

An Introduction to
Historical
Linguistics
Fourth Edition

Terry Crowley | Claire Bowern

OXFORD

UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Copyright © 2010 by Oxford University Press, Inc.

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, NY 10016

www.oup.com

Oxford is a registered trademark of Oxford University Press.

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording, or otherwise,
without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data
Crowley, Terry.

An introduction to historical linguistics /
Terry Crowley and Claire Bower—4th ed.

p. cm.

Includes bibliographical references and index
ISBN 978-0-19-536554-2 (pbk.)

I. Historical linguistics. I. Bower, Claire, 1977—

II. Title.

P140.C76 2009

417'.7—dc22 2008048617

Printed in the United States of America
on acid-free paper

Preface to the Fourth Edition

Like Terry Crowley, I was an undergraduate at the Australian National University and I did honors in historical linguistics just over 20 years after his time there. Terry was still a legend in the Department, the student who was invoked as the exemplar of the Golden Age of ANU. It was therefore with a great deal of trepidation that I approached the task for preparing a new edition of the text. As Bill McGregor has written in his introduction to his 2006 book, it is no easy task to edit a work where the authors have died, where there is no possibility of discussing intentions or various possible additions or subtractions from the text.

The field of historical linguistics has changed a great deal since the first draft of this book was written in the early 1980s, with much more access to data on understudied languages and much more progress in reconstruction in many areas outside Europe. We are seeing the increasing use of computational modeling within historical linguistics and interdisciplinary research is not the exotic enterprise it used to be. Grammaticalization is a much larger part of the field than it was in 1983 and now forms a substantial link between historical linguistics and typology. Therefore my main aim in preparing the new edition has been to *add* rather than *change* (although, of course, in order to do that, some sections of the previous edition had to be removed). I have kept stylistic changes to a minimum (for example, I retained Terry's "I" rather than changing it to "we"). I have also tried to broaden the appeal of the book from a text concerned primarily with Oceania and Australia to one presenting examples from all over the world. However, I hope readers agree that the new

text is not so far removed from those aims and that the Australiasian focus is still strong. This is still Terry's book.

In addition to the correction of some data errors, I've updated the suggestions for further reading and have included more articles from journals, as well as introductory materials. I have added some new exercises and some additional datasets. I have also reorganized the text and omitted and condensed some of the original chapters (for example, the chapter of causes of change is now condensed into chapter 1, and the methods for glottochronology have been omitted). The chapter on "problems with traditional assumptions" has been incorporated into other parts of the text. I have added sections on long-distance relationships and computational methods in historical linguistics and expanded the sections on historical morphology and syntax. I have also altered certain sections to bring the text more in line with current consensus thinking. In doing so, I have no doubt introduced things that Terry would not have agreed with, to which I can only say that I wish I could have discussed them with him.

Claire Bowern

Preface to the Third Edition

Having taught various linguistics courses at the University of Papua New Guinea (UPNG) and since then at the University of the South Pacific (USP), it has become apparent to me that the English used by writers of nearly all standard textbooks in linguistics was far too difficult for English-as-a-second-language speakers. This seemed to be especially true in books dealing with historical linguistics. Also, foreign words and phrases typically abound in textbooks on comparative linguistics, and beginning students are arrogantly assumed to know what is meant by *Umlaut*, *Lautverschiebung*, *spiritus aspi-rate*, *un système où tout se tient*, *sandhi*, and so on. Another problem with standard textbooks for South Pacific students was that the examples chosen to illustrate points and arguments often involved languages that students had never heard of or had no familiarity with—usually ancient European languages and sometimes modern North American Indian languages.

Those of us teaching linguistics at UPNG—mainly John Lynch and myself at the time—decided to remedy these faults for our students by producing our own series of textbooks. John produced a series of notes on linguistic analysis, and my contribution was a set of notes on historical linguistics. In these notes we tried to simplify the language and to explain linguistic concepts in a straightforward manner, yet without simplifying the concepts themselves. We also tried to draw examples as far as possible from languages of this part of the world (rather than from the Northern Hemisphere), as well as from English (this being the language of education with which all tertiary students in the Pacific were familiar).

Contrary to my intentions and expectations, the original UPNG printed notes "Introduction to Historical Linguistics" ended up being used also by students taking comparative linguistics at the Australian National University in Canberra and the University of Auckland in New Zealand. This meant that a set of materials that would have lasted 20 years at UPNG (with our class sizes at the time) was rapidly sold out. This gave me a welcome opportunity to revise the 1981 edition, and a substantially revised second edition appeared under the same title in 1983, in the same UPNG printery format as the first. Again I was pleasantly surprised to find that our stocks were rapidly exhausted, so it was decided to produce a third edition, this time in publisher-produced format.

UPNG Press and the Institute of Pacific Studies (at USP) agreed to publish the volume jointly, and I provided a text based largely on the 1983 version but with some revisions necessitated by the broader audience. This third edition of *An Introduction to Historical Linguistics* appeared in 1987. I would not recommend that anybody try to publish a book in the way that volume was produced, with one publisher in Port Moresby (Papua New Guinea), the other in Suva (Fiji), the typesetter in Auckland (New Zealand), the printer in Suva, and the author by then in Vila (Vanuatu). While the volume received very favorable comment, the results of the geographic dispersal of those involved in the production process are clear to anybody who has used it, as phonetic symbols ended up being cobbled together—some satisfactorily and some less so. Worse, a considerable number of typesetting errors went uncorrected or were even compounded before printing. Many people found that it was difficult to get hold of the volume, as the publishers were not well known among mainstream distributors of academic texts in Europe or North America (copies were difficult to obtain even in Australia and New Zealand). Despite these problems, however, the supply from this print run was also exhausted within a couple of years.

Clearly, in producing this text I had stumbled across a need that was waiting to be met, so I decided to prepare a further edition of this volume. I have taken the opportunity to correct all typographical, factual, and stylistic errors in the previous edition that have come to my attention. I have also taken into account the experience of my peers who have used the previous edition in substantially revising the text itself. I have broadened the content, added a number of sections, and reorganized the presentation of other sections. However, I have consciously decided to maintain the Pacific bias in exemplification. In doing this, I hope that linguists who are schooled in the Western tradition of the English Great Vowel Shift (which I do not mention) and Grimm's Law (which I mention only in passing) are not disappointed. Rather, I hope that this volume makes it possible to show students that the comparative method has universal applicability.

Of course, this is not to say that the model of language change that is assumed by the comparative method described in this volume is universally accepted by modern linguists. There is a substantial—and growing—coterie

of scholars who find many inherent weaknesses in this model. My own work on Pacific pidgins and creoles has left me with many similar doubts. These doubts notwithstanding, I feel that it is probably easiest to show students how languages change by first teaching them the traditional comparative method, just as it is easier to teach classical phonemics than it is to launch straight into underlying phonological representations and morphophonemic rules. Those who are more adventurous or more skeptical can build on the basis provided in this volume to show students how they think languages really change.

Hamilton, New Zealand
October 1991

Terry Crowley

Acknowledgments to the Fourth Edition

I would like to thank my teachers of historical linguistics, who have done a great deal by their example and encouragement. In particular, my advisors Harold Koch and Jay Jasanoff are largely responsible for my being a historical linguist at all. My friends and colleagues Bethwyn Evans (who very kindly gave me her lecture notes as a model for my first historical linguistics course), Lyle Campbell, Simon Greenhill, Russell Gray, David Nash, Luisa Miceli, Mark Donohue, and Paul Sidwell also discussed aspects of the text with me. Many thanks to Rémy Viredaz for providing me with his list of typos and errors, to Malcolm Ross for his notes on the Oceanic reconstructions, to Harold Koch for permission to quote from his lecture handouts, and to Alice Harris for providing the Georgian and Udi problem sets.

Claire Bower

Acknowledgments to the Third Edition

The present form of this volume owes much to a lot of people. First and foremost, I would like to thank my own students of historical linguistics at the University of Papua New Guinea while I was lecturing there between 1979 and 1983. It was largely with their help that I was able to locate areas of inadequacy in exemplification and explanation in earlier versions of this work. In particular, I would like to thank Kalesita Tupou and Sam Uhrle for checking and correcting the Tongan and Samoan data in chapter 5.

Other people have provided a great deal of input. Bill Foley, now at the University of Sydney, helped more than he originally intended in 1979 by providing me with copies of his own comparative linguistics lecture notes and problems, and some of his material has found its way into this book. John Lynch of the University of Papua New Guinea has read and commented on various versions of this work and made specific comments to improve examples and explanations. I would like to thank colleagues in a number of institutions who have used previous editions of this volume as a text in their undergraduate teaching of historical linguistics and others who have made specific suggestions for improvement, in particular Peter Mühlhäusler, Jeff Siegel, Mathew Spriggs, Ray Harlow, Liz Pearce, Lyle Campbell, and Julie Auger. I would also like to thank the reviewers of the previous edition of this book, Scott Allan (*Tē Reo* 32 [1989]: 95–99), Brian Joseph (*Language*

66/3 [1990]: 633–34), and Robert Blust (*Oceanic Linguistics* 35/2 [1996]: 328–35). Despite the various problems associated with the earliest edition, they provided words of both encouragement and constructive criticism. Thanks are also due to Vagi Bouauka of the University of Papua New Guinea and Frank Bailey of the University of Waikato for assistance in the preparation of maps and diagrams.

Finally, I must also make a formal acknowledgment to Jean Aitchison of the London School of Economics. Her published texts *The Articulate Mammal*, *Teach Yourself Linguistics*, and *Language Change: Progress or Decay?* are, to me, a model of clear and simple expression and of ideological soundness. She shows it can be done—I just hope I have achieved it.

Terry Crowley

Contents

| | |
|--|-----------|
| List of Tables | xxi |
| List of Figures | xxiii |
| List of Maps | xxv |
| How to Use This Book | xxvii |
| 1 Introduction | 3 |
| 1.1 The Nature of Linguistic Relationships | 3 |
| 1 How and Why Do Languages Change? | 11 |
| 1.2.1 Anatomy and Ethnic Character | 12 |
| 1.2.2 Climate and Geography | 13 |
| 1.2.3 Substratum | 13 |
| 1.2.4 Local Identification | 14 |
| 1.2.5 Functional Need | 15 |
| 1.2.6 Simplification | 16 |
| 1.2.7 Structural Pressure | 17 |
| 1.3 Attitudes to Language Change | 18 |
| 2 Types of Sound Change | 23 |
| 2.1 Lenition and Fortition | 24 |
| 2.2 Sound Loss | 26 |
| 2.2.1 Apheresis | 27 |
| 2.2.2 Apocope | 27 |

- 2.2.3 Syncope 28
- 2.2.4 Cluster Reduction 28
- 2.2.5 Haplology 29
- 2.3 Sound Addition 29
 - 2.3.1 Excrescence 30
 - 2.3.2 Epenthesis or Anaptyxis 31
 - 2.3.3 Prothesis 32
- 2.4 Metathesis 32
- 2.5 Fusion, Fission, and Breaking 33
 - 2.5.1 Fusion 33
 - 2.5.2 Unpacking or Fission 35
 - 2.5.3 Vowel Breaking 36
- 2.6 Assimilation 36
- 2.7 Dissimilation 43
- 2.8 Tone Changes 45
- 2.9 Unusual Sound Changes 46
- 3 Expressing Sound Changes 55
 - 3.1 Writing Rules 55
 - 3.2 Ordering of Changes 58
- 4 Phonetic and Phonemic Change 65
 - 4.1 Phonetic Change without Phonemic Change 66
 - 4.2 Phonetic Change with Phonemic Change 67
 - 4.2.1 Phonemic Loss 67
 - 4.2.2 Phonemic Addition 68
 - 4.2.3 Rephonemicization 69
 - 4.3 Phonemic Change without Phonetic Change 73
- 5 The Comparative Method (1): Procedures 78
 - 5.1 Sound Correspondences and Reconstruction 78
 - 5.2 An Example of Reconstruction: Proto-Polynesian 79
 - 5.2.1 Setting Out the Data 79
 - 5.2.2 Finding the Cognates 81
 - 5.2.3 Sound Correspondences 82
 - 5.2.4 Reconstruction Principles 85
 - 5.2.5 Residual Issues 93
 - 5.3 Reconstruction of Conditioned Sound Changes 94
 - 5.4 The Reality of Protolanguages 102
- 6 Determining Relatedness 108
 - 6.1 Finding Families 108
 - 6.2 Subgrouping 110
 - 6.3 Shared Innovation and Shared Retention 111
 - 6.4 Long-Distance Relationships 115

- 7 Internal Reconstruction 121
 - 7.1 Using Synchronic Alternations 122
 - 7.2 Internal Reconstruction and Indo-European Laryngeals 125
 - 7.3 Limitations of Internal Reconstruction 127
 - 7.4 Summary: Procedures for Internal Reconstruction 130
- 8 Computational and Statistical Methods 136
 - 8.1 Distance-Based versus Innovation-Based Methods 136
 - 8.2 Lexicostatistics 137
 - 8.2.1 Basic Vocabulary 137
 - 8.2.2 Subgrouping Levels 139
 - 8.2.3 Applying the Method 139
 - 8.3 Criticisms of Lexicostatistics and Glottochronology 149
 - 8.4 Subgrouping Computational Methods from Biology 151
 - 8.4.1 Inferring Correspondence Sets 151
 - 8.4.2 Inferring Subgrouping 152
 - 8.4.3 Some Definitions 153
 - 8.4.4 Selecting Data and Coding Characters 155
 - 8.4.5 Methods for Inferring Phylogenies 157
- 9 The Comparative Method (2): History and Challenges 162
 - 9.1 Background: The Neogrammarians 163
 - 9.2 Convergent Lexical Development 170
 - 9.3 Nonphonetic Conditioning 171
 - 9.4 The Wave Model and Lexical Diffusion 173
 - 9.5 Dialect Chains and Nondiscrete Subgroups 179
- 10 Morphological Change 184
 - 10.1 Changes in Morphological Structure 185
 - 10.1.1 Allomorphic Change 185
 - 10.1.2 Changes in Conditioning 186
 - 10.1.3 Boundary Shifts 186
 - 10.1.4 Doubling and Reinforcement 187
 - 10.1.5 Change in Order of Morphemes 188
 - 10.2 Analogy 189
 - 10.2.1 Analogical Change by Meaning 190
 - 10.2.2 Analogical Change by Form 191
 - 10.2.3 Analogical Extension and Leveling 192
 - 10.3 Doing Morphological Reconstruction 193
- 11 Semantic and Lexical Change 199
 - 11.1 Basic Meaning Changes 200
 - 11.1.1 Amelioration and Pejoration 200
 - 11.1.2 Broadening 200

- 11.1.3 Narrowing 200
- 11.1.4 Bifurcation 201
- 11.1.5 Shift 201
- 11.2 Influences in Direction of Change 201
 - 11.2.1 Metaphor 202
 - 11.2.2 Euphemism 202
 - 11.2.3 Hyperbole 202
 - 11.2.4 Interference 203
 - 11.2.5 Folk Etymology 203
 - 11.2.6 Hypercorrection 204
- 11.3 Lexical Change 205
 - 11.3.1 Borrowing 205
 - 11.3.2 Internal Lexical Innovation 210
 - 11.3.3 Shortening Words 210
- 11.4 Consequences of Borrowing and Irregular Lexical Change 212
 - 11.4.1 Semantic Change 212
 - 11.4.2 Borrowing and Copying 213
- 12 Syntactic Change 217
 - 12.1 Studying Syntactic Change 217
 - 12.2 Typology and Grammatical Change 218
 - 12.2.1 Morphological Type 219
 - 12.2.2 Accusative and Ergative Languages 226
 - 12.2.3 Basic Constituent Order 229
 - 12.2.4 Verb Chains and Serialization 232
 - 12.3 Grammaticalization 233
 - 12.3.1 Direction of Grammaticalization 237
 - 12.3.2 Grammaticalization and Reconstruction 238
 - 12.4 Mechanisms of Grammatical Change 239
 - 12.4.1 Reanalysis 239
 - 12.4.2 Analogy and Extension 240
 - 12.4.3 Diffusion or Borrowing 241
- 13 Observing Language Change 246
 - 13.1 Early Views 246
 - 13.2 Indeterminacy 248
 - 13.3 Variability 252
 - 13.3.1 Class-Based Variation 254
 - 13.3.2 Variation in Small Communities 257
 - 13.4 The Spread of Change and Lexical Diffusion 258
- 14 Language Contact 264
 - 14.1 Convergence 265

| | | |
|--------|---|-----|
| 14.2 | Language Genesis: Pidgins and Creoles | 273 |
| 14.2.1 | Some Definitions | 273 |
| 14.2.2 | Case Study 1: Tok Pisin | 274 |
| 14.2.3 | Case Study 2: Motu | 277 |
| 14.2.4 | Research on Pidgins and Creoles | 279 |
| 14.3 | Mixed Languages | 285 |
| 14.4 | Esoterogeny and Exoterogeny | 287 |
| 14.5 | Language Death and Language Shift | 288 |
| 14.5.1 | Causes of Language Death | 289 |
| 14.5.2 | Young People's Varieties: Structure Change during Language Shift | 290 |
| 14.5.3 | Speed of Language Death | 293 |
| 15 | Cultural Reconstruction | 299 |
| 15.1 | Archaeology | 299 |
| 15.2 | Oral History | 300 |
| 15.3 | Comparative Culture | 303 |
| 15.4 | Historical Linguistics | 305 |
| 15.4.1 | Relative Sequence of Population Splits | 306 |
| 15.4.2 | The Nature of Cultural Contact | 307 |
| 15.4.3 | Sequences of Cultural Contact with Respect to Population Splits | 309 |
| 15.4.4 | The Content of a Culture | 310 |
| 15.4.5 | The Homeland of a People | 312 |
| 15.5 | Paleolinguistics and Language Origins | 316 |
| 15.6 | The Reliability of Cultural Reconstruction | 321 |
| | Datasets | 327 |
| 1 | Palauan (Micronesia) | 327 |
| 2 | Nganyaywana (New South Wales, Australia) | 328 |
| 3 | Mbabaram (North Queensland, Australia) | 328 |
| 4 | Yimas and Karawari (East Sepik, Papua New Guinea) | 329 |
| 5 | Lakalai (West New Britain, Papua New Guinea) | 329 |
| 6 | Suena and Zia (Morobe Province, Papua New Guinea) | 330 |
| 7 | Korafe, Notu, and Binandere (Oro Province, Papua New Guinea) | 331 |
| 8 | Paamese (Vanuatu) | 331 |
| 9 | Motu (Central Province, Papua New Guinea) | 332 |
| 10 | Sepa, Manam, Kairiru, and Sera (Coastal Sepik, Papua New Guinea) | 333 |
| 11 | Burduna (Western Australia) | 334 |
| 12 | Québec French (Canada) | 335 |
| 13 | Tiene (Democratic Republic of Congo) | 336 |

