of the subject matter of dialectology, which might suggest a pessimistic view of the future of the field. On the other hand, many older dialects are sustained by a strong force of local identity that prevents their decline. Even in fully industrialized or post-industrial countries like Britain and Germany, some distinctive regional dialects, like those in the North of England (e.g., Tyneside, Yorkshire, or Lancashire English) or the South of Germany (e.g., Bavarian or Swabian German), continue to be spoken enthusiastically by millions of people and show little sign of disappearing anytime soon, even if their features are constantly modified by contact with standard and non-local speech. In other cases, like that of Denmark, which saw traditional dialects virtually disappear over the twentieth century (Kristensen and Thelander 1984; Pedersen 2003), local identity manifests itself today in subtler forms of variation, sustaining small regional differences in an otherwise homogeneous supra-regional or national type of speech that diffuses from cosmopolitan centers (Kristiansen 1998). Boberg (2005, 2008, 2010) observes a similarly fine-grained yet tenacious regional differentiation, which we might call micro-variation, in Canadian English, which is otherwise reputed to be remarkably homogenous over the country’s vast territory.

Moreover, if we look beyond traditional dialects, we find a proliferation of new dialects constantly emerging. This is true in many large urban centers, for example, whose populations are increasingly diverse: distinct ethnic and cultural subgroups in cities like London (Cheshire et al. 2011), Berlin (Wiese 2009), Stockholm (Kotsinas 1988) and Copenhagen (Quist 2000) mark their emerging social identities in linguistic as well as other ways, though intra-community variation of this kind is strictly speaking more a concern of sociolinguistics than of dialectology.

An even more important source of new dialects, however, is the most consequential linguistic development in the world today: the rise of English as a global language. This phenomenon, which has generated a whole new subfield of language variation and change called World Englishes, with its own conferences and journals, fundamentally involves the creation and development of dozens of new dialects of English (Crystal 2003; Schneider 2007; Kachru, Kachru, and Nelson 2009; Melchers and Shaw 2013). These were spoken at first in countries with historical ties to the former British Empire, where many native yet distinctive varieties of English are to be found, usually in multilingual settings (e.g., India, Ghana, Nigeria, Hong Kong, and Singapore; the “outer circle” of Kachru (1985)). Today, they are increasingly flourishing in countries with no such connection, where English has no historical status but is now used as a lingua franca for intercultural and international communication (Kachru’s “expanding circle”), especially in domains such as advertising, digital media, diplomacy, high technology, international commerce, international sport, popular entertainment, post-secondary education, scientific research, and of course tourism. In these contexts, new, non-native varieties of English, which may become native varieties in the future, exhibit distinctive features that reflect the native languages with which they coexist.

In Europe, for instance, where English has become the de facto working language of the European Union and is widely learned as a second language in primary school by children across the continent, the point at which future generations will regard Dutch English, Swedish English, even German English as legitimate dialects of English, spoken by bilingual populations, seems less remote every year (after all, Irish, Scottish, and Welsh English, now universally accepted as dialects of English, also began as second-language varieties, spoken by Britain’s Celtic populations). Assuming this scenario persists (which should not be taken for granted), a similar evolution of Russian English, Chinese English, and Japanese English dialects may not be far behind. Once these originally second-language varieties become semi-native, mutually intelligible regional types of English, they enter the legitimate domain of dialectology; from this perspective, the future of the field looks bright (if only for English dialectology!).
In short, the field of dialectology has grown and adapted in many ways and continues to respond to a changing environment today. While it no longer holds the central position in linguistic science that it enjoyed in the nineteenth century, it nevertheless remains a dynamic and relevant sub-discipline that continues to produce new scholarly work and attract new generations of students. From its origins in Europe it has now spread across the globe, with dialect studies available or in progress on languages spoken in every region of the world, as seen in Section 3 of this volume. Yet, despite all of these changes, most dialectologists today continue to focus on the central questions that gave rise to the field over a century ago:

1. How do languages vary across the territories in which they are spoken?
2. What are the common patterns in this variation, including the linguistic constraints that govern it, viewed across different languages?
3. How do settlement history, topography, social patterns, urbanization, and other non-linguistic factors explain the spatial distribution of linguistic features?
4. What is the nature of the transitions or boundaries between spatial distributions?
5. How do innovative features spread across new territory?
6. Is regional variation receding, stabilizing, or increasing over time?

Despite their long history, all of these questions remain relevant today, as they are addressed with the new methods described above and with new data, both from new communities and from previously studied communities that continue to change.

5 Rationale and Plan of This Book

Given the recent expansion and diversification of dialectological scholarship reviewed above, contemporary students of dialect at all levels of expertise now face a significant challenge. They must keep up with technical and theoretical advances in a wide range of different sub-disciplines, as well as with a constantly growing body of data on dialect variation in a wide range of languages. Moreover, as the demands of assimilating all of this new information grow heavier, it becomes more difficult—yet no less important—to maintain an intellectual connection between contemporary research and the scholarly achievements of the past. Foundational work should always be taught and re-taught as an underpinning for modern research, but also critically re-evaluated in light of new information and alternative, innovative thinking.

In light of these challenges, the field has become far too large for even the most senior and widely experienced scholar to have more than a passing acquaintance with all of its various sub-divisions, let alone for the junior scholar or beginning student who wishes to progress beyond the surveys available in short introductory monographs suitable for undergraduate courses. Yet those with an interest in developing a broad knowledge of dialectology, or in having access to such knowledge on an as-needed basis, have been faced with making their own surveys of a very large and in some cases inaccessible corpus of materials, which is simply not available in many libraries. Dialect atlases in particular are expensive and space-consuming luxuries found only in large or specialized collections, while much of the original work on non-English dialects, particularly that written before the late twentieth century, has not been translated out of its original languages and cannot be read by most English-speaking students even where it is available.

The present volume therefore seeks to provide both experienced practitioners and their apprentices with an overview of the field of dialectology—past and present—comprising three main aspects of the topic: principal theoretical approaches, methodological traditions, and sets of data. This trio of topics provides the main organizational basis for the book, reflected in the
three sections of the table of contents, each comprising 12 chapters. Because dialectology is and always has been fundamentally a data-driven field, committed to empirical investigation more than to theoretical speculation, or rather to basing the development of theory firmly on competently collected and analyzed sets of data, the methods of dialectology and the data of dialectology are just as important to any review of the field as the various aspects of dialectological theory. Accordingly, whereas the book’s first section gives a detailed account of the historical and contemporary development of dialectological thinking, including crucial concepts like the dialect dictionary, the dialect atlas, and the various interfaces with other areas of linguistics and non-linguistic sciences discussed above, the second section is concerned entirely with methodological matters—how dialect data are collected and analyzed—and the third with the data themselves, illustrated with descriptive overviews of dialect variation in the world’s most widely spoken languages and language families, particularly those that have produced the richest traditions of dialect study. Primary editorial responsibility for each of these sections was assigned to one of the three co-editors of the volume: Watt oversaw the section on theory, Nerbonne that on method, and Boberg that on data. Beyond this co-written general introduction, each section editor provides a specialized introduction to his section, which introduces and discusses the chapters it contains in more detail than is possible or appropriate here.

REFERENCES


Section 1 – Theory
Introduction

DOMINIC WATT

The first group of chapters in this book is devoted to matters of theory. We will, on the one hand, consider theories that have been elaborated within the field of dialectology itself. On the other, we will evaluate the contributions made to dialectology by theories from other domains of linguistics. Some of the latter set of theories have had a significant impact on the ways in which we conceptualize the notion of dialect, as well as how we describe dialects and classify them into superordinate groupings with respect to the languages of which dialects are said to be subspecies. We will see how dialectology and philology have contributed in their turn to theories of language, even in the case of theories whose exponents do not acknowledge, or may not even be aware of, that contribution. The following chapters will make it clear that linguistics as we recognize it today would probably be markedly different had so many paths in synchronic and diachronic language study not first been broken by scholars who devoted their careers to the study of dialect variation and relatedness, initially among European languages but later among languages from all the inhabited continents.

It has sometimes been claimed that dialectology is pre-theoretical, or atheoretical. Although the assertion that any sort of dialectology could be carried out in a theoretical vacuum could not be easier to counter, it is not hard to see why the belief persists. Until fairly recently, the professed mission of many dialectologists was simply to describe and catalog linguistic phenomena, rather than to try also to explain the genesis of these phenomena and the mechanisms by which they are transmitted from speaker to speaker across space and time. It was seen by many practitioners of dialectology to be more pressing in the immediate term to observe speech and language as they were actually being used out there in the world than it was to seek profound truths about the nature of language from more abstract philosophical perspectives. Translating this stance into action was to a significant degree motivated by a growing awareness and concern that the traditional dialects were vanishing, and that a failure to document them before they disappeared would mean irrevocably walling up a window onto the past. The prospect of imminently losing access to traditional forms of speech that could link us more directly to the world of our distant ancestors provided a powerful spur. It inspired scholars, individually or in teams, to go to huge lengths to set down, systematically and scrupulously, the rich detail of lexical, phonetic, phonological, morphological, and syntactic variation in the speech and language of people who were otherwise marginalized because of their lack of education, sophistication, or “breeding.” Dialectology gave, as never before, a voice to untutored rural dwellers, those seen as uncorrupted by modernity and urban life, and whose language was as untainted as it could be by the effects of the pressure to conform to institutionally imposed linguistic standards.
The erstwhile dialectological focus on the what, rather than the why and how, might be likened to large-scale data-gathering programs in disciplines such as zoology, botany, or astronomy. Cataloging the diversity of insects, flowering plants, galaxies, or exoplanets that have not yet been named or classified is a vital first step in understanding how the systems in which those entities operate are structured and how they function. It would be rash to assume that our theories of how those systems work are watertight until we have sampled the universe of variation as exhaustively as our finite resources will permit. Yet there are linguistic theorists who would argue that from the descriptive point of view, we essentially know all we need to know about variability in certain languages, and that dialect diversity in those languages is in any case largely irrelevant to the central enterprise of the field. It is perfectly valid, they would argue, to make pronouncements about the grammatical properties of an entire language by examining just one of its dialects—which is almost always the standard variety, if one exists—and not consulting any speakers at all. Why would one go to all the trouble of asking other people about their language, if one can use one’s own native-speaker intuitions as a source of data, or get the information one needs from published sources?

Dialectology, according to views of the above kind, has little to offer “serious linguistics.” The dialectologists’ preoccupation with regional and social variation in language production has been viewed as a harmless enough trait, but spending even part of a career studying this aspect of performance (which I have heard dismissed, in paraphrase of John McCarthy’s aphorism, as mere “froth on the surface of language”) has hitherto generally not been thought a worthy pursuit for the theoretician. The proper subject matter of linguistics, on this view, is the set of abstract principles that govern how sentences are constructed, or the constraints that determine how phonological units such as syllables or feet or tonemes may be strung together. The focus on lexis in many dialectological surveys bolstered a perception that dialectologists and philologists were not really concerned with phenomena in the phonological or grammatical domains, and given that in many cases, the questionnaires designed to elicit dialect lexis dwelt on terminology pertaining to occupations such as agriculture, animal husbandry, or traditional arts and crafts, it is easy to see why many researchers working in other subareas of linguistics formed the impression that the methods and underlying conceptual framework of dialectology had stalled decades earlier. Accusations of “golden ageism” were also frequently leveled at dialectologists, probably owing to a tendency in some quarters for ideologies concerning the supposed purity or uninterrupted lineages of traditional dialects to bias scholars in such a way that they downplayed the influence of relevant societal factors on the historical trajectories of languages and their dialects (e.g., marginalizing the evidence suggesting that geographical and social mobility were considerably more prevalent in former centuries than we often assume, such that we tend to believe today that mobility is a modern phenomenon; see e.g., Long 2013; Whyte 2000; also Milroy 1992).

The criticisms of dialectology discussed above are not wholly without foundation. But the field has moved on a great deal in recent decades. New theoretical insights and methodologies from cognate disciplines such as sociolinguistics, human geography, and social and evolutionary psychology (e.g., Buchstaller and Alvanides 2013; Cohen 2012) are helping contemporary dialectology to flourish, and the findings of studies in archaeology and population genetics (e.g., Heggarty et al. 2005; Winney et al. 2012), along with innovations in mapping and ‘Big Data’ analysis techniques (e.g., Huang et al. 2015; Grieve, this volume), are being fused with dialectological data in ways that mutually strengthen all of the contributing disciplines. It is of benefit to linguists of every stripe that new theories of language evolution, acquisition, structure, and use are crystalizing out of the interplay among these diverse approaches to human characteristics and behaviors.

We commence the Theory section of this volume by taking the long view of dialectology’s emergence from the earlier philological tradition, and how historical linguistics simultaneously grew as a parallel offshoot, with all three disciplines cross-fertilizing one another.
In his chapter, Raymond Hickey argues that the boundaries of dialectology are a matter of terminological convention rather than the result of the field having intrinsically sharp edges, and that the scope of dialectology is in any case subject to change as interconnections with other disciplines are forged. Similar perspectives are offered by Jacques Van Keymeulen in his discussion of the history and current status of the dialect dictionary (Ch. 2). The advent in recent years of searchable large-scale multimedia repositories of dialectal data has transformed the dialect dictionary from a static, cut-and-dried record of painstaking scholarship carried out over years or decades, to a dynamic “living” resource that can be regularly updated with new research findings and data gathered from the general public using innovative methods such as crowdsourcing apps (e.g., Goldman et al. 2014).

The potential for producing graphical depictions of data held in textual or numerical form by harnessing the power of mapping techniques, in particular contemporary cartographic software, is explored by Bill Kretzschmar, whose chapter (Ch. 3) deals with the evolution and future of the linguistic atlas. Theoretical claims in support of Kretzschmar’s contention that speech and language represent complex, non-linear, scale-free fractal systems become ever more testable in the era of prodigiously large datasets of the kind collected for linguistic atlas projects. Several of the themes Kretzschmar develops are reprised in the fourth chapter, on structural dialectology, by Matthew Gordon. In his exposition of the influences that the structuralist tradition in linguistics and dialectology have had upon one another, Gordon points to the design properties of the Atlas of North American English (ANAE; Labov, Ash, and Boberg 2006), arguing that ANAE instantiates many of the precepts of the American and European structuralist approaches to phonology. For example, ANAE organizes vowels into subclasses (long/short, ingliding/upgliding, and so forth), which are then used as a means of bundling North American regional dialects into supercategories on the basis of how vowel types are realized phonetically and how their qualities alter in response to homeostatic pressures of the sort that regulate chain shifting (see further Docherty and Watt 2001; Gordon 2013).

The impact upon dialectology made by the revolution in linguistic thinking brought about by the Chomskyan generativist paradigm that largely supplanted the American structuralist school in the 1950s (see relevant entries in Allan 2013; but cf. Salmons and Honeybone 2015) is examined in detail by Frans Hinskens (Ch. 5). Hinskens is at pains to point out that the relationship between dialectology and “formal linguistic theory” is by no means a one-way street; by analogy with the fable of the blind man and the lame, who pool their complementary faculties so as to overcome their individual disabilities, Hinskens contends that mainstream linguistics owes as much to dialectology as dialectology does to mainstream linguistics. Measures to ensure the continuing healthy symbiosis of dialectology with another of its sister fields—this time variationist sociolinguistics—are endorsed by Tore Kristiansen in his chapter on sociodialectology, this field of inquiry being the product of what happened, as he puts it, “when dialectology moved from the countryside into town.” In seeking to understand the mechanisms by which the embedding of innovative forms in linguistic structure takes place, Kristiansen believes we must integrate the explanatory frameworks developed within the Labovian sociolinguistic paradigm with those that have come down to us via other tributaries to the stream of research that has emerged from more than two centuries of scholarly interest in dialect variation.

One strand in this braid, to echo Kristiansen’s riverine metaphor, is the variety of dialectometry practiced by Hans Goebl (Ch. 7), for which he suggests the title “atlantometry” owing to the fact that the statistical modeling methods he employs are based exclusively on dialect atlas data. The geolinguistic focus of the early European dialectologists is hardly accidental, when one considers how energetically the new nation states went about surveying their territories and standardizing the ways in which the observations they made were to be quantified. Figuring out how to handle the multidimensional richness of dialect data would present
its own challenges, of course. From our contemporary perspective, it is only too easy to underestimate the hurdles that the early dialectometrists had to overcome in terms of bootstrapping into existence a set of metrics that would permit them to capture linguistic distance in a principled, systematic, and consistent fashion.

Even if dialects were stable and unchanging, the tasks faced by the atlantometrist or any other investigator of dialect would be daunting enough. But, as David Britain points out in his chapter on dialect contact and new dialect formation, new dialects arise all the time. The role of contact between speakers of different dialects in the emergence of new subvarieties of a language has been the subject of considerable debate in recent times, for example in terms of the extent to which identity factors might bias interactants toward or away from the adoption of certain linguistic behaviors (cf. Trudgill 2008; Hickey 2013). Although divergence between dialects is frequently observed and does not presuppose an absence of contact between the speakers of the varieties in question (e.g., Auer, Hinskens, and Kerswill 2005), it is also clear that human beings seem to be possessed, as Britain has it, of an overwhelming “urge to converge” linguistically. This predisposition is played out on a continent-sized stage and over more than a millennium of history in the European contexts explored by Peter Auer (Ch. 9), whose discussion focuses chiefly on continental Germanic languages, but applies in large part also to members of the other European language families. The grand themes of leveling and convergence are illustrated time and again in dialectological and sociolinguistic studies of language varieties across Europe, such that it becomes unviable to dispute Auer’s assertion that a march toward linguistic uniformity is the major defining characteristic of the European dialects over the course of the last century.

What is surprising is that in many parts of the world non-linguists cleave quite so strongly to their beliefs in the persistence of dialect forms that are either extinct or shortly to become so. Tapping into these beliefs and attitudes is the goal of perceptual dialectology (also known as “folk linguistics”), the field of which Dennis Preston (Ch. 10) has become the most prominent champion. Close scrutiny of what laypeople have to say about the language forms they and their fellow speakers use—or are believed to use—can tell us a good deal about how languages are changing or are liable to change in the future, insofar as the metalanguage of non-linguists often yields insights into their proclivity to adopt or reject incoming changes or to retain traditional ones. Laypeople may justify their views about dialects and accents in terms of how readily the speech patterns of other speakers and speaker groups can be understood, a habit which, as Charlotte Gooskens observes in Chapter 11, is no mere matter of casual academic interest. It may help language planners to decide whether a dialect is treated as a distinct language or not, with all the implications this decision might have for the rights and opportunities of minority populations. It is also true to say that stigmatizing certain language varieties for their (ostensibly) low intelligibility can have very tangible knock-on consequences for the variety’s speakers, who may suffer disadvantage as a result of negative attitudes toward their vernaculars held by those in positions of power in the educational or legal systems. In the final chapter in this section (Ch. 12), I deal further with perception issues, but this time in the context of strategies that speakers might apply if they wish to disguise their regular speech patterns. Leading listeners to believe that one’s feigned dialect or accent is authentic is generally a perfectly benign deception—we expect this ability to be well-developed in professional actors or individuals who have received elocution lessons, for instance—but where the aim is to gull listeners into thinking they are hearing a different person talking (e.g., so as to obtain money fraudently, or to impersonate a murder victim so as to mislead others into thinking the victim is still alive) the law may view the attempt as a criminal offence.

Essays on other applications of dialectological data could easily fill a separate volume, and it is encouraging to see that publishing projects of just this type are underway as part of the current drive to expand and enrich the already very sizeable body of literature on dialect variation. It is our hope that the chapters in this Handbook will quickly come to represent an invaluable and authoritative resource for all involved in these new dissemination initiatives.
REFERENCES


1 Dialectology, Philology, and Historical Linguistics

RAYMOND HICKEY

1.1 Introduction

The term “dialect” is understood today to refer to a geographically delimited form of language. The purpose of the present chapter is to trace the history of this meaning of the word and to outline the rise of dialectology, which is the historical study of dialects in this sense (Fisiak ed. 1988). Furthermore, this study seeks to set dialectology in relation to the disciplines of philology (Turner 2014; Momma 2015: 1–27) and historical linguistics. These latter two are closely related in that the former fed into the latter. Indeed, the modern discipline of linguistics arose at the end of the eighteenth and beginning of the nineteenth centuries out of earlier concerns of philology, which is the study of the textual records of languages.

The etymology of “dialect” can be traced back to Classical Greek, in which the word διάλεκτος originally referred to discourse, conversation, or way of speaking, and later came to mean a regional variety of a language. It is this last meaning that initiated the modern understanding of the word (the older meaning of “investigative discussion” can still be recognized in the term “dialectic”). However, one cannot say that once the meaning of “regional variety” was established one had a usage similar to that today. The essential difference is that nowadays “dialect” stands in a contrastive relationship to “standard,” a form of language favored in the public domain and employed in compiling official documents in a country. The reference to “country” is important here: the modern sense of “standard,” with all its prescriptive connotations, is essentially an artifact of modern nation states. Thus, the often negative connotations of dialect did not hold until the notion of a preferred form of language arose, a form that enjoyed preference in writing, education, and public speaking. How early this preference occurred historically is difficult to say with certainty. True, there were historical constellations of language varieties in which one was used more than others. This applied in the Hellenistic period of Greek (roughly three centuries before the beginning of the common era), when the dialect of Attica (including the city of Athens) was used widely as a koiné or common form of language in the eastern Mediterranean (Woodard 2008). In England, during the later Old English period, the language of the West Saxon region was employed in written documents (Gneuss 1972), such as religious or legal texts, and thus enjoyed a similar status to Attic Greek in ancient Greece. But in neither case did later attributes of standard forms of language apply, above all codification and prescriptivism, which involved the censure of dialect forms of the same language.
To trace the history of dialects and their study, that is, dialectology, one should distinguish three aspects of this complex: (i) awareness of dialects, (ii) attitudes toward dialects, and (iii) the description and study of dialects. These aspects stand in chronological order: first awareness arises, and mention of dialects is found in the textual record. Somewhat later, attitudes toward dialects seem to have developed. In the Western world these are invariably negative, with mention made of a preferred form of the language in question. Later still, one finds descriptions of dialects, usually of one particular language with which an author has a specific connection, either by birth or interest.