- 14 Cypriot Arabic 337
- 15 Nyulnyulan (Western Australia) 339
- 16 Aroma, Hula, and Sinaugoro (Central Province, Papua
New Guinea) 340
- 17 Proto-Gazelle Peninsula and Daughter Languages
(New Britain, Papua New Guinea) 342

Notes 345

Language References 355

References 359

Index 369

List of Tables

- 1 1 Some Words in Widely Separated Languages 5
- 5.1 Data from Four Polynesian Languages 80
- 5.2 Table of Correspondences 92
- 6.1 Numbers in Assorted European Languages 110
- 8.1 The Swadesh List 141
- 8.2 Data from Central Province Languages, Papua
New Guinea 142
- 8.3 Cognate Sets for Data in Table 8.2 143
- 10.1 Verb Forms in Related Languages Bardi and
Nyikina 193
- 13.1 [ɪ] Variation by Social Class in New York
City (%) 254
- 14.1 Words of Foreign Origin in English with Irregular
Plurals 268

List of Figures

- 0.1 The IPA alphabet xxxii
- 1.1 Wala speakers on an island 8
- 1.2 Walo, Peke, and Puke 8
- 2.1 Average fundamental frequency values (Hz) of vowels following English stops (Hombert et al. 1979: 39) 46
- 8.1 Network of Quechua and Aymara varieties (Heggerty 2008:44) 154
- 8.2 Karnic network 158
- 9.1 Isoglosses in Paama 176
- 9.2 German dialect differences 176
- 9.3 Further isoglosses in Paama 177
- 9.4 Wave model of change 179
- 9.5 Bandjalang varieties in Australia showing lack of mutual intelligibility 181
- 9.6 Language subgroups in part of Vanuatu 182

- 12.1 Typological cycle of language transformation 225
- 13.1 Percentage of [ɪ] tokens in two classes in New York City (Labov 1972) 255
- 13.2 Percentage of [ɪ] tokens in three classes in New York City (Labov 1972) 255

List of Maps

- 0.1 Pacific languages referred to in the text xxix
- 0.2 Australian languages referred to in the text xxx
- 0.3 Papua New Guinea languages referred to in the text xxx
- 0.4 Insular Melanesian languages referred to in text, and major Melanesian lingua francas xxxi
- 14.1 Spread of uvular 'r' in France 270
- 15.1 Areas of the Time of Darkness in Enga Province, Papua New Guinea 302
- 15.2 Using the Age-Area Hypothesis in an imaginary language family 314
- 15.3 Languages of Papua New Guinea 315
- 15.4 Languages of Oro Province, Papua New Guinea 316

How to Use This Book

I would like to think that this book will prove useful to teachers of historical linguistics at all undergraduate levels. I have written it on the assumption that students have already completed at least one basic course in descriptive linguistics, so I have not bothered to define terms such as phoneme, morpheme, or suffix. Some familiarity is also assumed with a distinctive feature analysis of phonology. More specialist linguistic terminology, such as ergative or exclusive pronoun, however, is introduced at its first appearance in the text in small caps and is always explained (and generally also exemplified) for the benefit of students. The linguistic terminology in this volume is used in the same way as in Crowley, Lynch, Siegel, and Piau, *The Design of Language: An Introduction to Descriptive Linguistics*. The bold page numbers in the index indicate where definitions are located.

I have attempted to cover the kinds of topics in historical linguistics that are dealt with in most courses on this subject, as well as enough areas of side interest so that lecturers will be able to follow some of the more specialist aspects of this subdiscipline as well. However, it should be kept in mind that *An Introduction to Historical Linguistics* is just that—an introduction. I have deliberately aimed at breadth rather than depth, and students should be encouraged to use other textbooks for wider reading in order to look at different topics or to look at different interpretations of the same topics. At the end of all chapters, I have included a list of supplementary readings where students can begin this wider reading. I have referred students to readings that are available in fairly well known textbooks, on the assumption that they

will be able to find these in university libraries. For more advanced courses, readings in specialist journals or more advanced textbooks may be necessary. I would suggest that if a higher-level course is being taught, lecturers compile their own supplementary reading lists. [I have added to the reading suggestions, including supplying a number of more up-to-date suggestions but also incorporating my own supplementary reading lists.—C.B.]

I have included at the end of each chapter a set of Reading Guide Questions. Students may want to test their understanding and retention of the material in a chapter by working through these questions. I have not included answers to these questions—if students do not feel confident about a particular answer, they should refer to the material in the chapter or ask the lecturer for help.

Each chapter includes exercises based on some of the concepts discussed in that chapter. These can be used in a number of ways. As a lecturer you may want to use these data as illustrative material in lectures. You may want to ask students to work through these materials in class, as a way of ensuring that they are able to apply the concepts discussed in that particular chapter. Finally, you may want to use the materials as a basis in formulating problems for your students for assessment (for that reason, I have not provided answers to the questions that are given). A number of exercises in this volume involve the same set of basic information on particular languages on which students are asked to perform different sorts of tasks. Rather than repeat this information in each chapter, I have collected the data in a series of datasets at the end of the volume. Students should refer to the datasets for these forms whenever an exercise requires it. [I have retained this format, but I have edited a number of the datasets and introduced a few others. Note also that I have corrected a number of typographical errors in problem sets, and so the versions of problems that appear here may be somewhat different from those in previous editions.—C.B.]

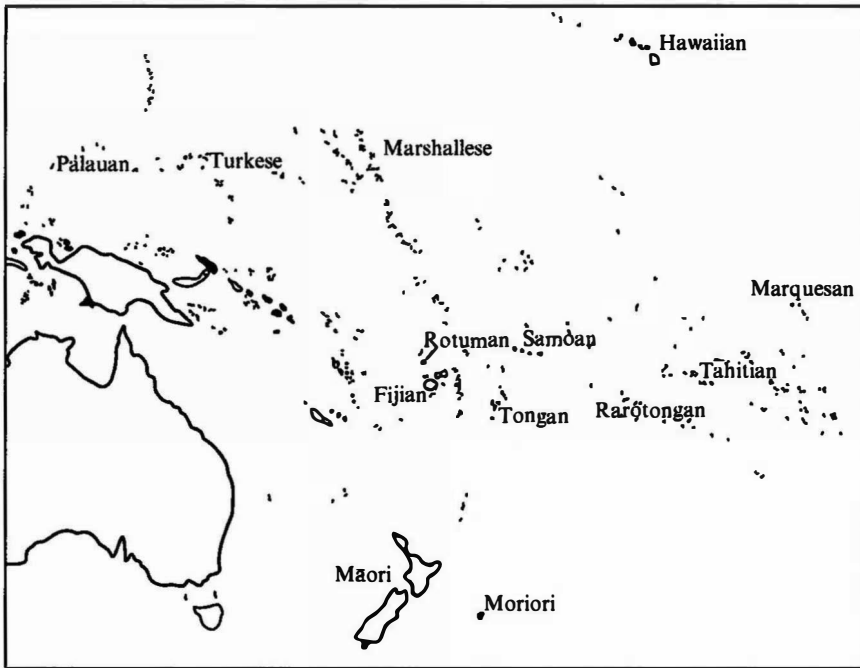
Many examples in this volume are taken from Austronesian languages, Australian languages, and the non-Austronesian languages of the Papuan area. Since this is a textbook of historical linguistics rather than an introduction to Austronesian linguistics (or those of other areas), I hope that specialist readers will accept the occasional simplification—or other kinds of misrepresentation—of data in the spirit that it is intended: that is, as an introduction to principles of historical linguistics.

Readers of this volume should note that I have used phonetic symbols that correspond to those used in Crowley, Lynch, Siegel, and Piau (1995). These symbols are widely used by linguists and correspond for the most part to standard IPA symbols. Conventions that are not widely used are explained as they are introduced. Otherwise, I have used the symbols that are set out on the following page. Readers should also note that English words are generally transcribed to reflect the pronunciations that are typical in Australian, New Zealand, and South Pacific English rather than the pronunciations of

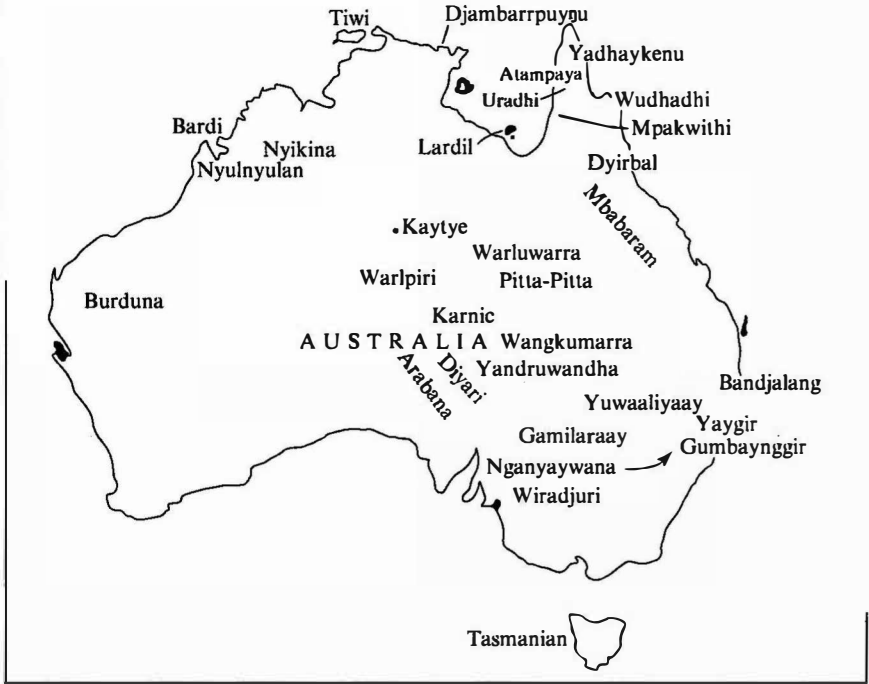
North American and British speakers. North American and British readers, however, should experience little difficulty with most transcriptions.

Material is cited in the text in IPA symbols surrounded either by phonetic brackets or phonemic slashes. For examples that are cited without surrounding brackets or slashes, the phonetic versus phonemic status of the forms is not relevant to the particular point being made. Forms cited orthographically appear in italics.

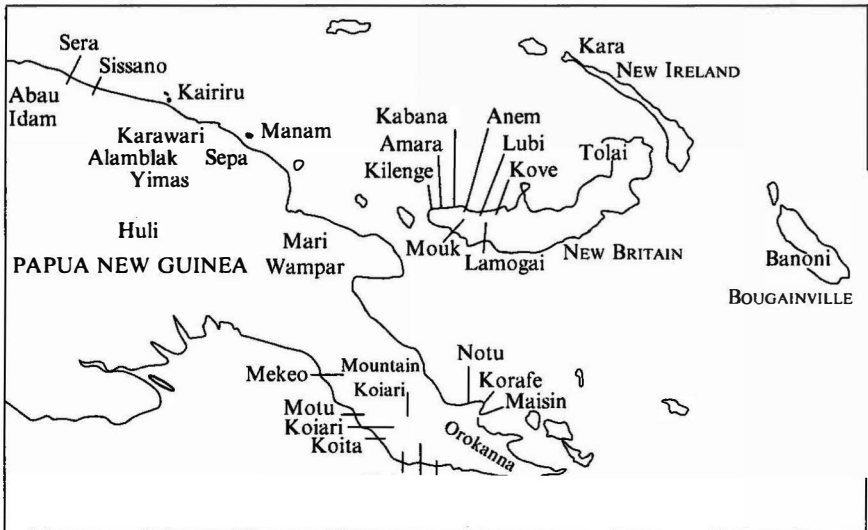
In the following maps I have indicated the location of languages that may not be known to the general reader of this volume. I am assuming that readers will be aware of where the better-known (or iconically named) world languages (such as French, Bahasa Indonesia, Afrikaans, and Icelandic) are spoken. I have indicated the location of lesser-known languages that are spoken outside the areas covered by the following maps in the body of the text.



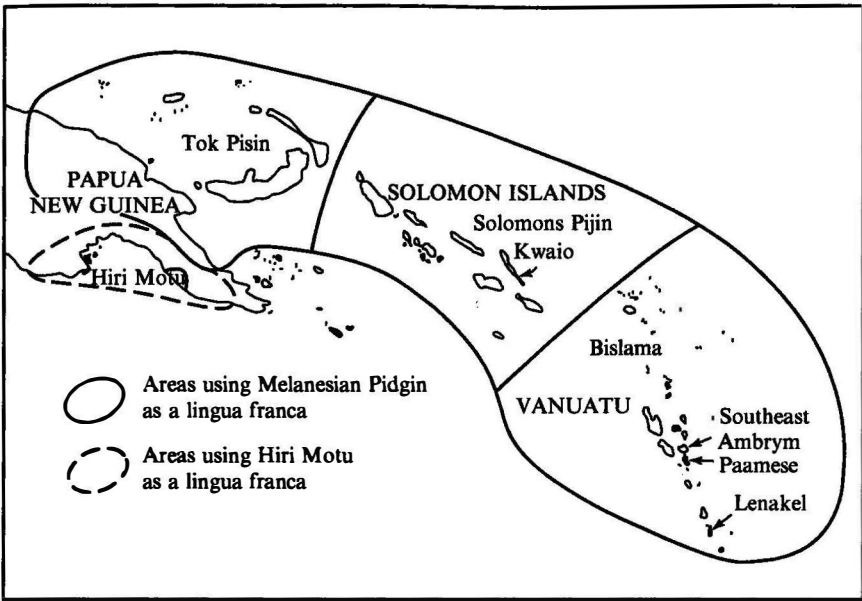
MAP 0.1 Pacific languages referred to in the text



MAP 0.2 Australian languages referred to in the text



MAP 0.3 Papua New Guinea languages referred to in the text



MAP 0.4 Insular Melanesian languages referred to in text, and major Melanesian lingua francas

THE INTERNATIONAL PHONETIC ALPHABET (REVISED TO 2005)

CONSONANTS (PULMONIC)

2005 IPA

| | Bilabial | Labiodental | Dental | Alveolar | Postalveolar | Retroflex | Palatal | Velar | Uvular | Pharyngeal | Glottal |
|-------------------|----------|-------------|--------|----------|--------------|-----------|---------|-------|--------|------------|---------|
| Plosive | p b | | | t d | | ʈ ɖ | c ɟ | k ɡ | q ɢ | | ʔ |
| Nasal | m | ɱ | | n | | ɳ | ɲ | ŋ | ɴ | | |
| Trill | ʙ | | | ʀ | | | | | ʀ | | |
| Tap or Flap | | ⱱ | | ɾ | | ɽ | | | | | |
| Fricative | ɸ β | f v | θ ð | s z | ʃ ʒ | ʂ ʐ | ç ʝ | x ɣ | χ ʁ | ħ ʕ | h ɦ |
| Lateral fricative | | | | ɬ ɮ | | | | | | | |
| Approximant | | ʋ | | ɹ | | ɻ | j | ɰ | | | |

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

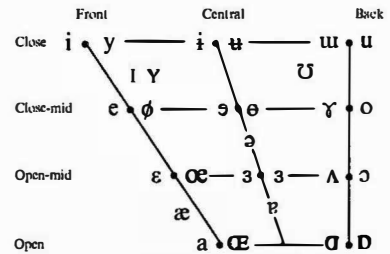
CONSONANTS (NON-PULMONIC)

| Clicks | Voiced implosives | Ejectives |
|--------|-------------------|-----------------------|
| ʘ | ɓ | ʼ |
| ǀ | ɗ | Examples: |
| ǃ | ɟ | ɓʼ Bilabial |
| ǂ | ɠ | ɗʼ Dental/alveolar |
| ǁ | ʄ | ɟʼ Dental/alveolar |
| | ɠ | kʼ Velar |
| | ʛ | sʼ Alveolar fricative |

OTHER SYMBOLS

- ʌ Voiceless labial-velar fricative
- ʷ Voiced labial-velar approximant
- ɥ Voiced labial-palatal approximant
- ʜ Voiceless epiglottal fricative
- ʕ Voiced epiglottal fricative
- ʡ Epiglottal plosive
- ɕ ʑ Alveolo-palatal fricatives
- ɺ Voiced alveolar lateral flap
- ɥ Simultaneous ʃ and x
- Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

SUPRASEGMENTALS

- ˈ Primary stress
- ˌ Secondary stress
- ː Long
- ˑ Half-long
- ˚ Extra-short
- ˌ Minor (foot) group
- ˎ Major (intonation) group
- ˑˑˑ Syllable break
- ˑˑˑ Linking (absence of a break)

kp̚ ts̚

DIACRITICS Diacritics may be placed above a symbol with a descender, e.g. ɲ̩

TONES AND WORD ACCENTS



FIGURE 0.1 The IPA alphabet (reprinted with permission from the International Phonetic Association, copyright 2005 by International Phonetic Association)

AN INTRODUCTION TO
HISTORICAL LINGUISTICS

Introduction

1.1 THE NATURE OF LINGUISTIC RELATIONSHIPS

Many linguists trace the history of modern linguistics back to the publication in 1913 of the book *Course in General Linguistics* by students of the Swiss linguist Ferdinand de Saussure. In this book, the foundation was laid for the scientific study of language. Saussure recognized, as we still do today, that language is made up of a collection of units, all related to each other in very particular ways, on different levels. These different levels are themselves related in various ways to each other. The primary function of language is to express meanings and to convey these to someone else. To do this, the mental image in a speaker's head has to be transformed into some physical form so that it can be transferred to someone else, who can then decode this physical message and have the same mental image come into his or her head.

One of the points that Saussure stressed was the fact that we need to make a distinction between studying a language from a **DIACHRONIC** point of view and from a **SYNCHRONIC** point of view. Up until the time of Saussure, linguistics had been focused primarily on the diachronic study of languages. Languages at a particular point in time were viewed not so much as systems within themselves but as "products of history"; as such, historical considerations could be used in making arguments about synchronic structure. Saussure disputed this interpretation and said that all languages

could (and, indeed, should) be described without reference to history. When we describe a language synchronically, we describe what are the basic units that make up the language (that is, its phonemes, its morphemes, and so on) and the relationship between these units at that time, and that time only. He therefore proposed a rigid boundary between diachronic and synchronic linguistics, which has been part of linguistics since his time (though lately, many linguists have come to question the need for such a rigidly stated view). This book introduces you to the concepts and techniques of diachronic linguistics.

Another important concept that Saussure stressed was the fact that the mental image in a speaker's head and the physical form used to transfer this image are completely arbitrary. This accounts for the fact that a certain kind of domestic animal is called a [sisia] in the Motu language of Papua New Guinea, a [huli] in the Paamese language of Vanuatu, a [ʃiē] in French, and a [dɒg] in English. If there were any kind of natural connection between a word and the thing it denotes, we would all use similar words for similar objects.

Saussure would not have denied that some parts of a language are strongly iconic, or natural. All languages have onomatopoeic words like *rokrok* for 'frog' and *meme* for 'goat' in Tok Pisin in Papua New Guinea, or *kokoroku* for 'chicken' in Motu. However, words such as these are usually very small in number and not an important consideration in language as a whole. Such words are also concentrated in certain meaning categories, such as bird or animal names.

If we compare two different words used by two different groups of people speaking different languages, and we find that they express a similar (or identical) meaning by using similar (or, again, identical) sounds, then we need to ask ourselves this simple question: Why? On the one hand, maybe it is because there is some natural connection between the meaning and the form that is being used to express it (such as between the word *meme* and the sound that a goat makes). On the other hand, maybe the similarity says something about some kind of historical connection between the two languages.

Let us go on a diversion for a moment and look at the topic of stories in different cultures of the world. Probably all societies in the world have some kinds of stories that are passed on from generation to generation, telling of the adventures of people and animals from a long time ago. Often, these stories are told not just for pure interest and enjoyment but also as a means of preserving the values of the culture of their tellers. The fact that all societies have such stories is not particularly surprising. Even the fact that societies have stories about animals that speak and behave like humans is not particularly surprising, as all humans of whatever culture are able to see similarities between animals and humans.

However, what if we found that two different peoples had a particular story about a person who died, and who was buried, and from whose grave grew a tree that nobody had seen before? This tree, the story goes, bore large

TABLE 1.1 Some Words in Widely Separated Languages

| | Bahasa Indonesia | Tolai (PNG) | Paamese (Vanuatu) | Fijian | Māori |
|---------|------------------|---------------|-------------------|-------------|---------------|
| 'two' | <i>dua</i> | <i>aurua</i> | <i>elu</i> | <i>rua</i> | <i>rua</i> |
| 'three' | <i>tiga</i> | <i>autul</i> | <i>etel</i> | <i>tolu</i> | <i>toru</i> |
| 'four' | <i>empat</i> | <i>aivat</i> | <i>ehat</i> | <i>va:</i> | <i>fa:</i> |
| 'five' | <i>lima</i> | <i>ailima</i> | <i>elim</i> | <i>lima</i> | <i>rima</i> |
| 'stone' | <i>batu</i> | <i>vat</i> | <i>ahat</i> | <i>vatu</i> | <i>kofatu</i> |

green fruit right near its top, but nobody knew what to do with this fruit. A bird then came along and pecked at the fruit to indicate to the people that its thick skin could be broken. When it was broken open, the people found that the fruit contained a sweet and nutritious drink.

This story can be recognized by coastal peoples nearly 3,000 kilometers apart, from Vanuatu through to many parts of Papua New Guinea. Surely, if two peoples share stories about the origin of the coconut which contain so many similar details, this cannot be accidental. The fact that the stories are widely dispersed can only be interpreted as meaning that there must be something in common in the history of these different peoples.

Getting back now to language: if we were to come across two (or more) different languages and find that they have similar (or identical) words to express basically the same meanings, we would presumably come to the same kind of conclusion. Look at the forms from a number of widely scattered languages in the table above.

These similarities must be due to more than pure chance. Of course, we do find chance similarities between words in different languages. After all, languages used a fairly small number of sounds, so it is not surprising that the odd word might end up sounding similar in different languages. In such cases, however, there are never SYSTEMATIC SIMILARITIES. Compare the English glosses for the words in the table: the word 'two' is somewhat similar to *dua*, but none of the other words are similar. We must presume that there is some kind of historical connection between these five widely separated languages (and, we might suspect, some of the intervening languages as well), but not between them and English. This connection (and the connection between the stories about the coconut that we looked at earlier) could logically be of two different kinds. First, it could be that copying (or BORROWING) is involved: four of these five languages could have copied these words from the fifth, or they could have copied various words from each other, or all five could have copied from a sixth language somewhere.

Second, it could be that these forms all derive from a single set of original forms that has diverged differently in each case. Since these four languages are spoken in widely separate areas, we could guess that the speakers have had little or no opportunity to contact each other until very recent times. Anyway, even if these people were in contact in ancient times, there would seem to be

little need for people to copy words for things like basic numbers and the word for 'stone'. These are the sorts of things that people from almost all cultures must have had words for already.

It might be understandable if the words for 'coffee' or 'ice' were similar, as these are certain to be introduced concepts in these areas. Originally, these things would have had no indigenous name. When people first come across things for which they have no name, they very frequently just copy the name from the language of the people who introduced the concept. Since traditionally people in the Pacific did not grow coffee (as this drink was introduced by Europeans, who themselves learned of it from the Middle East), we would expect that the word for 'coffee' in most of the languages of the Pacific would have been copied from the language of early European sailors and traders who first appeared in the Pacific in the past 200 years or so. Thus, the word for 'coffee' in most Pacific languages today is adapted to the sound systems of the various languages of the region and comes out something like *kofi* or *kopi*. (In areas of the Pacific where the French rather than the British were influential, of course, we find words like *kafe* or *kape* from French *café*.) The English word itself is borrowed from Turkish (perhaps via Italian) and comes ultimately from Arabic.

Getting back to the words for 'stone' and the numbers 2, 3, 4, and 5 that we saw in table 1.1, the most likely explanation for their similarity in these widely dispersed languages is that each of these sets of words is derived from a single original form. This brings us to the important concepts of LANGUAGE RELATIONSHIP and PROTOLANGUAGE. These ideas were first recognized in modern scholarship by Sir William Jones, who was a British judge in colonial India. Jones had studied a wide variety of languages; in 1786, he delivered a speech about Sanskrit (one of the languages of ancient India), and his words have since become very famous. In this speech he said, among other things:

The Sanskrit language, whatever be its antiquity, is of a wonderful structure; more perfect than the Greek, more copious than the Latin, and more exquisitely refined than either, yet bearing to both of them a stronger affinity, both in the roots of verbs and in the forms of grammar, than could possibly have been produced by accident; so strong indeed, that no philologist could examine all three, without believing them to have sprung from some common source, which, perhaps, no longer exists: there is similar reason, though not quite so forcible, for supposing that both the Gothic and Celtic, though blended with a very different idiom, had the same origin with the Sanskrit; and the Old Persian might be added to the same family. (Jones 1798: 422–23)

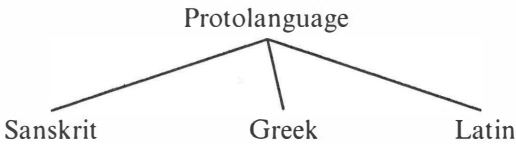
This statement added two significant advances to the understanding of language change at the time. First, Jones spoke of the idea of languages being related. Until then, people had tried to derive one language from another, often with ridiculous results. For instance, people had tried to show that all

modern languages of the world ultimately go back to Hebrew, the language of biblical times. Kings of Europe even went to the extreme of separating newborn babies from their parents to see what language they would speak naturally if they were left alone and not taught. The results varied from Dutch to Hebrew (and none of these claims is believable). The similarities between Sanskrit, Latin, and Greek that Jones was talking about were often explained before he delivered his speech by saying that Sanskrit developed into Greek, and that Greek then developed into Latin:

Sanskrit → Greek → Latin

Jones, however, introduced the idea of “parallel” development in languages. That is, he introduced the idea that there might have existed other languages that have disappeared without leaving a record. The concept that he was introducing was therefore the concept of language relationship. He was saying that if two languages have a common origin, this means that they belong to a single family of languages. (The idea that one language can change into another was long known. Dante, for example, discusses in *De Vulgari Eloquentia* the idea that French, Italian, and Spanish are modern descendants of Latin; see Shapiro 1990.)

Second, Jones spoke of the concept of a protolanguage (without actually using the term, as this did not come into general use until modern times). When he said that these three languages, and possibly the others he mentioned (and he was later shown to be correct), were derived from some other language, he meant that there was some ancestral language from which all three were descended by changing in different ways. So, the model of language and relationship that he proposed to replace the earlier model looks like the model that we use today:¹



The concepts of protolanguage and language relationship both rest on the assumption that languages change in certain systematic ways. In fact, all languages change all the time. It is true to say that some languages change more than others, and faster than others, but all languages change nevertheless. But while all languages change, the change need not be in the same direction for all speakers. Let us imagine a situation as in figure 1.1. We will assume that there was an area on this island occupied by a group of people who spoke a language called Wala. Perhaps under pressure from population density, perhaps because of disputes, or perhaps out of pure curiosity, some of the Wala people moved out across the river and some across the mountains,

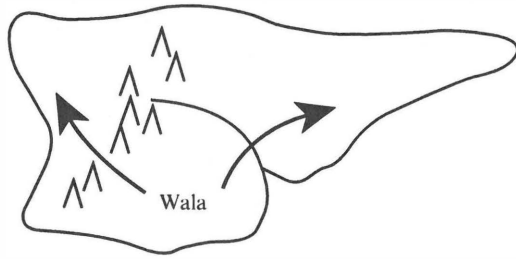


FIGURE 1.1 Wala speakers on an island

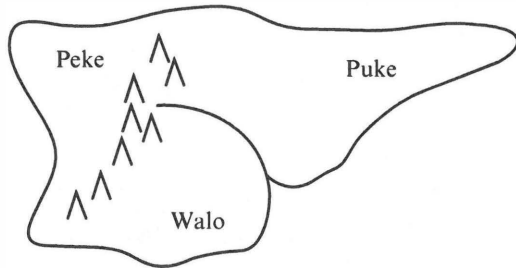
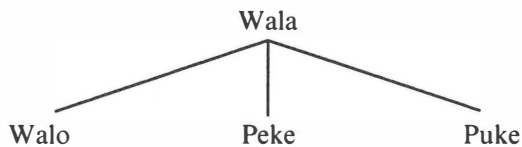


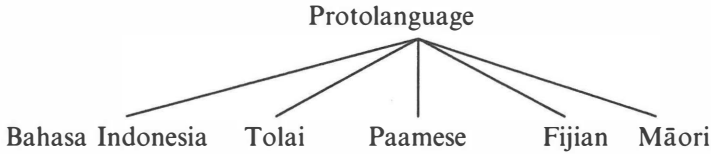
FIGURE 1.2 Walo, Peke, and Puke

and they settled in other areas. As I have said, all languages change, and the Wala language was no exception. However, the changes that took place in the Wala language across the mountains and across the river were not necessarily the same kinds of changes that took place in the original Wala homeland. Eventually, so many changes had taken place in the three areas that people could no longer understand each other. The Wala people in their homeland ended up calling themselves the Walo people, rather than their original name, Wala. Across the river the people came to call themselves the Peke, while the people on the other side of the mountains ended up calling themselves the Puke people. So, what we now have is a situation as in figure 1.2.

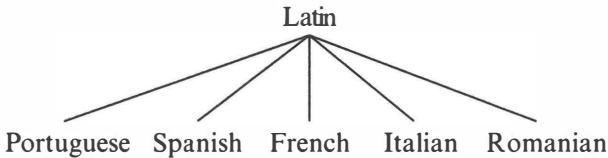
The three languages, Walo, Peke, and Puke, still show some similarities, despite their various differences. What we say, therefore, is that they are all related languages, all derived from a common ancestor, or protolanguage. We could therefore draw a family tree diagram for these three languages that would look like this:



We can say exactly the same kind of thing about Bahasa Indonesia, Tolai, Paamese, Fijian, and Māori. These are all related languages derived from a protolanguage that was spoken in the distant past at a time when writing was not yet known. Thus:



Generally, when a protolanguage evolves to produce a number of different daughter languages, we have no written records of the process. In the case of some of the languages of Europe, however, we have written records going back some thousands of years, and we can actually observe the changes taking place in these records. Latin was the language of most of western Europe at the time of Christ. However, as the centuries passed, Latin gradually changed in its spoken form in different parts of Europe so that it was quite different from the older written records. It is important to note that Latin changed in different ways in what is now Portugal, Spain, France, Italy, and Romania. The eventual result of this was that there are different languages in Europe that are today called Portuguese, Spanish, French, Italian, and Romanian. These languages are all similar to some extent, because they all go back to a common ancestor. In this case, we can draw a family tree to describe this situation, and here the protolanguage actually has a name that was recorded in history:



We should ask ourselves this question: Did Latin die out? The answer is that Latin did not die out in the same way that some languages have died out. Some languages die out because their speakers die out. The Tasmanian Aborigines, for instance, were badly affected by the diseases introduced by Europeans in the early 1800s, and many died. Many who did not die from disease were shot or poisoned by the Europeans. The last fully-descended Tasmanians died in the 1870s and 1880s, and knowledge of their languages died with them. (Contrary to popular belief, the Tasmanians did not become extinct. There are several thousand people in Tasmania today of partly Aboriginal descent who proudly identify themselves as Aboriginal Tasmanians, though their language is English.)

Other languages die out not because their speakers die out, but because they abandon their own language. Sometimes people abandon their own language as a result of having been forced to do so, while at other times people make the choice to switch to another language. In some parts of Australia and North America, for example, Aboriginal people were gathered together and the children were separated from their parents in dormitories and punished by missionaries or government officers if they were caught speaking anything other than English. The result is that many of these languages have disappeared, and the descendants of the original speakers now use only English.

There are parts of Papua New Guinea today, most notably in the area of the Sepik River, where parents are coming more and more to speak to their children in the national lingua franca, Tok Pisin, rather than their local vernacular. Some people have predicted that, within a generation or two, some of these vernaculars could be close to extinction, though in these cases the speakers are not being forced to give up their language. In these cases, there have been no movements of outsiders into these communities. People are making their own subconscious choice to switch from one language to another because Tok Pisin is associated with modernity and development, whereas the vernaculars are associated with tradition and backwardness.

But neither of these situations is true for Latin. Latin is not a dead language in the same sense that Tasmanian Aboriginal languages are dead. A protolanguage can in some ways be compared to a baby. A baby changes over time and becomes a child, then a teenager, and then an adult, and finally an old person. A baby does not die and then become a child, and so on. Similarly, Latin did not die and “become” French. Latin simply changed gradually so that it came to look like a different language, and today we call that language ‘French’. The name ‘Latin’ was not lost, either, as there is a little-known language spoken in Europe that is called ‘Ladin’. This is the modern form in that particular language of the old word ‘Latin’. One of the four official languages of Switzerland is also known as ‘Romansh’, which is a modern derivative of ‘Roman’. (Even further from Rome is Romania, but the Romanians also speak a language that is derived from Latin, and they have retained the original name of the Roman people who spoke Latin as the name of their language today.)

The changes between Latin and French (and Romansh and Romanian) were gradual. There was no moment when people suddenly realized that they were speaking French instead of Latin, in the same way that there is no single moment when a baby becomes a child or when a child becomes a teenager. After enough changes had taken place, people who compared the way they spoke with the older written forms of Latin could see that changes had occurred. But this is like looking at a photograph of ourselves taken when we were younger. We may look very different, but the person that we can see is definitely not dead!²

French and Romanian and Romansh have not stopped changing, either. The change continues into the present. French may well turn out to be the ancestor language from which a whole future family of languages is derived. So, too, may English, Bahasa Indonesia, Tolai, Paamese, Fijian, or Māori.

1.2 HOW AND WHY DO LANGUAGES CHANGE?

Our discussion in the preceding section supposes that languages change. But you might be wondering how we know that.

One way we know that languages change is that we have written records that show that stages of the language were different. If we pick up a book from 1400, or even 1700 or 1900, we can see differences in the language. The older the book, the more different the language. While most people can follow Shakespeare without too much difficulty (apart from some words that are no longer used), the same cannot be said of Chaucer or the Gawain poet, even when differences in spelling conventions are taken into account.

There aren't all that many places in the world where we can see that one language has split into several, but there are some. A few are in Europe: we have lots of documents for Latin, as well as historical information about the speakers of Latin who spread it. Another area where we have some record of diversification is in India, where we have long histories of records in languages from several families (going back about 3,000 years in the case of Sanskrit).

Another way is that we can see change happening at low levels, in vocabulary and sounds, and sometimes in other areas of language, too. For example, if you listen to an old recording (e.g., from a radio broadcast from the 1920s), you know immediately that the recording isn't a recent one. Part of that is because of the quality of the recording, but there are cues in the person's voice and the words they use that signal that the recording is old. Another example is in dialects. We know that most of the settlers who came to Australia were from southern England, but Australian English doesn't sound anything like any of the varieties of English spoken in the U.K. Something must have changed.

The fact of language change brings up another question: Why do languages change? Humans are creative creatures and they are constantly thinking up new words and new expressions. New technology is created (like telephones, computers, radar, and so on), and so we need names for these new things. New slang comes and goes, and so it isn't at all surprising that words should change over time. But there is more to language change than new words. Many reasons for change (some better than others) have been advanced over the years. Let's look at some of them.

1.2.1 Anatomy and Ethnic Character

In the nineteenth century, some scholars attempted to find an anatomical explanation for language change, concentrating in particular on sound change. At that time, cultural differences were often assumed to be related to anatomical differences, and different ways of thinking and behaving were often said to reflect the superior or inferior intellects of different peoples. (Such views, of course, are now regarded as racist nonsense, and I am only mentioning them here in the interest of historical accuracy.) Following the same line of thinking, change in language, and especially sound change, was sometimes related to cultural differences between peoples.