1.2 Dialect Awareness and Attitudes

An awareness of dialect differences in England goes back at least to the Middle Ages: Geoffrey Chaucer used Northern English (Hickey 2015) for the purpose of character portrayal in *The Reeve’s Tale* (Tolkien 1934; Wales 2006, 75). The north/south dichotomy is referred to by later authors on language, notably George Puttenham, who, in his *The Arte of English Poesie* (1589), states his preference for “our Southerne English,” which is the “usual speech at court and that of London and the shires lying about London within lx. myles and not much aboue” (Mugglestone 2007, 9). One of the earliest listings of dialect areas was made by Alexander Gil in his celebrated *Logonomia Anglica* (1619). On discussing the main features of the different dialectal areas, he mentions the northern lack of rounding in *beath* “both,” and the northern forms *sal* “shall,” *sud* “should,” *fula* “follow,” and *briks* “breeches.” There was also an awareness of the Englishes spoken in the Celtic regions: Shakespeare in the “Four Nations Scene” in *Henry V* portrays the speech of English, Welsh, Scottish, and Irish characters.

In France, the primary north/south dialect division was also characterized as early as 1284, by the poet Bernat d’Auriac. Here the forms of the keyword “yes” are essential, and have even resulted in the names of two large parts of France—Languedoc and Languedoeil—the former referring to the region south of the River Loire, the latter to that north of the Loire (which later developed into modern French).

There would seem to have been two attitudes to English dialects in the early modern period. One was neutral and the other decidedly in favor of southern speech. John Hart (d. 1574) spoke of “the flower of the English tongue,” referring to the language of the London court. The more neutral attitude is seen in dictionaries of the time, for example, William Bullokar (1616): “So every country hath commonly in divers parts thereof some difference of language, which is called the Dialect of that place,” a view echoed by Thomas Blount (1656): “Dialect is a manner of speech peculiar to some part of a Country or people, and differing from the manner used by other parts or people, yet all using the same Radical Language, for the main or substance of it.” But the great lexicographers of the eighteenth century—Johnson, Sheridan, Walker—showed no interest in regional variation, meaning that dialects were excluded from the emerging ideology of a standard in English.

1.3 The Description of Dialects

Early descriptions of dialects differ from later studies in that they largely consist of dialect words. The gathering of such words and their publication as lists has a long tradition in England and other European countries. The most significant early English work is John Ray’s *A Collection of English Words not Generally Used* (1674). Ray states that “in many places, especially of the North, the Language of the common people, is to a stranger very difficult to be understood” (Preface To the Reader), and was hence motivated to record northern words.
In some dictionaries, northern words that were not current in the south of England are given. For instance, John Palsgrave’s *Lesclarcissement de la Langue Francoyse* (1530) mentions words such as *sperre* “to shut” and and *that ylke day* “that same day” that are representative of “the northern language” (Palsgrave 1530, fo. CCC. lxviii; see also Ruano-García 2010, 109–128; Stein 1997). Another significant work in this respect is the unpublished compilation by Bishop White Kennett (1660–1728) titled the *Etymological Collections of English Words and Provincial Expressions* (1690s, MS Lansdowne 1033).

### 1.4 The Antiquarian Tradition

Kennett’s collection is representative of a genre of early dialect studies in which the ultimate origin and the subsequent history of English are of interest. This type of work is part of an antiquarian tradition that arose in the early modern period and has continued into modern times, albeit usually without the wild claims for the genesis of languages that were typical of early antiquarian works. The writers were well-meaning amateurs, often members of the clergy or military. An example of the latter was the English army officer Charles Vallancey (1721–1812), whose interest in dialects led him to compile a glossary of the archaic dialect of Forth and Bargy, Co. Wexford, in the southeast of Ireland (Hickey 2007, 66–84).

### 1.5 Dialects in the Age of Prescriptivism

The lack of academic concern with dialects prior to the nineteenth century could be attributed to the absence of a scientific framework for the study of language in general. However, in the eighteenth century one recognizes a deliberate neglect of regional features in English (Beal 2010a), and indeed severe condemnation of all language traits that do not correspond to “standard” usage, whatever the latter might mean. More neutral attitudes are visible in the detailed entries offered in some seventeenth-century English dictionaries. As we saw above, William Bullokar and Thomas Blount gave definitions of dialect that look very objective to modern readers. Bullokar’s entry continues:

…”[In] England the Dialect or manner of speech of the North, is different from that in the South, and the Western dialect differing from them both. The Grecians had five especial Dialects: as in The property of speech in Athens: 2 in Ionia; 3. In Doris; 4. In Eolia: and 5. that manner of speech which was generally used of them all.

(Bullokar 1616 [no pagination]).

Blount’s definition is yet more comprehensive:

In England, the Dialect, or manner of speech in the North, is different from that in the South; and the Western differs from both. As in this example: At London we say, *I would eat more cheese if I had it*, the Northern man saith, *Ay sud eat mare cheese gin ay hader*, and the Western man saith, *Chud ee’at more chiese on chad it*: *Chud ee’at more cheese un ich had it*. The Grecians had five especial Dialects … So every Country commonly hath in diverse parts of it some difference of language, which is called the Dialect or Subdialect of that place. In Italy, there are above eight several dialects or Subdialects…

(Blount 1656 [no pagination]).

Blount is remarkable in that he gives examples to illustrate different dialects of English. However, he is not followed by others. Somewhat later (1676), Elisha Coles published *An English Dictionary* in which he sees dialects as “Logick, speech; also a particular Propriety or
Idiom in the same speech,” with no reference to regions whatever. Some authors do at least specify that dialects are found in different parts of a country, for example, John Kersey, whose *Dictionarium Anglo-Britannicum* (1708) defines dialect as “a Propriety of manner of Speech in any Language, peculiar to each several Province or Country.” Reference to region is also found in Thomas Spence (Beal 1999) who says of dialect that it is “A polite manner of speaking, or diversity made in any language by the inhabitants in any part of the country where it is spoken; stile; speech” (Spence 1775 [no pagination]).

As noted earlier, the great lexicographers of the eighteenth century ignore dialect’s regional essence, at most referring to dialects of classical Greek. Instead, they concentrate on its meaning as a manner of expression. Samuel Johnson offers the following in his authoritative *Dictionary of the English Language* (1755):

**DIACET**

1. The subdivision of a language; as the Attic, Doric, Ionic, Æolic dialects.
2. Stile; manner of expression.
   
   When themselves do practise that whereof they write, they change their dialect; and those words they shun, as if there were in them some secret sting. Hooker, b.v.s. 22.
3. Language; speech.

Later language commentators and prescriptivists, above all, Thomas Sheridan (1719–1788) and John Walker (1732–1807), were content to adopt and repeat Johnson’s definition. Neither Sheridan nor Walker had time for dialectal variation, which directly conflicted with their standard English ideology; hence, their derisory comments on the regional speech of Britain and Ireland.

Besides the prescriptivism of authors like Sheridan and Walker, another motivation is recognisable in the eighteenth-century neglect of dialect. Consider this statement in George Puttenham’s *The Arte of English Poesie*: “After a speach is fully fashioned to the common vnderstanding, & accepted by consent of a whole countrey & nation, it is called a language.” This is the view of language as a unifying factor, in Puttenham’s case among the different regions of Britain. Here we have a very early reference to a “national language,” a notion picked up by later authors such as the Scot James Buchanan in his 1766 *Essay Towards Establishing a Standard for an Elegant and Uniform Pronunciation of the English Language Through the British Dominions*. Attention to dialects, let alone their valorisation, would not have been reconcilable with the desire for a “national language.”

By the close of the eighteenth century, notions of national language and standard language would seem to have merged, at least for many authors. The variation which was to be suppressed was regional (Beal 2010b), with the parallel promotion of the English of southeast England. As the variation was principally phonetic, one’s accent came to indicate one’s relative standardness as a speaker (Beal 2004), which remains the case today (Beal 2008).

### 1.6 The Denigration of Dialects

After the early eighteenth century the assessment of pronunciation appears to have changed (Beal 2010b). While Defoe in the 1720s could remark non-judgmentally on the attitude of Northumbrians to features of their pronunciation, after the mid-eighteenth century comments are far more critical. A vocabulary was adopted by authors on language that is condemnatory of all features that were not part of received southeastern English usage.

“Vulgar” is a censorious epithet used to describe a variety disapprovingly. The term was very common in the eighteenth and nineteenth centuries in evaluative treatments of
language like Savage’s *The Vulgarisms and Improprieties of the English Language* (1833). However, before the eighteenth century “vulgar” simply meant “of the people” (cf. Latin *vulgus* “common people”). John Walker is particularly keen to specify what he thinks merits the label “vulgar.” For instance, given that provincial speakers had to look to the capital for phonetic guidance, any “vulgarisms” used by Londoners are especially to be condemned. In his *Critical Pronouncing Dictionary* he lists “faults of the Londoners,” who, “as they are the models of pronunciation to the distant provinces, ought to be the more scrupulously correct” (1791, xii).

The early and late modern periods saw increasing divergence of English pronunciation and spelling, not solely as a result of the Great Vowel Shift (Pyles and Algeo 2004, 170–173). Several other developments contributed to this divergence, for example, the lowering and unrounding of short [u] to [ʌ] in the strut lexical set and vowel lengthening in bath words. Many changes in this period resulted in homophony and hence led to distinctions in spelling which did not correspond to pronunciation differences (e.g., the merging in southeastern English of the term and turn sets to a rhotacised schwa, which then simplified solely to schwa). But many dialects retained this distinction, which led to the stigmatisation of these varieties.

Twentieth-century scholars revisited the issue of social stigma (Lippi-Green 1997) both for those dialects with an ethnic basis—for example, African American English (Rickford 1999) or Chicano English (Fought 2006)—and those on the periphery of industrialised societies, for example, the “Ocracoke Brogue” of North Carolina’s Barrier Islands (Wolfram and Schilling-Estes 2004). The educational implications of the stigma experienced by speakers was often a primary concern (see Wolfram, Adger and Christian 1999; Baugh 2004).

### 1.7 From Philology to Linguistics

In scholarly literature the term “philology” has two similar but distinct meanings. The first is the study of older texts, whereas the second is the comparison of older stages of languages. The second meaning usually has the longer designation “comparative philology.” This scholarly activity was common in the nineteenth century, when the family relationships among languages, chiefly Indo-European ones, were reconstructed on the basis of older textual records.

Hale (2007) states in his handbook of historical linguistics that he sees philology as the scrutiny and analysis of historical artefacts, which do not represent language but are imperfect windows on language. He claims that “Philology is responsible for establishing the attributes of a text, many of which may be relevant for subsequent linguistic analysis” (2007, 21), and continues, “[t]here are two goals, related to one another, of this enterprise: to understand the linguistic structures present in the text itself (let’s call this the ‘local’ goal) and to understand the structures, entities, and processes which made the grammar of the ‘composer’ of the text (let’s call this the ‘ultimate’ goal)” (2007, 23).

In the present chapter it is the second meaning of “philology,” which is used. In this sense, “comparative philology” is synonymous with historical linguistics as practised throughout the nineteenth century. Because at the time the concern of linguists was overwhelmingly with members of the Indo-European language family, the field is commonly known as Indo-European studies (German *Indogermanistik*). It arose in the late eighteenth century, triggered by the work of Sir William Jones (1746–1794), who insisted on the relatedness of the Indo-European languages, then as now one of the major language families of the world and certainly the best researched. Jones was followed by others such as Rasmus Rask (1787–1832) and Jacob Grimm (1785–1863), who established the science of comparative philology at the beginning of the nineteenth century. It was the dominant school of linguistics until the advent of structuralism at the turn of the twentieth century.
1.8 Features of Indo-European Studies and Comparative Philology

1. The dissociation of linguistics and philosophy
2. The establishment of a sound foundation for etymology
3. The abandonment of attempts to prove putative relationships between Hebrew (the language of the Old Testament) and various European languages
4. The development of a descriptive apparatus for phonetics

In the latter half of the nineteenth century the Neogrammarian hypothesis was developed by a group of young linguists (the *Junggrammatiker*, lit. “new grammarians”) working in Leipzig. They assumed that language change proceeds gradually on a phonetic level, affecting all input sounds simultaneously. The Neogrammarians’ confidence in their assessment of sound change was fuelled by additional discoveries, most notably that by Karl Verner (1846–1896). Verner gave a satisfactory account—now known as Verner’s Law—of the apparent irregularity in many word forms in Germanic, which had been a concern since the days of Jacob Grimm. Subsequently, belief in the regularity of sound change, tempered by analogy, was fully established. The theoretical underpinnings of the Neogrammarian hypothesis were provided by Hermann Paul in his seminal *Prinzipien der Sprachgeschichte* (1886), in particular the much-debated view that sound laws were exceptionless.

The manner in which the Indo-European languages are assumed to have divided is envisaged by the *Stammbaum* “family tree” metaphor, a notion introduced by August Schleicher (1821–1868). A *Stammbaum* is an inverted tree, with branchings from top to bottom. At the top is Proto-Indo-European, and at the bottom the individual languages of the various branches. The tree representation has also been used to show the interrelationship of dialects.

Indo-European studies/comparative philology involved comparing cognate forms from genetically-related (usually Indo-European) languages with a view to reconstructing the proto-language from which the others were thought to have derived. This allowed scholars to trace superficially different forms back to a single (generally unattested) form. For instance, English *heart*, German *Herz*, Latin *cordia*, and Greek *kardia* can be shown to derive regularly from an Indo-European root *kerd*. The same principle was assumed to be possible when investigating dialects: comparing dialectal forms helped scholars in reconstructing earlier stages of languages, often because key forms missing from more standard varieties were attested in dialects. For example, the northern word *thole* “suffer, endure” is a continuation of Old English *þolian* (cf. German *dulden*), although the form does not exist in present-day standard English.

Indo-European studies/comparative philology was initially a German endeavor, but over the course of the nineteenth century scholars from other countries, for example, Scandinavian countries, the Netherlands, and France, became involved. In England, the main scholar was Henry Sweet (1845–1912), the author of books on phonetics and the history of English in general. He also developed a system of phonetic transcription, the *Romic Alphabet*, an important precursor of the International Phonetic Alphabet primarily promoted by Paul Passy (1859–1940) and a necessary instrument for the documentation of dialects.

1.9 The Dawn of Modern Dialectology: The Beginnings of a New Discipline

The systematic study of dialects began in the latter half of the nineteenth century, although as we have seen there is a long history of observation of dialect differences prior to this. The linguistic analysis of dialect variation is associated with the rise of historical linguistics,
which led to publications such as Walter Skeat’s overview of historical dialects (Skeat 1912). But it was the activity of scholars dedicated to documenting traditional dialects in danger of dying out, which was to prove more relevant for later dialectology. Investigations of dialects usually produced maps on which isoglosses were drawn. An isogloss is a line separating the occurrence of two different but related forms, for example, the isogloss which separates the occurrence of [ʊ] (in the north of England) from [ʌ] (in the south of England) in the strut lexical set (Wells 1982, 131–133), or that separating the regions in which non-prevocalic /r/ is present (the “rhotic” areas) versus absent (“non-rhotic” areas, which account for practically the whole of England other than the southwest and a small area of Lancashire). While isoglosses seem useful and present a neat picture of sound distributions, they only apply to speakers of traditional dialects and even then cannot depict the co-occurrence and statistical distribution of variants in transitional regions, let alone do justice to relevant factors such as age, gender, class, rural/urban divisions, and so on (all of which would be relevant to the documentation of a feature such as word-initial /h/ in English).

Nonetheless, isoglosses were an integral part of surveys of traditional dialects in the twentieth century before the advent of sociolinguistics. The prime example of such a survey in England is the comprehensive Survey of English Dialects (SED), initiated by Harold Orton (1898–1975), who had studied under Joseph Wright and Henry Wyld. He was appointed professor at Leeds after World War II, and together with Eugen Dieth (Zurich) supervised the SED. The survey involved over 1,000 questions per questionnaire, covering pronunciation, grammar, and vocabulary on topics such as farm life, nature, household matters, weather, health, and social activities. It appeared between 1962 and 1978 with the Basic Material being published as four volumes containing informants’ responses to interview questions. Further interpretative volumes were published based on the SED’s findings, for example, Kolb’s Phonological Atlas of the Northern Regions (1966), A Word Geography of England (Orton and Wright 1974), and The Linguistic Atlas of England (Orton, Sanderson, and Widdowson 1978).

Further works based on SED data are Upton, Sanderson, and Widdowson (1987), Viereck and Ramisch (1991), and Upton and Widdowson (1996).

Despite the collaborative work that characterizes the SED, there were also key individuals working in dialectology. In England, the most prominent was Joseph Wright (1855–1930), who set dialect study on a new footing in the early twentieth century. Wright studied in Heidelberg and Leipzig and at these centers came into contact with leading linguists of the day. Later he accepted a professorship at Oxford. He is now known for two works, the English Dialect Dictionary (5 vols., 1898–1905) and the sixth volume of this work, his English Dialect Grammar, all of which are still consulted today, although the coverage is incomplete. Wright’s predecessor, Alexander Ellis (1814–1890), was to become one of the foremost phoneticians and dialectologists of his day, and is remembered for his five-volume On Early English Pronunciation (1868–1889).

In Europe there were also early pioneers of dialectology. The main work of Jules Gilliéron (1854–1926), a French linguist who was instrumental in the development of modern dialectology and areal linguistics, was a multi-volume atlas of French dialects produced at the beginning of the twentieth century. Gilliéron sent out trained fieldworkers to conduct interviews and record data using a consistent phonetic notation. One of Gilliéron’s fieldworkers, Edmond Edmont (1849–1926), conducted no fewer than 700 interviews across France between 1896 to 1900, using questionnaires involving over 1,000 items. The results of his observations, chiefly drawn from male informants, together with results from Gilliéron and his other assistants, were subsequently published between 1902 and 1910 as the Atlas linguistique de la France. In Germany, similarly pioneering work was carried out by Georg Wenker (1852–1911). In 1876, he began sending out questionnaires of some 40 sentences to over 1,200 schoolmasters across the north of Germany, asking them to provide equivalents of words in their local dialect. Over a decade he received about 45,000 completed questionnaires. Wenker
transferred the information to maps that in 1881 were published as Sprachatlas des Deutschen Reiches “Linguistic Atlas of the German Empire,” covering north and central Germany. Wenker continued gathering questionnaires, and in 1926 the first volume of the Deutscher Sprachatlas, based largely on Wenker’s data, was published under the editorship of Ferdinand Wrede. Another notable German is Wilhelm Doegen (1877–1967), who had an interest in recording minority languages and dialects. Doegen studied phonetics in Berlin and later under Henry Sweet in Oxford, where he increased his knowledge of English and the anglophone world. He also became a member of the International Phonetic Association. Doegen’s original recordings of English dialect speakers were destroyed during World War II, but shellac copies survived and in the 1990s the Humboldt University in Berlin started digitizing this material to form the Berliner Lautarchiv corpus.

An exception to the general orientation of traditional dialectology and a precursor of modern studies of language and society is Louis Gauchat (1866–1942), a French-speaking Swiss scholar who in 1905 published a study of language use in the Alpine town of Charmey. His recognition that young people used different pronunciations from older ones, and that females led in the use of new variants, that is, are the vanguard in change, anticipated many of the insights of sociolinguistics as it developed in the 1960s and 1970s (Labov 1972).

1.10 Dialect Societies and Materials

Societies for the study of dialects arose in the nineteenth century in parallel to the activities of scholars. In England, the English Dialect Society was founded by Walter Skeat and lasted from 1873 to 1896, after which it was dissolved voluntarily. In this relatively short timespan the society published some 80 works on the dialects of England.

In America, a similar institution was founded in 1889. The American Dialect Society, mainly dedicated to the study of the English language in North America, published (and still publishes) the academic journal, American Speech, which has successfully adapted to modern developments in linguistics.

Journals dedicated solely to dialectology have sometimes had a precarious existence. The Belgian journal, Orbis, began in 1952 with a focus on dialectology, but went into sharp decline in the 1980s. The recently founded Journal of Linguistic Geography (2013–) is an online journal published by Cambridge University Press. Journals with a broader remit, mainly those that deal with variation from a contemporary sociolinguistic perspective (e.g., English World-Wide (1980–), World Englishes (1981–), and Language Variation and Change (1989–)), have been more successful.

The nineteenth century also saw the publication of dialect dictionaries, often dedicated to specific regions of a country. The North of England is the subject of Brockett’s Glossary of North Country Words (1825, 1846), whereas more restricted locales are treated in works such as Dinsdale’s Glossary of Provincial Words Used in Teesdale [Co. Durham] (1849), Nodal and Milner’s Glossary of the Lancashire Dialect (1875–1882), and Dickinson’s Glossary of Words and Phrases pertaining to the Dialect of Cumberland (1878–1881). Other dictionaries have a broader scope, for example, Pickering’s Vocabulary or Collection of Words and Phrases which have been supposed to be peculiar to the United States of America (1816), whereas some consist of extractions of dialect words from more general works (Wakelin 1987; Görlach 1995), for example, Axon’s English Dialect Words in the Eighteenth Century as Shown in the Universal Etymological Dictionary of Nathaniel Bailey (1883).