For instance, there are two sets of sound changes in some of the Germanic languages of northern Europe. The so-called First Sound Shift took place in the entire area in which the Germanic languages were spoken, while the more far-reaching Second Sound Shift took place only in the southern area (where German itself was spoken). A famous linguist in the nineteenth century, Jakob Grimm, tried to explain this:

It may be reckoned as evidence of the superior gentleness and moderation of the Gothic, Saxon, and Scandinavian tribes that they contented themselves with the first sound shift, whilst the wilder force of the southern Germans was impelled towards the second shift. (Grimm 1880: 306, quoted in Jespersen 2008: 858)

Such statements can easily become useful supports to politically sponsored racial beliefs and, indeed, have been used in this way in the past—for instance, in Nazi Germany.

It has also been suggested that there were significant differences between the languages of “civilized” people and those of “uncivilized” people with respect to language change. There was once a commonly held view among European scholars that modern civilization basically represented a corruption of a more pure and unspoiled form of human nature that was still to be found in the minds of what they came to call the ‘noble savage’. We can obviously question the kinds of presuppositions that are involved here, but that is beside the point for the moment. What is to the point is the fact that scholars at the time also attempted to find some kind of relationship between the fundamental nature of “primitive” languages as distinct from “civilized” languages. It was claimed that primitive languages contained more harsh, throaty sounds than civilized languages. Just as civilization was supposed to represent a degeneration of an original pure, natural state, so, too, was the supposed development of a preference for sounds produced further forward in the mouth. Such changes were equated with the laziness that characterized modern civilization. The ‘noble savage’, it was argued, maintained language in its more pure, guttural state!

Views such as these now warrant little further discussion. All that needs to be said is that it is quite impossible to relate any structural features of languages, whether they be phonetic features or grammatical features, to any differences in culture between two peoples. Such views represent pure racism.

1.2.2 Climate and Geography

In addition to some of the more bizarre nineteenth-century theories about language change that I have just discussed, there were some scholars who suggested that perhaps a harsh physical environment could produce harsh sounds in a language. What is meant by the “harshness” of a sound, of course, is not usually very clearly explained, though from the examples that people give, it appears that phonetic harshness involves the presence of many guttural sounds (i.e., glottal and uvular sounds) and the occurrence of many complex consonant clusters. The rugged terrain and harsh climate of the Caucasus Mountains in the former Soviet Union were sometimes said to have caused the languages of this region to develop such sounds.

It is not too difficult to prove that such views are nonsense. The Inuit of far northern Canada live in an environment that is as harsh as anywhere in the world, yet their phonetic system has been described by some scholars as “agreeable.” (You should note that it is just as unacceptable, however, to describe the phonetic system as “agreeable” as it is to say that it is “harsh.” Both represent nothing but value judgments.) Similarly, the Australian Aborigines of Central Australia live in a harsh environment of a different kind, yet they have a sound system that has been called “euphonic.” Evidently, what was meant by this was that these languages had relatively few consonant clusters, a fairly small number of phonemes, and relatively few guttural sounds.

1.2.3 Substratum

In the preceding two sections, I discussed some ideas for language change that have been thoroughly discredited. Let us now consider some more plausible causes of change.

The substratum theory of linguistic change involves the idea that if people migrate into an area and their language is acquired by the original inhabitants of the area, then any changes in the language can be put down to the influence of the original language. (In sec. 14.1, I discuss the question of how one language can influence another in its structure.) It is well known that a person’s first language will to some extent influence the way in which that person speaks a second language. We can all recognize foreign accents in our

own language. It is quite easy to tell whether someone is a native speaker of English or whether their first language is French, German, Chinese, or Samoan. If a large group of people switch from their original first language to a second language, they may carry over some features of their first language into their new language. This might take the form of words, or particular pronunciation features, or grammatical features. For example, Aboriginal English has a number of features of the sound systems of the Indigenous languages of Australia, such as no contrast between voiced sounds (e.g., [p] and [b]).

The problem with the substratum explanation of language change is that it is sometimes used to explain changes in languages where the supposed substratum language (or languages) have ceased to exist. The influence of the substratum in such cases can be neither proved nor disproved. One example of substratum influence that is often quoted involves the history of French. Before the time of the Roman Empire, what is now France was occupied by Celtic-speaking people (whose language was closely related to Welsh and Irish). France is now split into two major dialect areas, between the north and the south. Some scholars have suggested that this split corresponds to an earlier split in the original Celtic language and that these differences were carried over into the Latin they spoke when they switched languages. While this is a perfectly plausible theory, since the original Celtic language no longer survives in France, it can neither be proved nor disproved.³

1.2.4 Local Identification

The linguist Don Laycock once offered a different kind of explanation for why language change takes place, at least in some communities. In very small language communities, such as those found in Melanesia and in Aboriginal Australia, he suggested, languages may change simply to allow their speakers to distinguish things about their speech that are different from the speech of other people. People from linguistically very diverse areas, such as the Sepik in Papua New Guinea, have been reported as saying things like this: "It wouldn't be any good if we all spoke the same. We like to know where people come from." Linguistic diversity in this kind of situation is therefore a mark of identification for a community.

The urge for language to be used as a tool of identification can be particularly strong where the members of one ethnic group come to use the language of another ethnic group on a regular basis. What sometimes happens is that people will come up with their own distinctive vocabulary items and slang expressions as a way of signaling their distinct identity in what was originally a foreign language for them. Educated Papua New Guineans learned English

from their former Australian colonial “overlords,” yet nobody would mistake an ordinary Papua New Guinean speaking English for an Australian. For the most part, even the most fluent English-speaking Papua New Guineans do not want to sound like Australians. With this kind of attitude, people in Papua New Guinea have spontaneously come up with a number of colorful expressions which do not derive from Australian usage at all, such as the following:

| | |
|--|--|
| <i>That guy, he's really waterproof ia!</i> | ‘That guy doesn’t bathe very regularly.’ |
| <i>He's really service in greasing ladies.</i> | ‘He’s really good at chatting up women.’ |
| <i>Can I polish the floor at your place tonight?</i> | ‘Can I stay overnight on a mattress (or mat) on your floor tonight?’ |
| <i>She sixtied down the road.</i> | ‘She sped down the road.’ |

1.2.5 Functional Need

It is also true that some changes take place in language because a particular language *must* change in order to meet new demands that its speakers place on it. As the functional needs of a language change (i.e., the range of situations in which a language is used becomes wider), some aspects of the language may be lost, while others may be added. These kinds of pressures do not generally affect the phonology, or even the grammar, but they can have drastic effects on the vocabulary. Words referring to cultural concepts that have become irrelevant may be lost, while new words may flood into a language to express important new concepts. In chapter 11, I describe various aspects of lexical change arising from all sorts of different causes, so I will not go into this matter again at this point.

Some areas of lexical specialization develop for no particular reason, without any underlying cultural or environmental significance. When we compare the vocabulary of English with that of other languages, there are invariably areas of meaning in the other language that are encoded by single words for which we do not have single-word translation equivalents in English. For instance, in the Syle language of Vanuatu, we find words such as the following, though it would be difficult to find any particular cultural explanation for why they have words to express these meanings, while in English we don’t:

| | |
|-------------------|--|
| <i>elantvi</i> | ‘complain unjustifiably that something is insufficient or not good enough’ |
| <i>livinlivin</i> | ‘top of something that is teetering over an edge and is about to fall’ |
| <i>orvalei</i> | ‘touch something that is unpleasantly soft or mushy’ |

1.2.6 Simplification

Many of the sound changes that I describe in chapter 2 could be regarded as simplifying the production of sounds in one way or another. In dropping sounds, we are making words shorter, and therefore we need to exert less physical effort to produce them. The changes that come under the general heading of assimilation also clearly involve a change in the amount of effort that is needed to produce sounds as the degree of articulatory difference between sounds is reduced. Fusion, too, reduces the number of sounds in a word.

Despite the obvious appeal of this argument, there are also several problems with it. The first is that it is extremely difficult, perhaps even impossible, to define explicitly what we mean by SIMPLICITY in language. Simplicity is clearly a relative term. What is simple for speakers of one language may well be difficult for speakers of another. Kuman speakers in the Simbu Province of Papua New Guinea fused the two sounds [g] into a single velar lateral [L]. The principle of simplicity could be brought in as the causal factor, as this is an example of fusion. However, the velar lateral that results from this phonetic simplification is a sound that speakers of all other languages find almost impossible to produce to the satisfaction of Kuman speakers.

A second problem is that if all sound changes were to be explained away as being the result of simplification, we cannot explain why many changes do not take place. If it is easier to say [ʌŋkamd] than to say [ʌnkamd] for 'unkind', why don't all languages change [nk] to [ŋk]? Why do only some languages undergo this kind of simplification, and why only at some times? If language change were unidirectional, we should all be speaking basically the same kind of language now.

A third problem is that some sound changes clearly do not involve simplification. There is no way that the change called "metathesis" can be called simplification (though it does not make things any more complex, either). Exactly the same sounds are found before and after the change, and all that has been altered is the actual order in which the sounds occur. If phonetic fusion can be viewed as simplification, then surely phonetic unpacking must be just the opposite, as this creates two sounds from a single original sound.

Finally, simplification in one part of a language may end up creating complexities elsewhere in the system. For instance, the change known as "syncope" (i.e., the dropping of medial vowels) can be viewed as simplification in that it reduces the number of actual sounds in a word, but syncope often results in the creation of consonant clusters in languages that did not have them. While a particular word may end up being "simplified" as a result of syncope, the overall phonotactic structure of the language can be made much more complex. (By "phonotactics," I mean the statement of which phonemes can occur in what position in a word in a language, and which

other phonemes can occur next to them.) How can we say that a change from a CV syllable structure to a CCV syllable structure involves simplification, when the insertion of an epenthetic vowel between consonants to avoid CCV sequences is also called simplification?⁴

1.2.7 Structural Pressure

One explanation for sound change that has been put forward in recent years is the concept of STRUCTURAL PRESSURE. Linguists view languages as collections of units at various levels, and the units relate to each other in specific ways at each level in the system. Languages, therefore, operate in terms of systems. If a system becomes uneven, or if it has some kind of "gap," then (so the argument goes) a change is likely to take place as a way of filling that gap, so as to produce a neat system. For instance, imagine that a language has a five-vowel system:

| | |
|---|---|
| i | u |
| e | o |
| a | |

Now suppose that the vowel /e/ underwent a change such that it was unconditionally raised to /i/. This would result in the following system:

| | |
|---|---|
| i | u |
| | o |
| a | |

This is an unbalanced system, as the language has a contrast between a front and back vowel in the high vowels, but it has only a single mid vowel. There are many languages in the world that have three-vowel systems of the following type, but relatively few that have four-vowel systems such as that which I set out above:

| | |
|---|---|
| i | u |
| a | |

It would not be surprising to find that if a language had an unbalanced four-vowel system, speakers would then shift /o/ to /u/ to match the change that had produced the imbalance in the first place.

However, we cannot say that the pressure to fill gaps in systems like this is an overwhelming force in language change. The most that we can say is that languages that have gaps in their systems *tend* to fill them, but any attempt at a general explanation of sound change that contains the word

“tend” is of little value. Even a superficial examination of the world’s languages reveals that some of them have gaps in their systems, and there do not always seem to be changes taking place that would result in these gaps being plugged. In the Motu language of Papua New Guinea, for example, there are voiced and voiceless stops at the bilabial, alveolar, and velar points of articulation:

p t k
b d g

Motu also has nasals at the bilabial and alveolar points of articulation:

m n

However, there is no velar nasal in Motu. Although there is clearly a structural gap in the phonological system of the language, there is no indication that there are any changes taking place in the language that would result in the creation of a new phoneme that would occupy this empty slot in the phoneme inventory. Quite the opposite, in fact—we know from a comparison between Motu and closely related languages that it acquired this gap relatively recently, by unconditionally losing all of its velar nasals.

1.3 ATTITUDES TO LANGUAGE CHANGE

Since we are studying language change in this book, we should also think a bit about some of the common attitudes that people have toward the ways that languages change. As discussed in the preceding section, all languages are in a perpetual state of change. Sometimes, members of a particular society can observe changes that have taken place. In the case of written languages, people can see the language as it was written a number of generations ago, or even a number of centuries ago. In the case of unwritten languages, we obviously cannot observe how the language was spoken that far back in time, but very often people are able to recognize differences between the way the older people speak and the way the younger people speak.

It seems that in almost all societies, the attitudes that people have to language change are basically the same. People everywhere tend to say that the older form of a language is in some sense “better” than the form that is being used today. It is a common theme of language columns in newspapers, for example, that children are not learning to speak the language correctly. In most cases, if you ask people what they mean when they say these kinds of things, it turns out that they feel that the younger generation doesn’t use

some of the words that the older generation uses, that the younger generation speaks “sloppily,” or that they use slang.

In the preceding section, in the discussion of Saussure’s ideas, I said that forms in language are completely arbitrary. That is, there is no natural connection between a word and its meaning. This means that any sequence of sounds can express any meaning perfectly adequately, as long as members of the particular speech community agree to let those sounds represent that meaning. But people still like to insist that the earlier form of a language is better than the later form, and they still like to say that the newer ways of speaking and writing are “incorrect.” This applies to speakers of English, just as it does to any other language.

Language change is natural, and it is unstoppable, but that doesn’t stop people from attaching social judgments to various ways of talking. This should be unsurprising: after all, one of the functions of language is to index social information about the speaker and their identity. New markers come in, old ones go out, and items get adopted or rejected by different sectors of a society.

Reading Guide Questions

1. What statements did Ferdinand de Saussure make that influenced the course of linguistic science from his time on?
2. What is the significance of the discussion of stories told by people of different cultures in this chapter?
3. What possible explanations can we offer if we find that two languages express similar meanings by phonetically similar forms?
4. What do we mean when we say that two or more languages are genetically related?
5. What is a protolanguage?
6. What was the significance of the statement by Sir William Jones in 1786 about the relationship between Sanskrit, Latin, and Greek?
7. Does a protolanguage die out and then get replaced by its daughter languages? What, for example, is the nature of the relationship between Latin and Romanian?
8. How are people’s attitudes to language change and ideas of standard and nonstandard forms in language interrelated?
9. How do we know that language change is not caused by anatomical, cultural, or geographical factors?
10. Can a language be deliberately changed by members of a speech community?
11. To what extent is simplification a factor in causing language change to take place? What are some problems associated with this explanation of language change?
12. How might structural pressure cause a sound change to take place?

Exercises

1. What do you think is the importance to historical linguists of the fact that Sanskrit, Latin, and Greek were written languages? Would we have been able to make the same early advances in linguistic reconstruction if they were not?
2. Saussure and the modern linguists who followed him made a great deal of the arbitrary nature of language. How arbitrary is language? Examine the pairs of words below in a number of different languages. One word of the pair for each language means 'big' and the other means 'small'. Say which of each pair of words that you think means 'big' and which means 'small'. Compare the results across the class. Can you offer any explanation for what is going on? What do you think is the importance of such facts to the historical study of languages?

| | | |
|-------------------------|-----------------|-----------------|
| Paamese (Vanuatu) | <i>mari:te</i> | <i>titi:te</i> |
| Russian | <i>malenkij</i> | <i>bolfoj</i> |
| Fijian | <i>levu</i> | <i>lailai</i> |
| Bahasa Indonesia | <i>kətfil</i> | <i>bəsar</i> |
| Tagalog (Philippines) | <i>mali?it</i> | <i>malaki</i> |
| Kwaio (Solomon Islands) | <i>sika</i> | <i>ba?i</i> |
| Gumbaynggir (Australia) | <i>barwaj</i> | <i>junuj</i> |
| Samoan | <i>lapo?a</i> | <i>laiti:ti</i> |
| Dyirbal (Australia) | <i>midi</i> | <i>bulgan</i> |
| Lenakel (Vanuatu) | <i>ipwir</i> | <i>esua:s</i> |

(To find out which of these words mean 'big', refer to the answers at the end of these exercises.)

3. The word *tooth* in English has a long history in English writing, and it goes back to the same source as the German word *Zahn* [tsa:n] and the Dutch word *tand* [tant], indicating that these three languages are closely related. Latin also has a root for 'tooth' [dent-]. This is sufficiently different from the English, German, and Dutch forms to suggest that it is more distantly related to these languages. In written documents in English that are less than a few hundred years old, we start finding words such as *dental*, *dentist*, *trident* (a fork with three 'teeth'), and *denture*. What do you think this indicates about the historical relationship between Latin and English?
4. Look at the Lord's Prayer (King James version). Point out the expressions and constructions that would not normally be used in ordinary everyday speech today. Rewrite the prayer as it would be expressed in modern English. Why do you think people prefer to pray in an old-fashioned form of English that is sometimes hard to understand?

5. In his statement in 1786, Sir William Jones said that the various Indo-European languages that he was discussing must have “sprung from some common source, which perhaps no longer exists.” What did he mean by the comment that the original language perhaps no longer exists? Is he saying that the language became extinct? What sort of wording could you suggest that might more accurately reflect the actual situation?
6. For what sorts of reasons may a society give up its language and replace it with somebody else’s? Can you think of any examples from your own general knowledge where such a thing has happened or where it might happen in future?
7. Comment on Sir William Jones’s statement that Sanskrit, which resembles the protolanguage from which Latin and Greek were derived, “is of a wonderful nature, more perfect than the Greek, more copious than the Latin, and more exquisitely refined than either.”
8. French newspapers contain many English words, like *le football*, *le weekend*, *le camping*, and so on. Many speakers of French want to keep the language “pure” and to prevent the development of what they jokingly call *Français* (or *Frenghish*). There is even a government agency called the *Académie Française* (i.e., the French Academy), whose job it is to keep such words from appearing in the dictionary and to find good French words for all of these things. What comment would you make to members of this council?

Answers

The following are the words for ‘big’ from the forms that were given: Paamese *marite*, Russian *bolʃoj*, Fijian *levu*, Bahasa Indonesia *besar*, Tagalog *malaki*, Kwaio *baʔi*, Gumbaynggir *barwaj*, Samoan *lapoʔa*, Dyirbal *bulgan*, Lenakel *ipwir*.