The twentieth century saw comprehensive dialect dictionaries attempting complete coverage of a country or clearly delimited region (Penhallurick 2009). The five-volume Dictionary of American Regional English (DARE), compiled under the supervision of Frederick Cassidy and Joan Hall at the University of Wisconsin and published between 1985 and 2012 by
Harvard University Press, gives complete coverage of regional vocabulary in the United States. The comprehensive Dictionary of Newfoundland English covers dialect vocabulary in Newfoundland, Canada. It was compiled by George Story, William Kirwin, and John Widdowson, and first published in 1982.

The list of dictionaries could be extended considerably if those dealing with a single anglophone country were to be included (see Hickey 2014). For instance, the Dictionary of Canadianisms on Historical Principles (1967) is a major lexicographical work compiled under the supervision of Walter Avis (1919–1979). In 2006, a comprehensive revision was initiated at the University of British Columbia as the project DCHP-2, which contains much lexical material of dialect origin in the British Isles.

### 1.11 Dialect Studies

Monographs on dialects come in various guises. Apart from popular literature on local dialects there is academic literature, which can be for a general audience, for example, Trudgill (1990), Hughes, Trudgill, and Watt (2012), or for scholars, basically all other studies. Some studies are in a more traditional mode, for example, Brook (1978 [1963]), Petyt (1980), Wakelin (1972, 1977), Kirk, Sanderson, and Widdowson (1985), or Kolb et al. (1979). Other works have taken the insights of modern linguistics on board, for example, Milroy and Milroy (eds., 1993), Kortmann et al. (2004), and Dossena and Lass (eds, 2009). One can also mention the studies of forms of American English found in Wolfram and Schilling-Estes (2006) as well as, in a more popular vein, Wolfram and Ward (2005).

### 1.12 Data Collection Methods

The initial source of data for dialectology was the wordlist, a collection of words supposedly peculiar to the speech of a region. This plotted a trajectory for dialect studies that was characterized by lexical issues. The words sought were frequently those concerning traditional lifestyles (farming, crafts, and domestic issues in rural life). To glean such data, researchers devised lexical questionnaires containing questions like “What do you call the animal which builds dams in streams and rivers?” or, during interviews, “What do you call this part of the body?” with the interviewer touching his/her knee. The limitations of such methods are obvious, and clinging to them spelt oblivion for many dialect studies.

The techniques of modern sociolinguistics (Podesva and Sharma, eds. 2013), above all the rapid anonymous interview promoted by William Labov (Labov 1966), were often adopted to circumvent the observer’s paradox (speakers’ alteration of their speech while under observation by the linguist). But apart from short stretches of speech used for phonetic analysis, this method was not very suitable. For better or worse, informants were usually aware that they were being interviewed for a survey. Indeed, the increased attention paid to ethical issues demanded that the purpose of a survey be revealed to potential informants in advance. When collecting syntactic data, furthermore, much longer stretches of speech in which constructions might occur are necessary, and so anonymous recording would be impractical.

Other data collection issues moved centre-stage during the later twentieth century. One is randomness: for a survey to be representative, all speakers in a community must in principle have the same chance of being selected for a survey. This principle has been followed in surveys such as Telsur, which formed the basis for the Atlas of Northern American English (Labov, Ash, and Boberg 2005). Another is the issue of speech registers, which speakers have at their disposal. It may be of specific concern for a survey to determine how speakers alter their
speech as a function of formality. To capture this, some dialect studies record people speaking freely, reading a text passage, then a wordlist, to see where linguistic features fall on a cline of formality.

Parallel to these concerns, alternative methods of interfacing with informants were trialled. Interviews in groups generally led to a relaxation in speech style. The withdrawal of the interviewer, with informants being recorded on their own, offered another means of avoiding the observer’s paradox, albeit with new issues of reliability and control arising. These elicitation techniques were chiefly employed for the collection and later analysis of syntactic data (Buchstaller and Corrigan 2011; Walker 2013), which requires a considerable amount of informal material (Schilling 2013).

1.13 Accessibility of Data

Once data have been collected, the issue of presentation to and accessibility for the public is addressed. Traditionally, dialect material has been presented in print, often in several volumes of maps with entries for speakers illustrating specific dialect forms (see Wagner 1958–1964). Increasingly, such information is being presented in a digital format. Some surveys have combined print material with a CD-ROM/DVD, for example, Labov, Ash, and Boberg (2006), often containing clickable active maps linked to sound files associated with speakers from specific locations (see Hickey (2004) as an example). Websites dedicated to dialects/varieties act as sources of general information, for example, the International Dialects of English Archive at http://www.dialectsarchive.com, whereas others are more research-based, for example, Variation and Change in Dublin English (http://www.uni-due.de/VCDE).

1.14 Dialectology and General Linguistics

Dialect geography built upon the Neogrammarian hypothesis that sound change was regular, that is, rule-governed and exceptionless. Although the Neogrammarians’ claims concerning the rules of sound change are substantially correct, dialect situations are more complex and reveal that sound changes are not always exceptionless. Much discussion surrounding this issue has taken place (Labov 1981), with the level of language on which a feature is located playing a role, as well as the attitudes of speakers to incipient or ongoing change that can lead to their promoting or disfavoring changes.

Already by the nineteenth century the issue of how dialect features spread spatially was an issue addressed by scholars. Schleicher’s tree model (see above) regarded dialect diversity as arising through a process of binary branching. A later view is the wave theory developed by Johannes Schmidt around 1870. This sees language changes as spreading out from a center like concentric waves in water.

In the twentieth century, with the focus on urban rather than rural forms of language, new conceptions of feature spread arose. The cascade model of diffusion regards changes as spreading from one urban centre to another without affecting the intervening countryside. An example is the spread of TH-fronting to urban centers around England, which are distant from London, without the intervening rural areas being affected. The size factor seems to be important, with larger cities adopting changes before smaller ones (Britain 2012). There would appear to be some instances of spread in the opposite direction, as captured by the term “counterhierarchical diffusion” (the opposite of what usually happens), whereby features spread from rural to urban settings. An instance of such spread would be fixin’ to, which has been adopted into urban areas of Oklahoma.
1.15 Structuralism and Generativism

The heyday of American structuralism, between the 1930s and 1950s, saw some attempts to apply its principles to dialectology. The best-known of these is probably Weinreich (1954), in which the author argues for a “diasystem,” a superordinate level of structure above individual dialects, which would account for their perceived structural similarities (Weinreich 1954, 389–390). However, this notion proved untenable for many dialects, which developed as separate systems in geographically distinct areas, despite having common historic origins.

The ghost of a unifying underlying structure to dialects was not easily put to rest. In the 1960s and early 1970s attempts were made, for instance by Brian Newton investigating Greek dialects (Newton 1976) or by Martín Ó Murchú examining Irish dialects (Ó Murchú 1969), to show that the assumptions of generative linguistics could help explain dialect-relatedness. This strand of research was not very fruitful, however, and was discontinued. Ultimately, the historical relatedness of dialects was accepted as the source of present-day similarities, and abstractions across synchronic varieties were disfavored.

1.16 Dialectometry

Addressing the question of dialect relations without assuming a single superstructure continues via the approach known as “dialectometry” (Szmrecsanyi 2013). Essentially a European approach to dialectology—employed, for instance, in the analysis of Romance languages (Goebi 1982) or Dutch (Nerbonne and Heeringa 2010; Heeringa and Nerbonne 2013)—it uses numerical classification methods to analyze the apparent relatedness of dialects, and to measure the “distance” between them. The proponents of dialectometry highlight its ability to quantify dialect differences and to offer a measure of language change (Nerbonne 2003), whereas its critics see in its deterministic and mechanistic analysis of dialect variation a relative neglect of sociolinguistically determined variation.

Another approach to the grouping of dialects has become available with the increase in computers’ computational power. This makes use of phenograms, which are graphic representations of the structural similarities across groups of dialects (or languages; see Brato and Huber 2012) for a typical instance of such grouping based on African Englishes. Phenograms do not take the history of forms into account, only their synchronic manifestations. Thus, in the analysis of Spanish varieties discussed in Heggarty, McMahon, and McMahon (2005), Spanish in Madrid and in Bogotá show considerable similarities, but it is not clear whether this is due to accidentally shared developments or to continuity of the original Spanish input to Colombia (Heggarty, McMahon, and McMahon 2005, 85).

1.17 The Rise of New Dialects

Many European languages experienced diversification as a result of colonialism in the key period from 1600 to 1900. New forms of Portuguese, Spanish, Dutch, French and English arose outside Europe. This development led several scholars to consider the processes by which new dialects of established languages arise and acquire specific profiles. There are two main models in this field: Trudgill’s “New Dialect Formation” (NDF; Trudgill 2004, 2008) and Schneider’s “Dynamic Model” (DM), see Buschfeld, Hoffmann, Huber, and Kautzsch (eds. 2014).

NDF is viewed as a historical process whereby a new focused variety arises from a series of dialect inputs, for example, in late nineteenth-century New Zealand. Trudgill postulates the following stages: (1) rudimentary leveling, (2a) extreme variability, (2b) further leveling, (3) focusing. Thus new dialect formation has as its beginning a mixture of dialects, and as its
endpoint a single new dialect. In New Zealand, new dialect formation followed the initial immigration of speakers from different regions of the British Isles. This was a process of dialect mixture from which, over just a few generations, a focused variety arose that was then uniform and distinct from other existing varieties of English. Whereas the progression from input to output is uncontroversial, the question of just what input features survived into the later focused variety has been a matter of debate. Trudgill’s stance is deterministic: the quantitative representation of features across speakers of input dialects (given in percentages) determines whether they become part of the output, with an appeal to linguistic markedness to explain the survival of minority variants such as schwa in, for example, *trusted*. If a feature was used by more than 50% of the English, Scottish, and Irish communities of early anglophone New Zealand, then it survived. For this to have worked, early anglophone New Zealand society would have had to be uniform, with contact among all speakers. Trudgill did not consider the status of immigrants and disputed the role of social factors for the young in following generations, for example, the fact that New Zealand was a British colony and hence, southeastern English features would have been favored by later generations; he also rejected any embryonic identity function for the combination of features that emerged in the later, focused variety (Hickey 2003). Additionally, there is no evidence that in a scenario where sociolinguistic factors apparently played no role the quantitative occurrence of a feature across the early communities would determine its survival. It might well be that in such a situation, if it ever obtained, the survival of features might be random.