Further Reading

- Anthony Arlotto, *Introduction to Historical Linguistics*, chapter 1, “The Scope of Comparative and Historical Linguistics,” pp. 1–10.
- Lyle Campbell, *Historical Linguistics*, chapter 1, “Introduction,” pp. 1–15.
- Jean Aitchison, *Language Change*, chapter 1, “The Ever-Whirling Wheel,” pp. 15–31; chapter 7, “The Reason Why,” pp. 111–28; chapter 8, “Doing What Comes Naturally,” pp. 129–43; chapter 9, “Repairing the Patterns,” pp. 144–55, chapter 10, “The Mad Hatter’s Tea Party,” pp. 156–69.
- Mary Haas, *The Prehistory of Languages*, chapter 1, “Introduction,” pp. 13–30.
- Nicholas Ostler, *Empires of the Word*, part I, “The Nature of Language History,” pp. 1–26.
- Claire Bowern, “Historical Linguistics.”

Hans Henrich Hock, *Principles of Historical Linguistics*, chapter 20, "Linguistic Change: Its Nature and Causes," pp. 627–62.

Hans Henrich Hock and Brian Joseph, *Language History, Language Change, and Language Relationship*, chapter 1, "Introduction" pp. 3–19.

Lyle Campbell and William Poser, *Language Classification, History* chapter 1. "How are Languages Shown to be Related to One Another?" pp. 1–12.

Diachronica is a journal specifically devoted to historical linguistics. It is worth looking through previous issues for topics of interest.

Types of Sound Change

While it may not be particularly surprising to learn that all languages change over time, you may be surprised to learn that different languages tend to change in remarkably similar ways. For instance, if you look at the history of the sound [p] in the Uradhi language of northern Queensland, you will find that it has undergone a change to [w] in the modern language:¹

Uradhi

- *pinta > winta 'arm'
- *pilu > wilu 'hip'
- *paṭa > waṭa 'bite'

Now, if you look at the history of the same sound [p] in a completely different language, one with no known historical connection with Uradhi, you will find that exactly the same change has taken place. Let us look at the Palauan language of Micronesia. Ignore all sounds for now except for those in bold type.

Palauan

- ***pa**qi > waʔ 'leg'
- ***pa**qit > waʔəð 'bitter'
- *qatə**p** > ʔaðow 'roof'

It is easy to find examples in other languages of the world of the sound [p] changing to [w].² But we also find repeated examples of [p] changing to other sounds—for instance, [f], or [b], or [v]. However, it would be very difficult to find an example of a language in which [p] had changed to [z], [l], or [e]. Let us now look at common (and likely) sound changes and distinguish these from unlikely sound changes. We will also classify the various kinds of attested sound changes in the languages of the world.

2.1 LENITION AND FORTITION

The first kind of sound change that I will talk about is lenition, or weakening. Many people would intuitively judge the sounds on the left below to be “stronger” in some way than those on the right:

Stronger Weaker

| | |
|---|---|
| p | b |
| p | f |
| f | h |
| x | h |
| b | w |
| v | w |
| a | ə |
| d | l |
| s | r |
| k | ʔ |

If you’ve studied phonology, you’ve probably heard of the term *SONORITY*. The generalizations that can be made regarding these correspondences are that voiced stops can be considered “stronger” than voiceless sounds. Similarly, stops rank higher than continuants in strength, consonants are higher than semivowels, oral sounds are higher in rank than glottal sounds, and front and back vowels rank higher than central vowels. These generalizations about the relative strength and weakness of sounds are equivalent to the “sonority hierarchy” in synchronic phonology. Sonority and strength is a complex combination of loudness of the sound, pitch, and the articulatory effort. This hierarchy is as follows, with the most sonorous sounds to the left and the least sonorous sounds to the right:³

a > e, ε > o > i, u > rhotics > laterals > nasals > voiced fricatives > voiceless fricatives > voiced stops > voiceless stops

The kinds of changes that I have just presented, therefore, tend to involve a shift from less sonorous to more sonorous sounds. This is called LENITION. It should be noted, however, that some of the commonly encountered changes listed here are difficult to account for purely in terms of loss of sonority, so the notion of phonetic weakening is a bit more complex than I have indicated.

When phonetic change takes place, it is often in the direction of a strong sound to a weak sound. That is, we would be more likely to find a change of [k] to [ʔ], for example, than the other way around, with [ʔ] becoming [k]. Changes of the reverse order are possible, of course, though less likely. These rarer sorts of sound changes could be referred to as strengthening (or FORTITION) to contrast them from lenition. For example, in the history of German, all stop consonants at the end of a word have become voiceless. There is no contrast between /g/ and /k/ at the end of a word in German; all the /g/ phonemes have become [k]. Note: // signals that the sound is a phoneme, whereas [] means it is a phone.

I will now give examples of phonetic lenition, or weakening, in different languages. The change of [b] and [p] to [f] in the Kara language of New Ireland (in Papua New Guinea) is one good example of lenition:

Kara

| | | | |
|---------|---|-------|-------------|
| *bulan | > | fulan | 'moon' |
| *tapine | > | tefin | 'woman' |
| *punti | > | fut | 'banana' |
| *topu | > | tuf | 'sugarcane' |

Similarly, the change from [p] to [w] in the Uradhi and Palauan examples given in the introduction to this chapter illustrate lenition.

One particular kind of lenition goes under the name of RHOTACISM. The term "rhotic" is often used to cover all types of *r* sounds (trills, flaps, glides, and so on), as distinct from all types of *l* sounds (which are together referred to as "laterals"). Laterals and rhotics collectively make up the phonetic class of LIQUIDS. The change known as rhotacism refers to the lenition of [s] or [z] to a rhotic between vowels. This kind of change took place in the history of Latin:

Latin

| | | | |
|-------------|---|----------|------------------|
| *ami:ko:som | > | amīcōrum | 'of the friends' |
| *genesis | > | generis | 'of the type' |
| *hono:sis | > | honōris | 'of the honor' |
| *flo:sis | > | flo:ris | 'of the flower' |

There is even evidence in the spelling of modern English that rhotacism has taken place in the history of this language. The plural form of the verb [wɔz] 'was' is [wɜːz] 'were' (though in many English dialects it is pronounced [wɛː] or [wɪ]). Assuming that the spelling of English more closely reflects an earlier pronunciation than the modern pronunciation, it seems that the final *e* of 'were' represents an earlier plural suffix, and that the root was probably something like [wase] or [wese] and there was later lenition of the [s] to [ʃ], to give [wæʃ] or [weʃ]. It is from this form that the modern form [wɛː] is derived, and some dialects have undergone another change and have dropped [ʃ] at the end of a syllable.

2.2 SOUND LOSS

A common kind of sound change that takes place in languages is the loss of one or more sounds. This can be viewed as an extreme case of lenition: the weakest a sound can be is not to exist at all. An example from modern English of a sound being lost altogether would be illustrated by the variable pronunciation of a word such as 'history'. While some people pronounce this as [hɪstɔːrɪ], other people simply say [hɪstɪ], dropping out the schwa vowel [ə]. We saw another example in the preceding section, when talking about the different pronunciations of 'were', some with [ʃ], some without. Here are some more examples of that change:

| Written Form | Australian English | American English |
|--------------|--------------------|------------------|
| card | kaːɪd | kaɪd |
| father | faːðə | faðəɪ |

It is common in languages of the world for sounds at the ends of words to be lost. In many languages of the Pacific, for example, final consonants are regularly dropped, as shown by the following changes that have taken place in the history of Fijian:⁴

Fijian

| | | | |
|--------|---|------|-----------|
| *niuR | > | niu | 'coconut' |
| *taŋis | > | taŋi | 'cry' |
| *ikan | > | ika | 'fish' |
| *bulan | > | vula | 'moon' |
| *tasik | > | taði | 'sea' |
| *lajaR | > | laða | 'sail' |
| *laŋit | > | laŋi | 'sky' |

This change has also taken place in the history of some other languages, including French. The consonants are still there in the written language, which was standardized at a time when the consonants were still pronounced:

| Written Form | Spoken Form | |
|--------------|-------------|--------------|
| chat | ʃa | ‘cat’ |
| fil | fi | ‘son’ |
| sont | sɔ̃ | ‘(they) are’ |

There are some kinds of sound loss that are covered by particular terms. These special terms are described and illustrated below.

2.2.1 Apheresis

Initial segments are sometimes dropped. We can refer to this as **APHERESIS**, pronounced [ə'fɛɪsɪs]. The following examples of apheresis come from the Angkamuthi language of Cape York Peninsula in Australia:

Angkamuthi

| | | |
|--------|--------|----------|
| *maji | > aji | ‘food’ |
| *nani | > ani | ‘ground’ |
| *ŋampu | > ampu | ‘tooth’ |
| *ŋukal | > uka: | ‘foot’ |
| *yantu | > antu | ‘canoe’ |
| *wapun | > apun | ‘head’ |

This is a common change in Australian languages, but it seems to be less frequently found in other parts of the world.

2.2.2 Apocope

APOCOPE, pronounced [ə'pɒkəpi], is the name you will come across in textbooks for the loss of word-final segments (that is, both vowels and consonants). This is a common change in languages all over the world, and examples are easy to find. For example, look at the following changes that have taken place in the history of the language of Southeast Ambrym in Vanuatu.⁵

Southeast Ambrym

| | | | |
|-------|---|-----|-------------|
| *utu | > | ut | 'lice' |
| *aŋo | > | aŋ | 'fly' |
| *asue | > | asu | 'rat' |
| *tohu | > | toh | 'sugarcane' |
| *hisi | > | his | 'banana' |
| *use | > | us | 'rain' |

You can see in this dataset that the reconstructed final vowel⁶ is lost in all the Southeast Ambrym words.

2.2.3 Syncope

This term **SYNCOPE**, pronounced [sɪŋkəpi], refers to a process similar to apocope, but syncope refers to the loss of segments in the middle of words, not the ends. It is syncope that often produces consonant clusters in languages that did not formerly have them when medial vowels are lost. You've already seen an example of syncope in our preceding discussion of history (sec. 2.2). Another example is from the Kiput language and the subgroup it belongs to, Proto-North Sarawak. The following data and reconstructions are from Robert Blust (2002). In this language, at some point in its history, schwa sounds were deleted in the environment CV__CV. There were subsequent later changes which do not concern us here.

Proto-North Sarawak

| | | | |
|---------|---|--------|----------------------|
| *eledaw | > | *eldaw | |
| *baqeRu | > | *baqRu | 'new' |
| *eRezan | > | *eRzan | 'notched log ladder' |

We have examples of sporadic syncope in English. For example, *sprite* and *spirit* go back to the same word (a borrowing from Latin *spiritus*). The first shows syncope (and subsequent diphthongization of [i:] to [ai]). The second is a **DOUBLET** without the syncope.

2.2.4 Cluster Reduction

When consonants come together in a word without any vowels between them, we call them **CONSONANT CLUSTERS**. Often, such clusters are reduced

by deleting one (or more) of the consonants. This is one kind of change that has taken place word-finally in English words ending in [mb] and [ŋg], such as *bomb* and *long*, where the spelling reflects the earlier pronunciation, though the modern pronunciations are [bɒm] and [lɒŋ]. This change is still spreading in English, as word-final stops in clusters of [nd] are now being lost. Words such as *hand* are often pronounced as [hæn] rather than [hænd], especially when there is a following consonant. Thus, *handgrip* is frequently pronounced by many people as [hæŋgrip], not as [hændgrip].

Cluster reduction has also occurred in the middle of many words in English. Although the word *government* is derived from the root *govern* with the following suffix *-ment*, the resulting cluster [nm] is normally reduced simply to [m]. So, instead of saying [gʌvənmənt], we normally just say [gʌvəmənt]. For many people, this is further reduced by syncope to [gʌvmənt], and consonant cluster reduction sometimes again applies to produce [gʌmən]!

2.2.5 Haplology

Haplology is a kind of change that is rare and tends to be fairly sporadic in its application. This term refers to the loss of an entire syllable, when that syllable is found next to another identical, or at least very similar, syllable. For some reason, people find it difficult to pronounce sounds when they are near other sounds that are identical or very similar. This is why people so easily make mistakes when they try to say tongue-twisters such as *She sells sea shells by the sea shore* very quickly.

Haplology is the process that is involved when we pronounce the word *library* as [laɪbrɪ] instead of [laɪbrɪəri]. The word *England* [ɪŋglənd] was originally *Anglaland*, meaning the land of the Angles. (The Angles were a group of people who settled in Britain over 1,000 years ago, bringing with them the ancestor of the modern English language.) The two last syllables in *Anglaland* were reduced by this process of haplology, and now we have only one *l* in the name *England* as a result.⁷

2.3 SOUND ADDITION

While lenition, particularly the total loss of sounds, is a common kind of sound change, you will also find that sounds are sometimes added rather than dropped. On the whole, however, sound addition is rather rare, but there are some environments (contexts) where it is quite common. In modern English, you can see evidence of this kind of change taking place when we hear people saying [sʌmpθɪŋk] instead of the more widespread [sʌmθɪŋ] for 'something'. There are also examples such as [nɒp] 'nope' and [jɛp] 'yep' instead of [nɒʊ] 'no' and [jɛə] 'yeah', in which a final [p] has been added.

Sound addition often takes place at the end of words with final consonants, where many languages add a vowel. Many languages tend to have a syllable structure of consonant plus vowel (represented as CV), allowing no consonant clusters and having all words ending in vowels. If a language adds a vowel to all words ending in a consonant, then it is moving in the direction of this kind of syllable structure. So, for instance, when words in Māori are borrowed from English, vowels are always added after consonants at the end of the word and the words conform to the syllable structure of other Māori words. Here are some examples from Māori. A similar process occurs in loan words in Japanese.

Māori

| | |
|--------|---------|
| ka:fe | 'calf' |
| ko:ti | 'court' |
| korofa | 'golf' |
| kuki | 'cook' |
| mapi | 'map' |
| miraka | 'milk' |
| raiti | 'light' |

There are lots of examples of this type of change in loanwords in the world's languages, but it's also found in words that aren't borrowed.

If you have two related languages, one of which has final vowels and the other doesn't, it can be hard to tell whether one language has undergone apocope (Sec. 2.2.2) or whether the other language has added vowels to the ends of words. One way to tell which change has happened is to look at the types of vowels that occur at the end of the word. Can you predict what vowel it will be? Is it always /a/? Is it always the same as the second-last vowel? If so, that is good evidence that there has been an addition of vowels, not a subtraction. (It's not always sufficient evidence, and in some cases it might be misleading, though.)

Some kinds of sound addition are known by specific names in the literature of historical linguistics. These terms, with examples of the process that they refer to, are presented below.

2.3.1 Excrescence

EXCRESCENCE refers to the process by which a consonant is added between two other consonants in a word. Although this change operates against the general tendency in languages to produce consonant plus vowel syllable structures, in that it creates even longer consonant clusters, it is nevertheless

a fairly common kind of change. The insertion of [p] in the middle of the cluster [mθ] in the word, something that I mentioned earlier, is an example of excrescence. Excrescence has also taken place in other words in the history of English, and the added consonant is now even represented in the spelling system:

English

- *æ:mtig > ɛmpti 'empty'
 *θymle > θɪmbl 'thimble'

The excrescent stop that is inserted in these examples has the same point of articulation (or is homorganic with) the preceding nasal in all of these examples. The stop is added to close off the velum (which is open during the production of the nasal) before going on to produce the following nonnasal sound (i.e., a stop or a liquid). This is a very common change. Another example is the Spanish word for *hombre* 'man', from Latin *hominem* (the [b] in the Spanish form is excrescent).

A change known as OCCLUSIVIZATION has occurred in Cypriot Arabic. In this language, clusters of certain consonants and [j] have developed an excrescent [k]. The following data are from Borg (1985:21):

Cypriot Arabic

- *pjara > pkjara 'wells'
 *safje > safkje 'ash water'
 *θjep > θkjep 'water'
 *mejje > mefkje 'walking (fem.)'

2.3.2 Epenthesis or Anaptyxis

The term EPENTHESIS (or anaptyxis)⁸ is used to describe the change by which a vowel is added in the middle of a word to break up two consonants in a cluster. This change therefore produces syllables of the structure CV (i.e., consonant plus vowel), again illustrating the common tendency for languages to avoid consonant clusters and final consonants. Speakers of some varieties of English often insert an epenthetic schwa [ə] between the final consonants of the word [film] 'film', to produce [fɪləm].

Epenthesis has also taken place in the history of Slavic languages. In the following Ukrainian words, an epenthetic vowel has been inserted after the liquid (data are from Blevins and Garrett 1998:522):

Ukrainian

| | | | |
|--------|---|--------|--------|
| *dervo | > | dévevo | 'tree' |
| *soldŭ | > | sólod | 'malt' |
| *gordŭ | > | hórod | 'city' |
| *melko | > | molokó | 'milk' |

2.3.3 Prothesis

PROTHESIS is another term used to refer to a particular type of sound addition—the addition of a sound at the beginning of a word. In the Dravidian language Kannada, for example, the words *ondu* 'one' and *eradu* 'two' have acquired prothetic consonants and are pronounced as [wɔndu] and [jeradu], respectively.⁹ In the Motu language of Papua New Guinea, when a word began with an [a], a prothetic [l] was added before it, as shown by the following examples:

Motu

| | | | |
|-------|---|------|-----------------|
| *api | > | lahi | 'fire' |
| *asaŋ | > | lada | 'gills of fish' |
| *au | > | lau | 'I, me' |

2.4 METATHESIS

The change known as METATHESIS [mɛ'tæθəsis] is a fairly uncommon kind of change. It does not involve either the loss or addition of sounds or a change in the appearance of a particular sound; it is simply a change in the order of the sounds. Someone mispronouncing the word *relevant* as "revelant" is an example of spontaneous metathesis.

Metathesis has taken place in the history of some English words, and the changed form has become accepted as the standard. The English word *bird* [bɜ:d] was originally pronounced as [bɪd]. This then became [bɪd] by metathesis, and this is the form that we still represent in our spelling system. Of course, the sounds [ɪ] have undergone further changes in some dialects to become [ɜ:]. In some dialects of English, such as American, Scottish, and Irish English, the original [ɪ] is still clearly pronounced, although other changes to the vowel have occurred.

Metathesis doesn't affect all sounds equally. It is particularly common with liquids (that is, *l* and *r* sounds).

Although metathesis is a rare sort of change, generally occurring in only one or two words in a language, there are still some cases of regular metathesis. In the Ilokano language of the Philippines, for example, there has been fairly consistent switching of word-final [s] and initial [t], as shown by the following comparisons with Tagalog, the national language of the Philippines (which reflects the original situation):

| Tagalog | Ilokano | |
|---------|---------|----------|
| tajis | sa:ɲit | 'cry' |
| tubus | subut | 'redeem' |
| tigis | si:git | 'decant' |
| tamis | samqit | 'sweet' |

2.5 FUSION, FISSION, AND BREAKING

2.5.1 Fusion

Phonetic fusion is a fairly frequent kind of sound change, in which two originally separate sounds become a single sound. The resulting single sound carries some of the features of both of the original sounds. This is known as FUSION.

Before I go on to give examples of fusion, I need to clarify what is meant by the term FEATURE. "Feature" is a technical term as well as a regular English word. All sounds can be viewed as being made up of a number of particular features, which determine different aspects of the nature of the sound. The sound [m], for instance, contains the following features (among others):¹⁰

[m]

[+ consonantal]

[+ voiced]

[+ labial]

[+ nasal]

The sound [a], on the other hand, contains the following features:

[a]

[- consonantal]

[+ voiced]

[+ low]

When two sounds are changed to become one in the process of fusion, some of the features of one sound and some of the features of the other sound are taken and a new sound is produced that is different from both, yet which also shares some features of both of the original sounds. I take an example of a change of this type from French. French developed a set of nasalized vowels from a former sequence of vowel + nasal. The actual changes involve several stages, because in addition to the nasalization, there was a change in the height of the vowel. The sequences are as follow:

*yn > *ȳ > *õ̃ > œ (e.g., *yn > œ̃ 'one')

*ɔn > õ̃ (e.g., *bɔn > õ̃ 'good')

*in > *ĩ > ĩ̃ (e.g., *vin > vĩ̃ 'wine')

*an > ā̃ (e.g., *blank > blā̃)

(The symbol ~ is known as a TILDE and is placed over the vowel to indicate that the vowel is nasalized, with the air coming out through the nasal passage as well as through the mouth.) The generalization we can make here is that

*Vowel + nasal > nasalized vowel

Expressing this in terms of features, we can say that the [– consonantal] feature of the first sound has been kept, while the [+ nasal] feature of the second sound has been kept, and a single new sound combining both features has been created:

[– consonantal]

[+ nasal]

A second example of fusion can be quoted from the Attic Greek (the dialects of Ancient Greek spoken in and around Athens). Examine the data below:¹¹

Attic Greek

| | | | |
|-------------------------|---|----------|-----------|
| *g ^w ous | > | hous | 'cow' |
| *g ^w atis | > | basis | 'going' |
| *g ^w asileus | > | basileus | 'king' |
| *leik ^w o: | > | leipɔ: | 'I leave' |
| *jek ^w ar | > | hɛ:par | 'liver' |

In the pre-Greek forms, there was a [g] or a [k] with the feature specification of velar stops. These were followed by a [w], which had the

feature specification for a semivowel with lip rounding. In the Greek fused forms, we find that the stop feature of the first sound has been taken along with the bilabial feature of the second sound to produce a bilabial stop. Thus, when there was an original voiced stop as in [g^w], the fused sound became the voiced bilabial stop [b], and when there was an original voiceless stop as in [k^w], the fused sound became the corresponding voiceless bilabial stop, [p].

A particular type of phonological fusion can be referred to as COMPENSATORY LENGTHENING. This kind of sound change is illustrated by the following forms from Old Irish:¹²

Old Irish

| | | | |
|-----------|---|--------|---------------------------|
| *magl- | > | ma:l | 'prince' |
| *kenetlo- | > | kene:l | 'gender' |
| *etno- | > | e:n | 'bird' |
| *ag-mo- | > | a:m | 'a moving back and forth' |

What has happened here is that a consonant has been lost and "in compensation" for this loss, a vowel has been lengthened. If we introduce the idea of phoneme space as a feature of a sound, we can treat this kind of change as another type of fusion. If each phoneme carries, among its collection of features, a phoneme space (i.e., the actual space it occupies in a word), then we could say that all features except this single feature of phoneme space can be lost and only this one feature is fused with the features of the preceding sound. This new sound therefore contains two features of phoneme space. This is reflected in the change in the preceding examples from a short vowel (i.e., one space) to a long vowel (i.e., two spaces).

2.5.2 Unpacking or Fission

FISSION is a phonetic process that is just the opposite of phonetic fusion. From a single original sound, a sequence of two sounds may develop, each with some of the features of the original sound. We saw earlier that, in French, vowels followed by nasal consonants underwent fusion to become nasalized vowels. It is also possible to find examples of languages in which the reverse kind of change takes place. In Bislama (the variety of Melanesian Pidgin spoken in Vanuatu), words of French origin that contain nasal vowels are incorporated into the language by unpacking the vowel features and the nasal features to produce sequences of plain vowels followed by the nasal consonant [ŋ]. Thus the French word for 'truck', *camion* (IPA [kamjō]), is borrowed as [kamioŋ], with [oŋ] instead of a nasal vowel.

Another example of fission has occurred in the Native American language Yurok (Blevins 2003, §5), where glottalized consonants such as $ʔm$, $ʔn$ and $ʔr$ (among others) are syllabified as clusters of glottal stop + m , n , or r :

keʔ.mow 'food'
 tʃpeʔ.roj.okʔ 'I listen'

2.5.3 Vowel Breaking

In the change known as **VOWEL BREAKING**, a single vowel changes to become a **DIPHTHONG**, with the original vowel remaining the same but with a glide of some kind being added, either before or after it. When a glide is added before the vowel, we call this an “on-glide”, but if a glide is added after the vowel, we refer to this as an “off-glide”. One of the more noticeable features of some varieties of American English is broken vowels. What is pronounced in most dialects of English as [bæd] ‘bad’ is pronounced by some Americans as [bæəd], or as [bæɪd], with an off-glide. One of the distinguishing features of Barbadian English in the West Indies is the palatal on-glide before the vowel [æ]. Instead of pronouncing [kæt] ‘cat’, people from Barbados say [kjæt].¹³

Vowel breaking is fairly common in the languages of the world. A good example of a language apart from American English that has undergone regular vowel breaking is the Kairiru language spoken on an island near Wewak in Papua New Guinea:

Kairiru

| | | | |
|--------|---|------|------------|
| *pale | > | pial | ‘house’ |
| *manuk | > | mian | ‘bird’ |
| *ɲamuk | > | niam | ‘mosquito’ |
| *ranum | > | rian | ‘water’ |
| *lako | > | liak | ‘go’ |

(Note that in these examples there is also evidence of apocope, or the loss of the final vowels.)

2.6 ASSIMILATION

Many sound changes can be viewed as being due to the influence of one sound on another. When one sound causes another sound to change so that the two

sounds end up being more similar to each other in some way, we call this ASSIMILATION. Since assimilation is by far the most common kind of sound change, I present here a fairly detailed discussion of the various subtypes of assimilation, along with numerous examples.

Before I do that, I need to define the concept of PHONETIC SIMILARITY. Two sounds can be described as being phonetically more similar to each other after a sound change if those two sounds have more phonetic features in common than they did before the change took place. If a sound change results in an increase in the number of shared features, then we can say that assimilation has taken place.

As an example, take a word that contains a consonant cluster of the form [np] in an imaginary language. The two sounds in this cluster each have the following phonetic features:

| | |
|--------------|--------------|
| [n] | [p] |
| [+ voiced] | [– voiced] |
| [+ coronal] | [+ labial] |
| [+ sonorant] | [– sonorant] |

We could assimilate one, or two, or all of the features of one of these two sounds in the direction of the other. For instance, the [n] could lose its nasal feature [+ sonorant] and replace it with the stop feature of the [p] next to it. This change would have the following effect:

*np > dp

If, instead of assimilating the nasal feature to the following stop, we were to assimilate the place of articulation of the nasal to that of the following stop, we would have the following change:

*np > mp

Finally, if the voiced feature of the nasal were to acquire the voicelessness of the following stop, this change would show up as follows:

*np > ṅp

(The [ṅ] with a circle beneath it represents a voiceless alveolar nasal. Such a sound is rare in the world's languages, and the last change that I referred to would be less likely to occur than the previous two changes.)

The changes that I have just presented all involve the assimilation of only a single feature. It is, of course, possible to assimilate two features at a time, as in the following examples:

*np > bp

(keeping only the voicing of the nasal, but assimilating it to the following sound, both in its manner of articulation and its place of articulation)

*np > tp

(keeping only the alveolar place of articulation of the nasal, but assimilating it to the following [p] both in its voicelessness and in its manner of articulation)

*np > mp

(keeping only the nasal feature, but assimilating it to the [p] in its voicelessness, as well as in its place of articulation)

All of these changes are examples of **PARTIAL ASSIMILATION**, because the changed sound always retains at least one of the original features by which it is distinguished from the unchanged sound. If all of the features are changed to match those of another sound, then the two sounds end up being identical and we produce a **GEMINATE** (or phonetically double) sound. When assimilation produces geminate sounds in this way, we can speak of **TOTAL ASSIMILATION**. In the case of the cluster [np], an example of total assimilation would be a change of [*np] to [pp].

There is yet another dimension that we should discuss regarding this kind of assimilation. All of the examples that I have just presented are what are called **REGRESSIVE ASSIMILATION**. This means that the “force” of the change operates “backward” in the word—that is, from right to left. It is the features of the following [p] in all of the examples given that influence the features of the preceding [n], which is why we call this regressive assimilation. This kind of assimilation can be represented in the following way:

A ← B

(The symbol ← indicates the direction of the influence of one sound over the other.)

There is, of course, a second possibility, in which the direction of the change is reversed, and it is the preceding sound that exerts its influence over the sound that follows it. This kind of situation could be represented by the symbol facing forward in the word like this:

A ⇒ B

Such a situation, in which the features of a following sound are changed to match those of a preceding sound, is called **PROGRESSIVE ASSIMILATION**. Of the two types of assimilation, regressive assimilation is by far the more commonly encountered in the world’s languages, although progressive assimilation does also occur.

If we take the same cluster [np] and this time treat the [n] as the influencing sound rather than the [p] as before, we find that the following changes can all be regarded as examples of partial progressive assimilation:

- *np > nb (with assimilation of voicing)
- *np > nt (with assimilation of place of articulation)
- *np > nm (with assimilation of manner of articulation)
- *np > n̥p (keeping only the voiceless feature of the [p])
- *np > n̠p (keeping only the bilabial feature of the [p])
- *np > nd (keeping only the stop feature of the [p])

Progressive assimilation can be total, as well as partial, so there is also the following final possibility:

- *np > nn (keeping none of the features of [p])

With two sounds that have only three different features each, you can see that there are 14 possible changes that can all be classed as assimilation. This concept therefore covers a wide range of possible sound changes, and as I said at the beginning of this section, most sound changes that take place in the languages of the world involve assimilation in one way or another.

Rather than continuing to talk about assimilation in the abstract as I have been doing, I now give concrete examples to show how this process works. Let us look at the history of some words in the Karnic languages of the Lake Eyre Basin in Australia (data are from Austin 1990):

| | Yawarrawarrka | Yandruwandha | Diyari |
|------------|---------------|--------------|--------|
| 'language' | patpa | parlpa | — |
| 'eyebrow' | pitpa | pirlpa | pirlpa |
| 'hole' | witpa | — | wirlpa |
| 'whistle' | witpi | — | wirlpi |

Yandruwandha and Diyari show the unchanged forms of the words, while in Yawarrawarrka the lateral has become a stop in these clusters. That is, the stop feature from the [p] has been copied by the previous consonant.

As I have already mentioned, progressive assimilation is much less common than regressive assimilation and examples are much harder to find. However, in the history of Icelandic, the following are examples of very regular total progressive assimilation:

Icelandic

| | | | |
|---------|---|-------|------------|
| *fīndan | > | finna | 'find' |
| *gulθ | > | gull | 'gold' |
| *halθ | > | hall | 'inclined' |
| *munθ | > | mun | 'mouth' |
| *unθan | > | unna | 'love' |

Examples of partial assimilation are more common than examples of complete assimilation. Partial assimilation can involve a wide range of possibilities, as we have already seen, with the changes involving the place of articulation (including the high, low, front, and back features of vowels, as well as the features referring to the place of articulation of consonants), manner of articulation (whether stop, fricative, nasal, lateral, and so on), and voicing (whether voiced or voiceless). Assimilation may also involve any combination of these various features.

Assimilation of place of articulation is a common change. You can see the results of this change in modern English with the varying forms of the negative prefix [ɪn-] 'in-'. This is normally pronounced with the variant [ɪm-] before bilabial consonants, [ɪŋ-] before velars, and [ɪn-] before all other sounds (including vowels), as in the following:

| | |
|---------------|-----------------|
| m-dəvɪzəbəl | 'indivisible' |
| m-bælns | 'imbalance' |
| ɪŋ-kənsɪdərət | 'inconsiderate' |
| m-ədmsɪsəbəl | 'inadmissible' |

The [n] has assimilated in its place of articulation to the following consonant—that is, the alveolar feature has been replaced with the feature for the place of articulation of the following sound when the next sound is bilabial or velar.

The change that is known as PALATALIZATION is also an assimilatory change. By this change, a nonpalatal sound (i.e., a dental, an alveolar, a velar, and so on) becomes a palatal sound, usually before a front vowel such as [i] or [e] or before the semivowel [j]. Sounds that we can class as palatal include the palatoalveolar affricates [tʃ] and [dʒ] and the sibilants [ʃ] and [ʒ] (as well as some other consonants that are less common). This change can be described as assimilatory because the palatal feature of the vowel (the fact that it is front rather than back) is transferred to the neighboring consonant.

One good example of palatalization is the change from [t] to [tʃ] before the vowel [i] in many dialects of Fijian. For example, where Standard Fijian

has [tinana] 'his/her mother', many of the local dialects have palatalized the initial consonant to produce [tʃinana]. There are examples of palatalization having taken place in the history of English, too. The velar stops [k] and [g] became palatalized to [tʃ] and [j], respectively, when there was a following front vowel, as shown by the following examples:

English

| | | | |
|---------|---|-------|----------------|
| *kinn | > | tʃin | 'chin' |
| *kɛ:si | > | tʃi:z | 'cheese' |
| *geldan | > | ji:ld | 'yield' |
| *gearn | > | ja:n | 'yarn, thread' |

(Note that the change of [g] to [j] probably involved palatalization of [g] to [dʒ] first, and then the [dʒ] underwent lenition to [j].)

Sometimes, a palatal that is produced as a result of this kind of assimilation can undergo lenition to become [s]. For example, in Motu in Papua New Guinea, [t] has shifted to [s] in a similar kind of palatalizing environment to that described above for Fijian, even though [s] is a postalveolar sound rather than a palatal sound. Note the following examples:

Motu

| | | | |
|--------|---|------|---------------|
| *tama | > | tama | 'father' |
| *taŋis | > | tai | 'cry' |
| *tubu | > | tubu | 'grandparent' |
| *topu | > | tohu | 'sugarcane' |
| *tolu | > | toi | 'three' |
| *tina | > | sina | 'mother' |
| *qate | > | ase | 'liver' |
| *mate | > | mase | 'die' |

In addition to assimilation involving changes in the place of articulation, changes in the manner of articulation of a sound to make two sounds phonetically more similar to each other are also common. In the Warluwarra language of northern Queensland (in Australia), Proto-Warluwaric *g has become [ɣ]. For example, the word for 'one' is *yaryulila* (a cognate word in Yanyuwa, a related language, is *yargu*). Proto-Warluwaric *milga 'side' is *milya* in Warluwarra. The velar stops in these examples have changed to become voiced fricatives at the same place of articulation. This can be viewed as the assimilation of two of the features of the original stops to the features of the surrounding segments.

Another common type of change that can also be viewed as a special kind of assimilation is the change called FINAL DEVOICING. Sounds at the end of a word, especially stops and fricatives (but sometimes also other sounds, even vowels), often change from being voiced to voiceless. In German, the devoicing of final stops has been very regular, for example:

German

| | | | |
|-------|---|------|--------|
| *ba:d | > | ba:t | ‘bath’ |
| *ta:g | > | ta:k | ‘day’ |
| *hund | > | hunt | ‘dog’ |
| *land | > | lant | ‘land’ |
| *ga:b | > | ga:p | ‘gave’ |

In a case like this, the voiced feature of the original sound is changed to voiceless to match the voicelessness of the following silence at the end of the word. This can also be thought of as a type of fortition.

There is a further aspect to assimilation that I have not yet touched on. This is the contrast between what we call “immediate assimilation” and “assimilation at a distance”. In the examples of assimilation that I have presented so far, it has always been a case of one sound being influenced by the sound either immediately preceding or immediately following it. These are, therefore, all examples of immediate assimilation.

In the case of assimilation at a distance, however, a sound is influenced by another sound not immediately to the left or the right of it, but further away in the word, perhaps even in another syllable altogether. In the Southern Highlands of Papua New Guinea, when speakers of the Huli language adopt the Tok Pisin word *piksa* ‘picture’ into their language, it is sometimes pronounced by older people as [kikiɖa] rather than [pikiɖa] as we might expect. What has happened is that the [p] of the first syllable has assimilated (at a distance) in place of articulation to the [k] of the second syllable. Another example of this is the English word *orang utan*, which is pronounced by many people as [ɔraŋ utɑŋ], with two velar nasals instead of an alveolar nasal.

Sometimes assimilation at a distance like this is a very regular feature of a language, and some type of assimilation may even apply over an entire word. When this happens, we call this HARMONY. Many languages have what we call VOWEL HARMONY, which means, basically, that there is assimilation of one (or more) features of one vowel to some (or all) of the other vowels in the same word. To see how this works synchronically, consider the following Turkish words:

| | | |
|-------|------|------------|
| tavuk | -lar | 'chickens' |
| ayı | -lar | 'bears' |
| ev | -ler | 'houses' |
| köpek | -ler | 'dogs' |

If the noun ends in a back vowel, the form of the plural ending is *-lar*, but if the stem ends in a front vowel, the plural is *-ler* instead.