The Dynamic Model was devised by the Austrian-German linguist Edgar Schneider to account for the development of English in former British colonies. It stresses the manner in which overseas varieties of English have evolved in specific ecologies and strives to account for which combinations of features have emerged. The model stresses the essential interaction of social identities and linguistic forms, the nature of which accounts in large measure for the profiles of post-colonial Englishes. Contact occupies a central position in Schneider’s model, both between dialects present among settlers, as well as between English speakers and indigenous-language speakers in various colonial locations. Contact-induced change produced differing results depending on the social and demographic conditions under which it occurred, that is, on the local ecology, and on its linguistic triggers (e.g., code-switching, code-alternation, bilingualism, or non-prescriptive adult language acquisition).

The model was first presented in Schneider (2003) and later in more detailed form in Schneider’s (2007) monograph *Postcolonial English*. It assumes that former colonies underwent various stages, which can lead ultimately to the development and differentiation of independent endonormative varieties of English, though this stage has not been reached in all cases. Schneider also proposes that there is a shared underlying process—a unilateral causal implication—driving the formation of postcolonial Englishes, as follows: sociohistorical background > identity of early groups > sociolinguistic conditions of communication and contact > resulting features of the emerging post-colonial variety.

Schneider identifies five stages in the development of post-colonial Englishes: Phase 1: foundation—dialect mixture and koineisation (for locations with multiple dialect inputs); Phase 2: exnormative stabilization—a “British-plus” identity for the English-speaking residents when the colony is established and has secured its position *vis-à-vis* the home country, mostly England (or the United States in the case of the Philippines); Phase 3: nativization involving the emergence of local patterns, often associated with political independence or the striving for this; Phase 4: endonormative stabilization, for example, “national self-confidence,” with codification, usually soon after independence; and Phase 5: differentiation—the birth of new dialects, internal developments now linked to internal socioethnic distribution processes. Further issues considered in Schneider’s model include the distinction of settler and indigenous strands in the early stages of new varieties, the impact of accommodation, and the importance of identity formation.
1.18 Conclusion

Where does dialect study stand today? The scholarly investigation of dialects (Maguire, McMahon, and Dediu 2010) is as dynamic as ever, as evidenced by the present volume and by many recurring conferences such as Methods in Dialectology, and the orientation of research has changed considerably since the advent of modern sociolinguistics in the mid-twentieth century (Shorrocks 2000, 2001). The old style, in which older rural males formed the focus, has been abandoned completely, and is only referenced nowadays where a certain dialect shows nothing but literature of this type. Some of the stock components of older dialectology, such as the isogloss (see above), are no longer viewed as particularly useful as they rest on older, less inclusive conceptions of language variation.

The scope of dialectology has increased manifoldly with its exact extent resting ultimately on terminology. If dialectology is taken to encompass urban, sociolinguistic investigations—many authors speak of “social dialects” (Wolfram and Fasold 1974) or “urban dialects”—then its scope is wide indeed. But this would represent a weakening of the original focus of dialectology comparable to the overextended use of “pidgin” and “creole,” which is often found in the literature. In the sense of the linguistic study of regional forms of language, dialectology has matured considerably in the past half century and has proved its ability to adopt and incorporate insights from neighboring fields in linguistics. Examples are the compilation of corpora (Anderwald and Szmrecsanyi 2009) and the digitisation of existing literature, for example, Wright’s 1910 English Dialect Dictionary (Markus 2009; Markus, Upton, and Heuberger 2010), or the application of methods from the “language variation and change” paradigm (Chambers and Schilling 2013), which evaluates the social determinants of microvariation in speech communities. The ability of dialectology to be enriched by such inputs amply proves its vitality and robustness as a linguistic discipline in its own right.

REFERENCES


Gil, Alexander 1619. Logonomia Anglicana.


Phonology, and Sound Change. Berlin: de Gruyter.


2 The Dialect Dictionary

JACQUES VAN KEYMEULEN

2.1 Introduction

This chapter discusses the making of dialect dictionaries. We dwell on user-oriented meta-lexicographical considerations, and on the macrostructural and microstructural options that ensue. Special attention is devoted to fieldwork procedures for unwritten language varieties. We address the basic questions of fieldwork: what, where, who, how, and how much?

Dialect will be defined here as a geographically determined language variety (local or regional), which is essentially only orally transmitted and is relatively isolated from the roofing standard variety (if any). It is clear that this type of variety is quickly disappearing in most modern societies, under the pressure of standard language forms. We do not dwell on the question of whether a given language variety is a dialect or a language. Language as against dialect (of that language) is, in the first place, a sociological notion rather than a linguistic one, and has to do with the level of codification of pronunciation, orthography, and grammar, usage in official and formal situations, usage in writing, and acceptance by a community of speakers (see Haugen 1966). Some dialects, however, have reached the status of “regional language” (e.g., Low German in Germany, or Limburgian in the Netherlands) under the European Charter for Regional and Minority Languages, which was adopted in 1992 under the auspices of the Council of Europe. These enjoy some of the characteristics of “official” languages.

In defining dialect as a geographically determined language variety, for practical reasons we follow the continental European tradition, although the content of the term is widened—especially in the Anglo-American literature—to include any variety of a given language, even its standard form. For the same reasons, we do not include the lexicography of socially determined lexical varieties such as sociolects (see Green, in press), genderlects, ethnolects, technolects, or other specialized vocabularies, although we think that many of the observations below, especially those pertaining to fieldwork, may be useful for them as well. Nor do we include the lexicography of so called “pluricentric languages,” that is, “languages with several interacting centres, each providing a national variety with at least some of its own (codified) norms” (Kloss 1978, 66-67; see also Clyne 1992). Laureys (1997) coined the useful term natiolect for “a national standard of a language that is spoken in more than one state.” Natiolects (e.g., Irish English, Austrian German, or Belgian Dutch) do not only exist in Europe, but are very typical forms of geographically defined varieties within world languages such as English, French, Spanish, or Portuguese. Attitudes toward natiolectal
vocabulary, and hence its treatment, vary greatly. In some cases, separate dictionaries account for this type of geographical variation (e.g., the Dictionary of South African English on Historical Principles [DSEA] for South African English, or the Variantenwörterbuch des Deutschen of Ammon et al. 2004); sometimes labels are used in the comprehensive dictionaries of the standard languages (e.g., Belgisch Nederlands or België for the Belgian Dutch words in the Van Dale dictionary of standard Dutch). A recent example of a balanced approach to pluricentricity is the Prisma Handwoordenboek (Hofman 2014) for Dutch. Thanks to a thorough analysis of digital text corpora, the dictionary is able to accurately label the words which are natiolectally restricted on a continuum between only “Netherlandic Dutch” and only “Belgian Dutch.” In any case, considering natiolectal differentiation as dialectal nowadays is increasingly frowned upon as being politically incorrect, since natiolectisms are used by the educated part of a language community in formal oral and written settings. Natiolectal vocabulary is part of a co-standard, not of a sub-standard.

We will also focus on dialect lexicography proper, and not on the way in which dialect words are included (and labeled) in standard-language lexicography. Dialectal or regional words may be incorporated in the latter context for a large number of reasons (e.g., to make literature in substandard language varieties more comprehensible). A very early example of interest in the geographical dimension of a vernacular is the multilingual Nomenclator Omnium Rerum (1567) of Hadrianus Junius, who added regional labels to the Dutch words. This tradition—which was to be typical of Dutch lexicography thereafter—was first brought to a pinnacle in Kiliaan’s comprehensive dictionary. In his Dutch-Latin Etymologicum Teutonicae Linguae (1599), he not only translated the headwords into a number of European languages, alongside the definition in Latin, but in a number of cases he also added labels such as fland. (Flemish), holl. (Hollandic), fri. (Frisian), and the like, in order to clarify the geographical origins of his data (Figure 2.1).

In what follows, we will firstly discuss a number of meta-lexicographical considerations concerning dialect lexicography. We then turn to the way the overall wordlist of a dictionary is selected, which leads to its macrostructure. Following this, we devote our attention to the microstructure, that is, the information that is given about the headwords, and to data collection, which in our case boils down to a discussion of lexicographical fieldwork for unwritten language varieties. The final section of the chapter contains some remarks on new technologies and desiderata for the future. Examples are taken from different languages, but with a bias toward Dutch, and especially southern Dutch dialect lexicography.

2.2 The User’s Perspective: Meta-Lexicographical Considerations (Why, and for Whom)

Dialect dictionaries came into being from the eighteenth century onward, in the wake of emerging scientific interest in language history. Moulin (2010, 593) refers to the interesting lexicographic programme of Leibnitz (1697), which aimed at the examination of all German words, the regional ones included. Large-scale dialect lexicography emerged all over Europe—and especially in the German language area—from the end of the nineteenth century onward, together with modern dialectology. A philological scientific paradigm

Figure 2.1 Headword baeye, beye “berry” (Kiliaan 1599).
prevailed, reinforced by Romanticism, with its stress on naturalness, historical preoccupation, and nationalism. Some dialect dictionaries aimed to enrich the comprehensive national dictionaries, whereas others tried to account for both the language and the culture of mainly rural communities. The attitude toward regionalisms, however, was not always positive. In the wake of the rise of standard language ideologies, some dialect word collections came into existence as lists of “errors.” Rézeau (1990, 1471), for instance, mentions Desgrouais (1766), stating: “La source du mal, le patois” [“the source of evil, the dialect,” my translation] in “Les gasconismes corrigés.” Words labelled “Zuidnl.” (southern Dutch) in older issues of the standard Dutch dictionary *Van Dale* were put on lists by the Flemings themselves, in order to warn readers against them. Attitudes (and tolerance) toward geographical and other types of language differentiation may differ from culture to culture (see Rézeau (1990) and Hausmann (1990, 1501) on negative attitudes toward Belgian French words, and Görlach (1990) on changing attitudes toward varieties of English).

Dialect dictionaries may be written for societal as well as for scientific reasons. Dialect speakers may wish to have a reference work of their local or regional language variety for practical or for symbolic motives. A local or regional dictionary is obviously handy when a word is not understood by a non-dialect speaker. Local dialect dictionaries also have a codifying effect (or even purpose) because they usually use a dialect spelling for headwords, thus rendering the dialect “writeable” (see also Bernal and Aymerich, in press). The mere existence of a dictionary is very often considered proof of the respectability of the language variety it describes. Another popular motivation is the safeguarding of the dialect vocabulary for a supposedly dialectless future. With that in mind, it is striking that many amateur dictionaries featuring dialectal headwords do not contain some kind of bilingual (standard > dialect) index, which would indeed disclose the dialect vocabulary to the non-dialect speaker of the future. As things stand now, many dictionaries can only be used by good dialect speakers, because these users are the only ones who can decipher the (sometimes home-made) dialect spelling of the headwords. In any case, without an index by which the dictionary is reversed it is impossible to learn dialect words, because a question like “What is the dialect word for “butterfly?” cannot be answered without a standard > dialect index. Many dictionaries also want to describe the (disappearing) culture of the dialect-speaking community together with the vocabulary going with it, which sometimes entails an extensive encyclopedic component in the microstructure of the dictionary.