Sometimes you will find harmony involving features other than just vowel features. In the Enggano language (spoken on an island off the coast of southern Sumatra in Indonesia) there has been a change that we refer to as NASAL HARMONY. In this language, all voiced stops in a word became homorganic nasals and all plain vowels became the corresponding nasal vowels following any nasal sound in a word. So:

Enggano

| | | | |
|-----------|---|----------|-------------|
| *honabu | > | honāmū | 'your wife' |
| *ehĕkua | > | ehĕkūā | 'seat' |
| *eū?ada?a | > | eū?ānā?ā | 'food' |

One special kind of vowel harmony goes under the name of UMLAUT. This term is most frequently used in Germanic languages to refer to the fronting of a back vowel or the raising of a low vowel under the influence of a front vowel in the following syllable. Often, the following high front vowel that caused the change to take place in the first place was then dropped in these languages (by apocope) or reduced to schwa. Thus, the new front vowel became the only way of marking the difference between some words. The irregular singular and plural pairs of words such as *fool/feet* in English are the result of such vowel harmony, or umlaut. The original singular form was [fo:t], and its plural was [fo:t-i]. The [o:] was later fronted to the front rounded vowel [ø:] under the influence of the following front vowel [-i] in the plural suffix, so the plural came to have the shape [fø:t-i]. Later, the vowel of the suffix was dropped, and the front rounded vowel of the root was unrounded to become [e:]. So, while the singular was [fo:t], the plural had become [fe:t]. It was this alternation between [fo:t] and [fe:t] that was the source of the modern irregular pair *fool/feet*. (This kind of umlaut in the history of English is described in more detail in Sec. 4.3.)

2.7 DISSIMILATION

Now that we have studied at length the concept of assimilation, it should be a relatively simple matter to grasp the concept of dissimilation. This process

is precisely the opposite to assimilation. Instead of making two sounds more like each other, DISSIMILATION means that one sound changes to become less like some other nearby sound. Dissimilation, therefore, reduces the number of shared phonetic features between two sounds.

I have already mentioned in this chapter the difficulty that we have with tongue-twisters: if you say these fast enough, you will sometimes find yourself dropping out sounds that are similar to each other when they occur frequently in the same sentence. Another thing that happens when we say tongue-twisters is that we tend to make sounds more distinct from nearby sounds than they are supposed to be. If you say *Peter Piper picked a peck of pickled peppers* frequently, the chances are that you will end up saying “peckers” instead of “peppers.” This would perhaps be partly a case of the [p] in the word ‘peppers’ assimilating at a distance to the [k] in words such as ‘picked’ and ‘peck’, but at the same time the [p] is probably dissimilating from the other [p] sounds that are found near it in the same word.

I will mention one very famous example of dissimilation here, because it is frequently encountered in textbooks of historical linguistics, where it is often referred to as GRASSMANN’S LAW. This sound change, first recognized in 1862 by the German scholar Hermann Grassmann, took place both in the ancient Sanskrit language in what is now India and in the ancient Greek language. In both languages, there was a phonemic contrast between aspirated and unaspirated stops. However, when two syllables followed each other and both contained aspirated stops, the first of these lost its aspiration and became unaspirated. So, in Sanskrit, the earlier form [*bho:dha] ‘bid’ became [bo:dha], and in Greek, the form [*phewtho] with the same meaning became [pewtho]. This is clearly a case of dissimilation at a distance.

An example of immediate dissimilation (rather than dissimilation at a distance) can be found in Afrikaans, the language of one of the two major tribes of Europeans in South Africa (the other being English speakers). Observe the following changes:

Afrikaans

| | | | |
|----------|---|--------|------------|
| *sxo:n | > | sko:n | ‘clean’ |
| *sxoudər | > | skouər | ‘shoulder’ |
| *sxælt | > | skælt | ‘debt’ |

In the original forms, there was a sequence of two fricative sounds: [s] and [x]. In Afrikaans, the fricative [x] changed to a stop at the same place of articulation—[k]—so that there would no longer be two fricatives next to each other. Thus, the [x] dissimilated in manner of articulation to [k] from the fricative [s].

2.8 TONE CHANGES

One area of sound change which often gets forgotten is change in nonsegmental aspects of phonology, such as tone. But languages change in this area, too, just as segmental sounds like consonants and vowels change. TONE LANGUAGES use the pitch of the vowel to signal differences in meaning. Tone languages are found in North America, Papua New Guinea, Africa, and Asia. There are two main ways in which tone arises in languages. The first is a type of reanalysis. The second is borrowing of the category of tone from a neighboring language.

In the first case, in order to understand how tone arises, you need to know a bit of phonetics. You know, of course, that when people speak, their voices have different pitches. Some people have naturally higher voices than others, and of course people modulate their voices when they speak. However, there are also differences in the pitches of vowels in certain contexts. Many of these are automatic; we don't pay attention to them because they are conditioned by the environment the sounds occur in. For example, the natural pitch of a vowel following a voiced consonant is lower than the natural pitch of a vowel following a voiceless consonant. Figure 2.1 shows the pitch of a vowel immediately following a [p] and a [b], and the pitch 100 ms following.

Now, remember from sec. 2.5 that features can sometimes be reinterpreted. We've seen examples already where a feature has been reinterpreted because of its surrounding environment. In this case, what seems to have happened is that the automatic raising or lowering of a vowel's base pitch is reinterpreted as meaningful. If the triggering environment of voicing is wiped out, all that is left is the pitch difference on the vowel. What started as an automatic alternation is reinterpreted by subsequent generations as part of the phonology of the language.¹⁴

I mentioned that voiceless consonants can be reinterpreted as high tone when they precede the vowel in question and that voiced consonants produce low tone. But there is more to tone difference than this. In particular, many languages have CONTOUR TONES (such as tones that start high and fall to a low pitch, or tones that start low and the pitch rises over the syllable). The presence of a consonant in the coda of the syllable can also raise or lower pitch. In particular, coda glottal stops [ʔ] cause pitch raising (and therefore rising tones), and fricatives, particularly [h], cause falling tones.

Many tone systems are not purely pitch systems but also have particular phonation types associated with them. For example, one of the Vietnamese low tones occurs on vowels with breathy voice, and creaky voice often causes falling pitch. PHONATION TYPE, like initial consonant voicing, also plays a role in vowel pitch. In fact, some have argued that phonation type is more important in the creation of tone than voicing is.

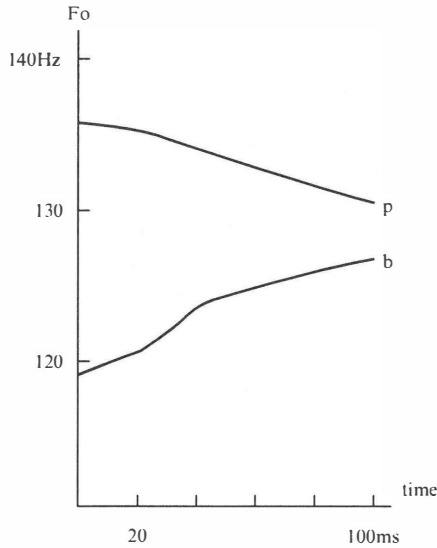


FIGURE 2.1 Average fundamental frequency values (Hz) of vowels following English stops (from Hombert et al. 1979: 39).

The second way that tone arises in languages is when they borrow the categories of tone from another language its speakers are in contact with. Tone occurs in areas. (I talk more about areas in chapter 14.) It is often argued that tones in Southeast Asian languages spread from Southern Chinese varieties into Tai-Kadai languages. However, the details of how a category like tone can be borrowed are complex. In some cases, lots of words are borrowed with their tones, which leads to the adoption of tone patterns in other vocabulary. In other cases, it appears that the patterns themselves are borrowed.

Tones may change over time. Like other types of sounds, they may merge and split, and these mergers and splits may be conditioned by environment or they may happen everywhere. For example, in Saigon Vietnamese, the low-falling and high-falling tones have merged into a single falling tone. (Thurgood [2002] and the references there provide more information about this subject.)

2.9 UNUSUAL SOUND CHANGES

In this chapter, I present a wide range of types of sound changes that you will come across in languages of the world. You have now seen examples of all of the most common types of change. However, there are numerous examples of sound changes in language that would appear, at first glance, not to obviously fit into any of the categories that I have set out above. For

instance, take the French word *cent* 'hundred', which is pronounced [sã]. This ultimately goes back to a form that can be reconstructed as [kɾɪtom] (with the first [m] being a syllabic nasal—that is, a nasal that can be stressed in the same way as a vowel).¹⁵ How can the change from [kɾɪtom] to [sã] possibly be described in terms of the types of changes that we have been looking at in this chapter?

The answer to this question comes in the observation that, while the differences between these two forms might appear to be immense (and therefore unlikely), we can usually reconstruct various intermediate steps between the two extreme forms that appear to represent quite reasonable sorts of changes. Let us imagine that the change from [kɾɪtom] to [sã] in fact took place through the following series of steps over a very long period of time:

*kɾɪtom > kemtom

(unpacking of features of syllabic and consonant to two separate sounds)

kemtom > kentom

(regressive assimilation of [m] to [t] in place of articulation) kentom > kent

(loss of final unstressed syllable)¹⁶

kent > cent

(palatalization of [k] to [ç] before front vowel)

cent > sent

(lenition of stop to fricative)

sent > sēt

(fusion of features of vowel and nasal to produce nasal vowel)

sēt > sāt

(lowering of vowel)

sāt > sã

(loss of final consonant)

Sometimes we find that an individual sound has changed in a rather unusual way. Although we should keep in mind the types of sound changes described in this chapter as being somehow more likely to occur than other kinds of sound change, students of languages will always come up against rare changes. For example, Proto-Algonquian *θ and *l fall together in some environments and can be reconstructed as *r to Proto-Eastern Algonquian (Goddard 1982: 21).¹⁷

For instance, in some languages—including Trukese—there have been regular changes of [t] to [w], and in the Mekeo language (spoken in the Central Province of Papua New Guinea), there has been a change of both [d] and [l] to the velar nasal [ŋ]. This latter change is illustrated by the following examples:

Mekeo

- *dua > ŋua 'two'
 *dau > ŋajau 'leaf'

How might we account for such changes? Again, it is possible to suggest a series of more reasonable intermediate stages that have left no trace. The Trukese change of [t] to [w] may have passed through the following stages, for example:

[t] > [θ] > [f] > [v] > [w]

Similarly, the Mekeo change of [d] and [l] to [ŋ] may have gone through the following steps:

[d] > [l] > [n] > [ŋ]

However, while in some cases there is evidence for intermediate stages in the change, in other cases there is no evidence for breaking down the change into intermediate steps.

Given a sufficient period of time, any sound can change into any other sound by a series of changes such as those discussed in this chapter. It is partly for this reason that the reconstruction of the history of languages by the method described in this volume has not really been able to go back further than about 10,000 years. Any changes beyond that time would probably be so great that, even if two languages were descended from a common ancestor, time would have almost completely hidden any trace of similarities that the languages may once have had.¹⁸

Reading Guide Questions

1. What is lenition?
2. What is rhotacism?
3. What is cluster reduction?
4. What is the difference between apocope and syncope?
5. What is the difference between haplology and metathesis?
6. What is the difference between excrescence and epenthesis?
7. What is the difference between aphaeresis and prothesis?
8. What is phonetic fusion?
9. What is meant by compensatory lengthening?
10. What is the difference between phonetic unpacking and vowel breaking?
11. How is assimilation different from dissimilation?
12. What is the difference between partial and complete assimilation?

13. What is the difference between assimilation at a distance and immediate assimilation?
14. What is palatalization, and how can this be viewed as assimilation?
15. What is final devoicing, and how can we view this as assimilation?
16. What is vowel or consonant harmony?
17. What is meant by the term “umlaut”?
18. What is Grassmann’s Law? What sort of sound change does this involve?
19. How does high tone develop?

Exercises

1. Some of the phonetic changes described in this chapter can be regarded as belonging to more than one of the named categories of changes. For instance, final devoicing was described in sec. 2.6 as a kind of assimilation, while devoicing in general was described in sec. 2.1 as lenition, or weakening. Can you find any other kinds of sound change that can be described under two different headings?
2. What do you think the spelling of the following words indicates about the phonetic history of English: lamb, sing, night, rough, stone, mate, tune, Christmas. Describe any changes that might have taken place in terms of the kinds of sound changes described in this chapter.
3. Many place names in England have spellings that do not reflect their actual pronunciations. From the following list, suggest the kinds of phonetic changes that may have taken place as suggested by the original spellings:

| | |
|--------------|----------|
| Cirencester | [sɪstə] |
| Salisbury | [səlzbɪ] |
| Barnoldswick | [bɑ:lɪk] |
| Leicester | [lɛstə] |
| Chiswick | [tʃɪzɪk] |
| Cholmondely | [tʃʌmli] |
| Gloucester | [glɒstə] |

4. Speakers of English for whom English is their first language pronounce the following words as shown:

| | |
|----------|-------------|
| society | [səsaɪəti] |
| social | [soʊʃəl] |
| taxation | [tækseɪʃən] |
| decision | [dəsɪʒən] |

Papua New Guineans speaking English frequently pronounce these words as [səsaɪti], [ʃoufəl], [tækfeɪʃən], and [dəʃɪʒən], respectively. What kind of phonetic changes do these pronunciations involve?

5. The following changes have taken place in Romanian. Should we describe these changes as phonetic unpacking or as vowel breaking? Why?

*pɔte > pwate 'he is able'
 *pɔrta > pwartə 'door'
 *nɔkte > nwapte 'night'
 *flore > flwarə 'flower'
 *ora > warə 'hour'
 *eska > jaskə 'bait'
 *ɛrba > jarbə 'grass'

6. The following changes took place in some dialects of Old English. Should we describe these as phonetic unpacking or as vowel breaking?

*kald > keald 'cold'
 *erða > eorða 'earth'
 *lirnjan > liornjan 'learn'
 *melkan > meolkan 'milk'

7. In the following data from the northern dialect of Paamese (Vanuatu), why do we say that assimilation has taken place? What particular kind of assimilation is involved?

*kail > keil 'they'
 *aim > eim 'house'
 *haih > heih 'pandanus'
 *auh > ouh 'yam'
 *sautin > soutin 'distant'
 *hauļu > houļu 'many'

8. In the following data from Toba Batak (Sumatra), what kind of assimilation has taken place?

*hentak > ottak 'knock'
 *kimpal > hippal 'lump of earth'
 *cintak > sittak 'draw sword'
 *ciŋkəp > sikkop 'enough'
 *pintu > pittu 'door'

9. In the following Italian data, what kind of assimilation has taken place?

- noktem > notte 'night'
- faktum > fatto 'done'
- ruptum > rotto 'broken'
- septem > sette 'seven'
- aptum > atto 'apt'
- somnus > sonno 'sleep'

10. In the following Banoni forms, there is evidence of more than one pattern of assimilation having taken place. What are these patterns?

- *manuk > manuya 'bird'
- *kulit > yuritsi 'skin sugarcane'
- *jalan > sanana 'road'
- *tañis > tañisi 'cry'
- *p^wekas > beɣasa 'feces'
- *boRok > boroyo 'pig'

11. Old English had a causative suffix of the form [-j] and an infinitive suffix of the form [-an], both of which have been lost in Modern English, and their original functions are now expressed in different ways. Examine the pair of words below from an earlier stage of English:

- drink-an 'to drink'
- drank-j-an 'to cause (someone) to drink'

The modern words *drink* and *drench*, respectively, evolved from these two words. What sort of change has been involved to derive the final consonant of *drench*?

12. In the Marshallese language of Micronesia, the following changes have taken place:

- *matapa > medan 'his/her eye'
- *dam^wapa > demwan 'his/her forehead'
- *masakit > metak 'pain'
- *madralsis > metal 'smooth'
- *sakaRu > tekaj 'reef'
- *madama > meram 'light'

How would you characterize the changes that have affected the vowels in Marshallese?

13. In Dataset 1, a series of sound changes in Palauan is presented. Try to classify these changes according to the types of sound change discussed in this chapter.
14. Examine the forms in Nganyaywana in Dataset 2. The original forms are given on the left. Try to classify the changes that have taken place.
15. Refer to the forms in Mbabaram in Dataset 3. Try to describe the kinds of changes that have taken place.
16. From the data in Yimas and Karawari given in Dataset 4, what kinds of changes would you say had taken place in each of these two languages?
17. Assume that in some language, the following sound changes took place. These changes all appear to be quite abnormal in that there is no simple change of features from one stage to the other. Can you suggest a succession of more reasonable sounding intermediate steps to account for these unusual results?

*b > h

*e > i

*k > r

*k > s

*p > w

*l > i

*k > h

*ɣ > ?

*s > ?

*s > r

*t > f

*b > l

18. Can we argue that there is some kind of “conspiracy” in languages to produce CV syllable structures? What kinds of sound changes produce this kind of syllable structure? What kinds of sound changes destroy this kind of syllable structure?
19. In the Rotuman language (spoken near Fiji) words appearing in citation (i.e., when the word is being quoted rather than being used in a sentence) differ in shape from words that occur in a natural context. Some of these different forms are presented below. Assuming that the contextual forms are historically derived from the citation forms, what sort of change would you say has taken place?

| Citation Form | Contextual Form | |
|---------------|-----------------|------------|
| laje | laej | 'coral' |
| kami | kaim | 'dog' |
| rako | raok | 'learn' |
| maho | maoh | 'get cold' |
| tepi | teip | 'slow' |
| hefu | heuf | 'star' |
| lima | liam | 'five' |
| tiko | tiok | 'flesh' |
| hosa | hoas | 'flower' |
| mose | moes | 'sleep' |
| pure | puer | 'rule' |

20. In Bislama (Vanuatu), the word for 'rubbish tin' is generally pronounced as [pubel]. Some speakers pronounce this in Bislama as [kubel]. What sort of change is involved here?
21. Compare the forms in Standard French and the French that is spoken in rural Québec in Dataset 12. Assuming that the Standard French forms represent the original situation, what kinds of changes have taken place in the French that is spoken in Québec?

Further Reading

- Hans Henrich Hock, *Principles of Historical Linguistics*, chapters 5–7 on sound change, pp. 61–147.
- Leonard Bloomfield, *Language*, chapter 21, "Types of Phonetic Change," pp. 369–91.
- Robert J. Jeffers and Ilse Lehiste, *Principles and Methods for Historical Linguistics*, chapter 1, "Phonetic Change," pp. 1–16.
- Hans Henrich Hock and Brian D. Joseph, *Language History, Language Change, and Language Relationship*, chapter 4, "Sound Change," pp. 111–48.
- Robert Blust, "Must Sound Change Be Linguistically Motivated?" 22(2): 219–69.
- Brian Joseph and Richard Janda, *Handbook of Historical Linguistics*, sound change, pp. 311–422. (Many of the articles in this book will provide more advanced reading on historical linguistics).
- Juliette Blevins and Andrew Garrett, "The Origins of Consonant-Vowel Metathesis."