Scientific motivations, such as safeguarding endangered language varieties for historical or ethnological reasons, coincide to some extent with popular ones. Monotopical dialect dictionaries do not in principle differ fundamentally from dictionaries of standard languages. Since every local dialect can be considered a language system of its own, the same macrostructural and microstructural options may be taken as for any standard language dictionary. The specificity of scientific dialect lexicography, however, resides mainly in the geographical dimension of the dialectal type of lexical variety. The *WVD*, for instance, considers as its minimal goal the recording of reliable (preferably orally collected) lexical information for every lexical area within the zone of investigation; the maximal goal is to obtain lexemes for every locality, in order to be able to draw word maps. The collection of exclusively orally transmitted vocabulary (and the lexico-geographical patterns therein) is complementary to the dictionaries of both the historical and present-day periods, which are mostly based on written text corpora.

Collecting dialectisms or regionalisms in order to warn against them in favor of the standard language is nowadays considered old-fashioned. Scientific dialect lexicography is normally strictly descriptive. Since the target users are generally situated in a hypothesized dialectless future, the metalanguage (i.e., the language used to define and discuss dialectal forms) is normally the standard language. The dictionary should also be structured in such a way that both words and meanings are interpretable for non-dialect speakers in the future.
In this sense, dialect lexicography is a kind of bilingual dialect/standard-language lexicography. In sum, a dialect dictionary has to be based on the assumed needs of the (future) user, and should ideally answer two main questions: (a) “What does word X mean?” and—perhaps more importantly for the future—(b) “How is concept/meaning X expressed?” Alongside the last question, a user may be interested in other microstructural elements (see below).

It goes without saying that the front matter (introduction) of a dictionary should be explicit about all meta-lexicographical options, and, as has already been pointed out, the back matter should incorporate a standard > dialect “bilingual” index.

2.3 Macrostructural Considerations

Dialect lexicography differs from standard-language lexicography in three main ways: (1) it is essentially based on fieldwork, (2) an oral language variety has to be rendered in writing, (3) the geographical dimension of this type of language variation has to be accounted for. Moreover, the lexicographer has to decide between a contrastive or a confrontative approach (the terms are from Wiegand (1977)) with regard to the standard language. In the first case, the dialect dictionary contains only the words—or meanings or other deviating elements—that do not occur in the standard variety. Such a dictionary was formerly known as an idioticon. In the latter case, the dialect dictionary is meant to account for the totality of the vocabulary, including all lexical and semantic similarities with the standard language.

Any macrostructural option (e.g., options concerning the inclusion of specialized vocabulary, collocations) that is possible for a standard-language dictionary is also conceivable for a dialect dictionary. Although the focus of many dialect dictionaries was, and is, on the oldest layer of the orally transmitted language, it is possible to bring in social parameters other than age alone. The Dictionary of American Regional English (DARE) compiled by Cassidy et al. (1985-2013) is a good example of what is possible in this respect. DARE takes account of age, gender, and educational, occupational, and ethnic groups.

2.4 Onomasiological or Semasiological Arrangement

Semasiological dictionaries present and explain the meanings of a given word or phrase, in contrast to onomasiological dictionaries, in which words or phrases are presented as expressions of semantically linked concepts (meanings, ideas, notions, word families, and similar relationships).

Dialect dictionaries may have a local or a regional geographical scope, but even “local” dictionaries, especially those for urban dialects, very often account for geographical differentiation, along with social variation of different types. The titles of regional dictionaries may sometimes mislead the user, however. The words included in the Westvlaams Idioticon (West Flemish Dictionary; De Bo 1873), for instance, are not necessarily general in the West Flemish dialect, but are recorded somewhere in the Belgian province of West Flanders. Sometimes it is also unclear whether the title of a dictionary is to be interpreted in a purely geographical sense, or refers to a linguistic entity proper. Many words in the Woordenboek van de Vlaamse Dialecten (WVD, Dictionary of the Flemish Dialects), for instance, are not Flemish but Brabantic, since the area of investigation in the east is based on the administrative boundary of the province of East Flanders. Brabantic dialects are spoken in the eastern fringe of that province.

The smaller the geographical scope of a dictionary, the more a semasiological (alphabetical) arrangement becomes evident. If the geographical scope is widened, onomasiology
comes into play. Moulin (2010, 598 ff.) mentions, for example, the Wörterbuch der obersächsischen Mundarten, in which an alphabetical arrangement is disrupted by “nests” of cognates (see below). Many regional dictionaries try to account for the geographical component of the vocabulary by resorting to the mesostructure (i.e., the cross-referencing system) of the dictionary.


The entry Grasmücke “warbler” taken from the Schlesisches Wörterbuch (SW), with a “nest” of words denoting the same concept (formerly) used in Silesia.

Sometimes word maps unburden the dictionary of an excessively cumbersome reference system. Alternatively, dictionaries may simply take the shape of an onomasiologically arranged word atlas for a large region, as is the case for the three regional dictionaries covering the three southern Dutch dialect groups (WVD for Flemish, WBD for Brabantic, and WLD for Limburgian dialects), which are—notwithstanding their titles—in fact not dictionaries but geographically oriented inventories of word usage. They are thematically assigned, each fascicle containing a row of concepts with lists of heteronyms assigned to every concept (see Van Keymeulen 2003). The word collection thus presented offers in the first place material for dialect-geographical research, but detailed semantic information (e.g., the polysemic structure of the meaning) is hard to render in such types of publication.

2.5 Onomasiological Arrangements

An onomasiological arrangement is “inspired by the idea that the reality around us can be roughly divided into a system of concepts [...] That system can be, for instance, logical, philosophical or pragmatic” (Van Sterkenburg 2003, 127–8). Below, we discuss as an example the overall onomasiological taxonomy of the aforementioned southern Dutch dialect dictionaries. The taxonomy is heavily indebted to the tradition that started with Hallig and von Wartburg (1952). Their Begriffssystem (“conceptual system”) was adapted for dialect lexicography (see van Keymeulen 2003) (Figure 2.2). Generally speaking, man is placed at the center of things, and reality is assigned to him in ever broadening circles:

An onomasiological taxonomy of “reality” in our case is a classification of concepts, not of words (or “things,” for that matter), and is thus in principle language-independent. It may be used and adapted for new cultural contexts—and for new dictionaries. The question of whether “chicken” is a bird or a piece of food (and should hence be placed in either Section 4.2 or Section 2.3 in Figure 2.2) can be solved by considering the taxonomy as a dynamic structure instead of a static one. The same concept, indeed, may occur in different classes, depending on the viewpoint one takes. Sometimes the same concept is lexicalized differently, depending on the section it belongs to: in some Flemish dialects a potato is called erpel when grown as a plant, but patat when cooked and eaten.

Any onomasiological taxonomy should be explicit about its specific purpose. In the case of the regional dictionaries of southern Dutch, the purpose of the arrangement is to account for “the concrete coherence of things in daily life” (Weijnen and Van Bakel 1967, 40; my translation) as experienced by the dialect-speaking community. The main framework of the conceptual classification in Figure 2.2 has, in the author’s opinion, universal value, and
adaptations with regard to specific cultural contexts are only needed at its lower levels. The functionality and frequency of the concept in the everyday life of the average dialect speaker are used as guiding principles in the assignment of a concept to a specific section. The concepts relating to sexuality, for instance, are therefore placed in Section 2.2 “Family life” and not in Section 1.1 “Man as a physical being.” A tomato is a plant, but is more frequently thought of as being eaten than grown; hence, “tomato” is in the first place assigned to Section 2.3 “Eating and drinking.”

Onomasiological dictionaries of the type mentioned above should be supplemented with alphabetical indices. The new technologies available offer many solutions for mesostructural needs of all kinds. For an example, see Figure 2.3.

2.6 Semasiological Arrangement

Most dialect dictionaries are semasiologically (i.e., alphabetically) arranged; the standard-language lexicography is—often implicitly—used as an example. In that way, they can efficiently render microstructural information, especially information pertaining to semantics. Dialect dictionaries covering more than one local dialect have to find a solution to account
Het sijsje (Carduelis spinus) is een kleine, geel en groen gekleurde vink met een zwarte kruin. Het houdt zich het liefst op in elzen en berken, waar het op zoek gaat naar zaden en insecten. Zie afb. 28.

Afb. 28. Sijsjes hangen soms ondersteboven in elzentakken.
for diatopical lexical differentiation (see above). Sometimes, though rarely, onomasiological re-arrangements (e.g., word fields) are annexed to alphabetical dictionaries. Moulin (2010, 598) notes for the major wide-area German dictionaries the same “skeletal structure”: namely, a “word-geographic principle” and “the organizational unit of a headword with grammatical information, definitions, illustrative examples, distributional information, and cross-referencing systems.”

2.7 Microstructural Considerations

Microstructural elements of all kinds, such as descriptions of meaning, parts of speech, example sentences, and so on, are best rendered when words, rather than concepts, are taken as the point of departure in the lexicographic description. Since in principle a dialect does not differ from a standard language, the same microstructural options may be taken as those chosen for any standard language dictionary (those concerning definitions, the inclusion of phonological or grammatical information, collocations, example sentences, etymology or labels of any kind). The metalanguage of the dictionary is usually the standard language, although there are exceptions to this rule. A dictionary author indeed may wish to demonstrate the value of a “substandard” variety by rendering both headwords and metalanguage in the very same variety.

The specificity of the microstructure of dialect lexicography resides mainly in the form of the headwords and in the treatment of pronunciation and the geographical diffusion of the vocabulary. In order to form dialectal headwords, a formerly unwritten language variety has to be taken down in writing. Some authors invent orthographies, usually based on the spelling of the standard language, with some adaptations and additions (usually diacritic symbols) in order to try and render the dialect pronunciation in the form of the headword itself, instead of adding it in the microstructure.

râschittink zn. v.; mv. –gen: (boog- schutter) rouwsschieting ter ere van over- leden lid. [FFP 22].

rastieël zn. o.: ruif voor paarden. [rasteel; Ofr. rastelier].

Example of homemade dialect spellings for headwords, in an attempt to render dialect pronunciation. The “Dutchified” (i.e., normalized toward Dutch orthography) headwords would be rouwsschieting and rasteel respectively (Pletinckx 2003).

Dictionaries with a wide geographical scope have to abstract away from local dialects in order to summarize across a sometimes very large amount of phonological variance in the form of the headword itself. The standardization of the entry form sometimes takes the shape of “framing” the spelling of the dialect headword as if the word belonged to the standard language.

Dialect lexicography should ideally pay detailed attention to the geographical component of all microstructural elements, especially on the lexical, phonological, and semantic levels, if the sources permit. In many cases, the geographical component of the microstructure leaves much to be desired, since sources or preparatory studies are simply lacking (see also Niebaum 1989–1991, 663). In the absence of geographical labels, the user should be aware that words and microstructural elements (with definitions of meaning included) may not apply across the whole of the area covered by the dictionary.
Only a tiny number of microstructural elements, for example the labels for the parts of speech, can be added by the lexicographer him- or herself. Example sentences are usually invented by the author, instead of being taken from sources, in order to illustrate the word meaning; the best example sentences also illustrate dialectal morphosyntactic particularities. Since many lexicographers want to reveal the culture of the dialect speaker, the encyclopedic component of the microstructure is often very extensive (e.g., technical descriptions, illustrations, etc.).

2.8 Data Collection by Fieldwork

2.8.1 Introduction

Dialect lexicography may be based on oral or written sources. Some “regional languages” of different types do enjoy a—sometimes very old—written tradition, the quantity and quality of which varies from language to language (e.g., Low German, Scots, Occitan, Jamaican English, Papiamento). Dialect texts can be used as sources for dictionaries, but the lexicographer has to be aware of the fact that the “dialecticity” level of the texts may be unclear. The language may vary between dialects with standard language interference and standard language with a dialect flavor. Transcripts of dialectal audio material could in principle serve as a source as well, but it would be extremely time-consuming to establish a large enough corpus for a comprehensive dictionary. Sometimes dictionaries take older dictionaries as their basic material, or already existing wordlists are revised and enlarged. We do not discuss this “plagiarism in alphabetical order”—as it has been jocularly called—here.

In what follows, we focus on synchronic fieldwork for a purely oral language variety. We try to propose answers to the five main questions concerning lexicographical data collection: what? where? who? how? and how much?

2.8.2 What?

For a purely oral language variety, both the words and the microstructural elements have to be collected by fieldwork, depending on the selections based on the meta-lexicographical options. The lexicographic focus is normally on words and meanings; dialect lexicography is also typically preoccupied with pronunciation. For some microstructural elements (e.g., morphology), lexicographers very often try to rely on already existing research, if there is any.

2.8.3 Where?

The geographical scope of a dictionary should be well defined. Many dictionaries have too ambitious a title in respect of geography. In a regional dictionary, geographical representativeness should be an important aim, that is, every lexeme/meaning within the area of investigation should be present. This can be guaranteed by establishing a premeditated geographical network of measuring points, which implies that a pilot study designed to determine the lexico-geographical pattern of the area has already been carried out. It is a scientific requirement to be as explicit as possible with regard to the localization of all data. A hybrid dictionary-cum-atlas ideally has one good data point for every locality. This ideal is seldom met, however.

2.8.4 Who?

Informants should be chosen in accordance with the macrostructural options. Lexical knowledge, as it happens, is distributed unevenly in the language community. In traditional dialectology, “NORMs” (Non-mobile Older Rural Males) were usually chosen as informants.
(Chambers & Trudgill, 1998, p. 29), since the main aim was to collect the lexicon of the oldest dialect layer. The selection of informants is not always an easy job. The lexicographer may have to call in the help of an intermediary who takes an interest in the project and is in a position to introduce the researcher to possible informants. Schoolteachers and clergymen/women are very suitable as intermediaries because they are normally highly respected and know a lot of people.

Every informant should meet both objective and subjective criteria. The objective criteria have of course to do with the sociological parameters with regard to the macrostructural options. In any case, the relevant biographical information for every informant should be carefully inquired after and noted down. The subjective criteria can be summarized as follows: interest in the dictionary project, intelligence, introspective capacity in matters of language and talkativeness. Cooperation with the researcher always implies some degree of education on the part of the informant: s/he has to understand the purpose of the project and, in the case of inquiries by correspondence, has to be able to write and read. Men are considered more suited as informants for substandard language varieties than are women; Chambers and Trudgill (1998, p. 61) call the observation that women on average use a greater frequency of higher-status variants than men “the most strikingly consistent finding of all to emerge from sociolinguistic studies in the industrialised western world.”

Informants should always be tested. They evidently want appreciation for their work (e.g., getting informed about the results of the project). Normally, informants are volunteers; if they are paid they tend to get too cooperative, and may deliver nonce answers.

2.8.5 How?

Data collection methods vary according to the language component investigated (phonology, morphology, syntax, lexicon), the relative validity of the elicited data, and the feasibility of data collection being carried out systematically. All methods aim to get as close to “spontaneous speech” as possible. Some methods, however, use experimental settings and produce highly structured data (for eliciting methods in dialectology in general, see also Seiler, 2010). In what follows, we briefly comment on the most important collection methods in lexicography.

2.8.6 (Self-)Observation

It is of course an important advantage if the lexicographer is describing his or her own dialect, although a sound distrust of one’s own knowledge is advisable; self-observation should be checked with good informants. This is especially the case for regional dictionaries, since it is impossible to have good intuitions about the totality of the geographical differentiation within the area of investigation. The lexicographer’s introspection is used to deliver or assess data (see Bergenholtz & Mugdan, 1990, p. 1613).

Many lexicographers are attentive listeners, and carry with them notebooks in order to note down words and sentences they overhear. This may yield useful information (e.g., lively example sentences), provided the overheard individual was indeed a representative speaker of the dialect. Many amateur dictionaries are based on material collected in this way, sometimes over many years. In many cases, dictionaries are based on the observations (and notes) of a whole group of volunteering dialect speakers.

Corpora of recordings of so-called “free speech,” whereby a good dialect speaker is invited to talk about something s/he is interested in, may render highly spontaneous language material, which is especially suitable for phonological or syntactic research. For lexicographical purposes, though, this method has only a limited value, as the resulting
corpora are too small (see e.g., Freiburg English Dialect Corpus (FRED), Kortmann & Wagner, 2005). The major disadvantage of observation is that, for obvious reasons, negative evidence is impossible: words that by chance do not occur in the corpus of course cannot be observed. Transcriptions may come in handy for example sentences, or for citations of the closed word classes (which are relatively highly frequent and hard to elicit via questionnaires).

Transcribing dialect recordings is very time-consuming. As time goes by, the transcription of existing collections (e.g., the recordings of the Dutch dialects at the Meertens Institute in Amsterdam and at Ghent University) becomes ever more urgent. Dialect recordings are best transcribed by a native speaker of the dialect, who understands not only the traditional dialect itself, but also the subject matter of the conversation. Transcribers who combine both skills are becoming rare.

2.9 Purposive Systematic Fieldwork

Nearly all major dialect dictionaries are based on purposive systematic fieldwork aimed at investigating the lexical knowledge of informants, using questionnaires administered orally (direct method) or by correspondence (indirect method). Purposive fieldwork is the best method of inquiring into passive dialect knowledge; in dialect loss situations, many words only survive in the memories of members of the oldest generation and have disappeared from active usage. The warning of Seiler (2010, 514) should always be taken to heart: “It must be kept in mind that elicitation procedures create artefacts (task effects, repetition effects, order effects) arising from the unnatural situation the informant is exposed to.” It is in fact these artefacts that are the source of information. Methods may differ according to the type of vocabulary investigated; some elements of the lexicon are more easily collected by explicit questioning than others (open versus closed word classes, say, or words with concrete versus abstract meanings), because the introspective capacity of a language user differs according to the different lexical types. The indirect method is considered better for taboo words. Purposive fieldwork normally is carried out thematically, although the ensuing dictionary is alphabetical. Below, we give a brief overview of fieldwork methods used in dialect lexicography. Fieldwork methodologies are also discussed in handbooks for anthropology or sociolinguistics.

2.10 Oral Investigation (Direct Method)

The advantages of oral investigations are evident. The fieldworker controls the interview situation: clarification, feedback, taking notes, and recording are possible. Good phonetic data can only be gathered by this method. The most important disadvantage is the time-consuming nature of the method; it is, however, nearly always used at the initial stages of a project, when the area of investigation is being explored and questionnaires tested. The so-called “observer’s paradox” (Labov 1972, 209) (i.e., “spontaneous” speech is influenced by the very presence of the researcher) can to an extent be overcome by putting the researcher in a minority position: interviewing a (small) group of informants, who know each other and are used to speaking the sought-after language variety among themselves, can prove very fruitful, especially because the informants can correct each other. The “intersubjectivity” of the group is thus used to yield valid data that is as close to “objective” as possible. Of course, the fieldworker has to take care to control the group dynamics (e.g., preventing certain informants from becoming too dominant, or inviting more reticent informants—often women—to participate).

The creation of both a social and a scientific “common ground” between interviewer and interviewee(s) is very important. They should—implicitly or explicitly—agree on the aims of
the interview and the definitions of central notions (e.g., what “dialect” is). The fieldworker should also have a sound knowledge of the preconceptions, motivations, and language attitudes of the informant, in order to understand that person’s reactions to questions.13 In the case of cooperation on a regular basis, an informant can ideally be turned into a language consultant through on-the-job training. The contact language may cause problems, however; sometimes intermediaries have to be asked to conduct the interview in the presence of the researcher.

A purposive systematic oral investigation is normally structured via a questionnaire (see below), which serves as a steering instrument for a conversation about a specific theme. The task of the researcher is to analyze this meta-lexicographical discourse after it has been collected. Van Keymeulen (1986) investigated some kinds of explanation strategies used by informants: translations into standard language, examples/situations of usage, pointing to things, antonyms/synonyms, naïve analytical definitions, and so forth. In the exploratory stages of the research, an informant should be given enough room to comment or to make associations.

2.11 Investigation by Correspondence (Indirect Method)

Although the direct method is always superior, investigation by postal or electronic correspondence is well suited to (large-scale) lexicographical research. It renders good, and comparable, data. Informants have time to think and the observer’s paradox is circumvented. Feedback is, however, not possible. A major problem concerns how to interpret the notation of dialect words by unskilled informants, who use the standard spelling system or a home-made orthography. The interpretation of these data entails a good knowledge of the dialect phonology on the investigator’s part.

2.12 The Questionnaire

Questionnaires are used in both the direct and the indirect method. In the former case, they structure a meta-lexicographical discourse between fieldworker and informant; in the latter, the questionnaire needs to be designed with the utmost care, since no immediate feedback is possible. Questionnaires are normally thematically organized, and for the most part the metalanguage is the standard variety, which may cause problems for a dialect-speaking informant. Some projects have made extensive use of intermediaries who assisted the sometimes illiterate informants in filling out questionnaires.

The preparation of a lexicographical questionnaire in the case of a relatively unknown dialect has the following phases: (1) exploration of the extra-linguistic reality (ethnological investigation); (2) establishing an inventory of concepts; (3) selection of lexically relevant concepts (e.g., the folk taxonomies in biology); and (4) establishing the way word meanings are stored in the mind of the informant, so as to construct good elicitation techniques. This preparation is normally done orally. Luckily, most lexicographical projects do not begin from scratch, although in some cases (e.g., specialized lexica of traditional crafts) it is a dialect dictionary which is the first description, not only of the words, but also of the craft itself.

2.13 The Structure of the Questionnaire

The structure of the questionnaire should demonstrate the empathic capacities of its compiler, who has to try to formulate questions and tasks in such a way as to lead the informant to a good answer. The informant’s biographical data (and that of the intermediary, if any)