Mark Hale, *Historical Linguistics*, part III, "Syntactic Change," has considerably more discussion. Up to now, we've been assuming that defining a "change" is quite straightforward.

Maira Yip, *Tone*, chapter 2, "Contrastive tone" pp. 35–39.

Jean-Marie Hombert et al., "Phonetic Explanations for the Development of Tones."

Graham Thurgood, "Vietnamese and Tonogenesis."

Expressing Sound Changes

3.1 WRITING RULES

When reading the literature of the history of sound changes in languages, you are almost certain to come across various rules written by linguists to express these changes. You will therefore need to know how to write and interpret such rules. This short section of the chapter tells you how to read and write rules.¹

When a sound undergoes a particular change wherever that sound occurs in a language, we refer to this as an UNCONDITIONED SOUND CHANGE. Comparatively few sound changes are completely unconditioned, as generally there are at least some environments (however restricted) in which the change does not take place or in which perhaps some other changes occur. One example of a completely unconditioned sound is that found in the Motu language of Papua New Guinea, where there has been an unconditioned loss of earlier [ɨ], as shown by the following forms (other changes have taken place too, as you can see):

| | Motu | |
|---------|--------|-----------------|
| *asaɨ | > lada | 'gills of fish' |
| *taɨis | > tai | 'cry' |
| *aɨin | > lai | 'wind' |
| *taɨiɨa | > taia | 'ear' |

Similarly, in Hawaiian there was an unconditioned change of [t] to [k], and another of [ŋ] to [n], as shown by the following forms:

| Hawaiian | | |
|----------|---|------------------|
| *tapu | > | kapu 'forbidden' |
| *taŋi | > | kani 'cry' |
| *taŋata | > | kanaka 'man' |
| *ŋutu | > | nuku 'mouth' |
| *tolu | > | kolu 'three' |

Unconditioned sound changes such as these are the simplest historical changes to express in terms of formal rules. The earlier form is given on the left and the later form on the right, with the two being linked by an arrow. So, the Hawaiian changes just described can be expressed simply as:

*t > k
*ŋ > n

The Motu change involving the loss of the velar nasal can be expressed as:

ŋ > ø

(The symbol ø represents the absence of any sound.)

A great many sound changes only take place in certain phonetic environments rather than in all environments in which the sound occurs. Such changes are referred to as **CONDITIONED SOUND CHANGES**, or sometimes as **COMBINATORY SOUND CHANGES**. Most of the sound changes that you saw in chapter 2 were conditioned sound changes. A sound change can be conditioned by a great range of different types of environments. Factors to consider include the position of the sound in a word (whether it is initial, final, or medial), the nature of the preceding and following sounds, the position of stress, whether or not the syllable is open, or perhaps some combination of such conditioning environments.

If a change takes place only in a specific phonetic environment, this environment is written following a single slash (/). The location of the changing sound with respect to the conditioning environment is indicated by a line (___). If a change takes place before some other sound, then the line is placed before the sound that conditions the change; if a change takes place after some other sound, then the line follows the conditioning sound. Some examples of rules expressing conditioned changes that we have looked at, with their expressions in words, are given here:

- *t > s / $_ V_{front}$ [t] became [s] before front vowels (in Motu)
 *x > k / s $_$ [x] became [k] after [s] (in Afrikaans)
 *p > v / V $_ V$ [p] became [v] between vowels (in Banoni)

The symbol V is the standard symbol to express any unspecified vowel. Similarly, any unspecified consonant is expressed by the symbol C.

To express the fact that a change takes place word-finally or word-initially, we use the symbol # to represent the beginning or end of a word, as follows:

- *p > w / # $_$ initial [p] became [w] (in Uradhi)
 C_{voiced} > C_{voiceless} / $_ \#$ final voiced consonants became voiceless
 (in German)
 V > \emptyset / $_ \#$ word final vowels were deleted (in Southeast
 Ambrym)

Elements that are optional (i.e., whose presence or absence does not affect the application of the rule) are placed in round brackets. Thus:

- V > / V_{nasal} (C) $_$ vowels were nasalized after nasal vowels, whether or not
 there is an intervening consonant (in Enggano)

When there are two different sets of sounds involved in a change, this can be represented by placing the sounds one above the other in curly brackets. The Enggano nasal harmony rule described in sec. 2.6 can actually be described more fully in the following way:²

$$\left\{ \begin{array}{c} V \\ \text{voiced stop} \end{array} \right\} > \left\{ \begin{array}{c} V_{nasal} / V_{nasal} \\ \text{Nasal} / \text{Nasal} \end{array} \right\} (C) _$$

A vowel or voiced stop became a nasalized vowel or a nasal consonant, respectively, when there is a preceding nasal vowel or nasal consonant.

Also, the change in Motu involving palatalization (and subsequent lenition) that I described earlier can be alternatively expressed as:

$$*t > s / _ \left\{ \begin{array}{c} i \\ e \end{array} \right\}$$

[t] became [s] before [i] or [e]

Although this is an alternative formulation for the change in Motu, it is considered to be a less "elegant" statement because it misses the

generalization that the conditioning environment is the class of non-low front vowels.

Rules should always be stated in as general a way as possible, without being too general. They are meant to be interpreted literally, so they should not point to changes that did not actually take place. So, while it is true to say that both [i] and [e] are unrounded vowels, we cannot represent this change in Motu as follows:

$$*t > s / _ V_{\text{unrounded}}$$

This would be incorrect because [a] is also an unrounded vowel and the change of [t] to [s] did not take place before [a]. Your rules need to cover all the sounds in the environment that undergo the change, but the rule must also exclude examples where the change did not occur.

3.2 ORDERING OF CHANGES

When a language undergoes a whole series of sound changes, it is sometimes possible to reconstruct not only the changes themselves but also the order in which the changes took place. Let us examine the following data from Hawaiian:³

| | Hawaiian | |
|---------|----------|-----------------|
| *taji | > kani | 'cry' |
| *ʔato | > ako | 'thatch' |
| *takele | > kaʔele | 'back of canoe' |
| *aka | > aʔa | 'root' |
| *pito | > piko | 'navel' |
| *paki | > paʔi | 'slap' |
| *tapu | > kapu | 'forbidden' |
| *taʔata | > kanaka | 'man' |
| *isu | > ihu | 'nose' |
| *sika | > hiʔa | 'firemaking' |

This set of data reveals that the following unconditioned changes have taken place:

$$\begin{aligned} *t &> k \\ *k &> ? \\ *ŋ &> n \\ *s &> h \end{aligned}$$

Of these four changes, we can say something about the order in which they applied. To begin, let us check the first two sound changes to see if we can decide whether [t] shifted to [k] first, or whether [k] first shifted to [ʔ]. If we were to assume that the [t] first shifted to [k] and that the other shift of [k] to [ʔ] took place after this, then changes like the following would have taken place:

- *takele > kakele 'back of canoe'
- *pito > piko 'navel'
- *tapu > kapu 'forbidden'

If [k] then shifted to [ʔ], these words would also have changed as follows, along with all of the other words that contained [k]:

- *kakele > ʔaʔele 'back of canoe'
- *piko > piʔo 'navel'
- *kapu > ʔapu 'forbidden'

The forms [ʔaʔele], [piʔo], and [ʔapu], however, are not the correct forms in Hawaiian, as these words should contain the [k] sound rather than glottal stops. So we must conclude that at the time that [k] shifted to [ʔ] in Hawaiian, there must still have been a distinction between [k] and [t], otherwise all original [k] and [t] would have ended up as [ʔ]. If we were to assume that these two changes applied in the opposite order, then we would get the correct results:

| Protolanguage | Stage 1 *k > ʔ | Stage 2 *t > k | Modern Hawaiian | |
|---------------|----------------|----------------|-----------------|-----------------|
| *takele | taʔele | kaʔele | kaʔele | 'back of canoe' |
| *aka | aʔa | | aʔa | 'root' |
| *pito | | piko | piko | 'navel' |
| *paki | paʔi | | paʔi | 'slap' |
| *tapu | | kapu | kapu | 'forbidden' |

We can represent this by placing one rule over another and linking the two in the following way:

$$\left(\begin{array}{l} *k > ʔ \\ *t > k \end{array} \right)$$

But what about the other changes that have taken place? Can we say anything about whether these changes took place before or after (or between)

the two changes that we have just looked at? In fact, we can only come to conclusions about the ordering of sound changes when the changed sound, or the sounds involved in the conditioning of a change, actually overlap in some way. In the shift of [t] to [k] and the shift of [k] to [ʔ], we were able to say something about the ordering of the two rules because the symbol [k] appears somewhere in the statement of *both* of these changes. In the Hawaiian data presented here, there were also two other changes involved:

$$\left(\begin{array}{l} *ŋ > n \\ *s > h \end{array} \right.$$

None of the symbols in these two rules appear in the statements for either of the changes that I have just been describing. As there is no overlap between the symbols involved in the statement of any of these rules, we cannot come to any conclusion about the ordering of these rules. It does not make any difference whether we apply these two rules first, last, or between the other rules: the end results will not be affected in any way. Historically, of course, these two changes must have applied at some period, either before the change of [k] to [ʔ] or after it, or perhaps at the same time as that change. However, on the evidence that we have, there is no way that we can find out when these other changes took place.

In listing the full set of changes for this set of data in Hawaiian, we can indicate the fact that there is no evidence that a particular change is ordered either before or after any other change simply by not linking them as we did above. So, the ordering of these four changes could be equally represented in any of the following ways:

$$\begin{array}{l} (*k > ʔ \quad *ŋ > n \\ *t > k \quad *s > h \\ *ŋ > n \quad *t > k \\ *s > h \quad *ŋ > n \end{array} \left(\begin{array}{l} *k > ʔ \quad *ŋ > n \\ *s > h \quad *s > h \\ *t > k \quad (*k > ʔ \\ *t > k \end{array} \right.$$

In fact, it does not matter in what order you write the rules for these changes, as the only changes that are linked in time are those that are marked with the special symbol that is used for indicating the ordering of sound changes. The placement of any other changes among a set of changes is purely a matter of convenience.

Let us now look at a more complicated example, in which conditioned sound changes are involved. The data come from the Banoni language of Bougainville (an autonomous region of Papua New Guinea):⁴

Banoni

| | | | |
|--------|---|---------|----------------------|
| *koti | > | kotʃi | 'cut' |
| *tina | > | tʃina | 'mother' |
| *puti | > | putʃi | 'pull out' |
| *mata | > | mata | 'eye' |
| *mate | > | mate | 'die' |
| *paʃan | > | βaʃana | 'add meat to staple' |
| *kulit | > | yuritʃi | 'skin sugarcane' |

The sound changes that I will look at are the following:

*t > tʃ / __ V_{high}

$$* \emptyset \left\{ \begin{array}{c} i \\ e \\ a \\ o \\ u \end{array} \right\} / \left\{ \begin{array}{c} i \\ e \\ a \\ o \\ u \end{array} \right\} C\#$$

The first rule changes [t] to [tʃ] before the high vowels [i] or [u]. The second rule involves the addition of a harmonizing vowel after a consonant at the end of a word. (Some other changes are indicated in these data, but they will be ignored at this point.)⁵

The question that you should ask yourself is: Can these two changes be ordered with respect to each other? According to what I said earlier, if two changes involve some common sound, either in the changing sounds or in the conditioning sounds, then we can test to see which applied first. Since these two rules both involve the symbol V referring to vowels, we can test them for ordering.

If we were to assume that the change of [t] to [tʃ] took place first, we could correctly predict the application of this change in all cases but one—the Banoni form of the original word *kulit 'skin sugarcane'. Because this form has no following vowel in the protolanguage, it does not meet all of the conditions for the application of the rule that changes [t] to [tʃ]. However, if the vowel addition rule were to apply only after the change of [t] to [tʃ], we would end up with [yuritʃi] for this word (assuming that we apply the other incidental consonant changes as well). The fact that the actual form is [yuritʃi] rather than [yuriti] means that there must already have been a high vowel after the [t] when the rule affecting the [t] applied. This shows that the rule adding a final harmonizing vowel must have applied before the rule changing the [t] to [tʃ]. So we can state the ordering of these two changes as follows:

(Harmonizing Vowel Addition
 *[t] > [tʃ]

Reading Guide Questions

1. What is meant by saying that rules should be written to be as general as possible but not too general?
2. What is meant by speaking of “ordered rules”?
3. How do we decide on the ordering of rules, and how do we show the relative ordering of rules?

Exercises

1. Make a summary chart of the rule notation that you have learnt in this chapter.
2. Express the following changes formally:
 - a. Intervocalic [s] undergoes rhotacism while [s] before consonants is deleted.
 - b. Word-initial consonants undergo weakening to [j].
 - c. Intervocalic [h] changes to glottal stop.
 - d. The second member of all consonant clusters is deleted.
 - e. An epenthetic [o] is added between the two members of a word-final consonant cluster.
 - f. Word-final high vowels are deleted while interconsonantal high vowels become schwa.
 - g. A prothetic [h] is added before [e] and [o].
 - h. A prothetic vowel is added to all words which start with a fricative; the vowel is identical to the vowel following the fricative.
3. Examine the Nganyaywana forms in Dataset 2.
 - a. Under what conditions are the vowels of initial syllables retained, and when are they lost?
 - b. Long vowels are shortened. Did this change take place before or after the loss of vowels dealt with in the previous question? Why?
4. Examine the Mbabaram forms in Dataset 3.
 - a. Some word-final [a] became [e], some became [o], and some remained unchanged. What are the conditioning factors?
 - b. Initial syllables were lost. Did this change take place before or after the changes affecting final [a]? Why?
5. Examine the Yimas and Karawari forms in Dataset 4.
 - a. Formulate explicit rules for the changes that have taken place in each of the two languages.

- b. Can you find any evidence concerning the ordering of any of these changes, in either Yimas or Karawari?
- c. Given the following original forms, what would you expect the modern Yimas and Karawari words to be?

*simari 'sun'
 *simasim 'sago'
 *nanɣun 'mosquito'

6. Examine the Lakalai forms in Dataset 5.
 - a. Write formal rules to account for all of the changes that have taken place.
 - b. Do any of these changes need to be ordered with respect to each other? Why?
7. Examine the changes in Motu in Dataset 9.
 - a. What are the rules that express the various changes that have taken place here?
 - b. What is the ordering of these rules?
8. Examine the Burduna forms in Dataset 11.
 - a. Write rules that express the changes that have taken place.
 - b. Is there any evidence that any of these changes must have taken place before any others? If so, say what they are.
9. Examine the following data from the Mpakwithi language of Cape York Peninsula in northern Queensland (Australia):

*maɪa > ʔa 'hand'
 *kuta > ʔwa 'dog'
 *pakaj > kaɪa 'down'
 *pama > ma 'person'
 *puŋku > gu 'knee'
 *ɲipima > pimi 'one'
 *muŋka > gwa 'eat'
 *cuma > mwa 'fire'
 *ɲaŋku > gaw 'that'
 *japi > paj 'forehead'
 *ɲampu > baw 'tooth'

- a. Describe in words the changes that have taken place in this language. (There is not enough data here for you to be able to write fully explicit rules.)
- b. Can you suggest anything about the order in which these changes have taken place?

10. Examine the standard French and rural Québec French forms in Dataset 12. Assuming that the standard French forms represent the original pronunciation, except that [ʁ] was originally pronounced as [r], write rules expressing the changes that have taken place in rural Québec French.

Further Reading

Francis Katamba, *Morphology*.

Mark Hale, *Historical Linguistics*, has extensive discussion of the status of “changes” in historical linguistics (that is, what it means to say that $*X > Y$).

Lyle Campbell, *Historical Linguistics*, chapter 2, “Sound Change,” pp. 16–49, has many examples of rule writing.

Phonetic and Phonemic Change

When a linguist describes the synchronic sound system of a language, she or he must be aware of the difference between a phonetic description and a phonemic description of the language. A **PHONETIC DESCRIPTION** simply describes the physical facts of the sounds of the language. A **PHONEMIC DESCRIPTION**, however, describes the way that these sounds are related to each other for speakers of that particular language. It is possible for two languages to have the same physical sounds yet to have very different phonemic systems. The phonemic description therefore tells us what are the basic sound units for a particular language that enable its speakers to differentiate meanings.

Just as it is possible to describe a language synchronically in both phonetic and phonemic terms, it is possible to make a distinction between a diachronic phonetic study and a diachronic phonemic study of a language. It is possible, therefore, for some sound changes to take place without altering the phonemic structure of a language, though many sound changes do alter it. But it is also possible for a phonemic change to take place in a language without there being a phonetic change.

Up to now, we have been talking about “sound change” without making it clear how the sounds relate to one another in the system of the protolanguage and its daughter languages. In this chapter, we look in more detail into sound changes that result in changes to systems.¹

4.1 PHONETIC CHANGE WITHOUT PHONEMIC CHANGE

Many phonetic changes take place in languages without in any way altering phoneme inventory or the relations between phonemes. Such change is therefore purely allophonic or subphonemic. All that happens is that a phoneme develops a new allophone (or changes its phonetic form slightly), or the distribution of existing allophones of a phoneme is changed.

One example of a subphonemic change in the history of English involves the phoneme /r/. This phoneme has always been spelled with the symbol *r*, right from the earliest records. This suggests that speakers of English have not perceived any change in this sound. However, we do know that earlier the phoneme /r/ was pronounced phonetically as a flap or trill (as is still the case in Scots English), rather than as the frictionless continuant [ɹ] that most speakers of English use today. Although this sound has changed phonetically, it has not caused any reanalysis of the phonological system to take place. The same words that used to be distinguished in meaning from other words by a flap or a trill are now distinguished, instead, by [ɹ]. This change could be represented as:

$$/r/ : [r] \sim [r] > /r/ : [ɹ]$$

Another example of phonetic change without phonemic change from the history of English involves the short high front vowel phoneme. In most dialects of English this is pronounced as [ɪ]. In the New Zealand dialect of English, however, this has been centralized in the direction of [i]. The change from [ɪ] to [i] has again not caused any new meaning contrasts to develop. The same words are distinguished in New Zealand English as in other varieties of English, only by a slightly different phonetic form. Again, this purely allophonic change can be represented as:

$$/ɪ/ : [ɪ] > /ɪ/ : [i]$$

The final example that I give here of subphonemic change comes from the Motu language of Papua New Guinea. The previous two examples from English involve a change in the phonetic form of the phoneme wherever it occurs: that is, they are examples of unconditioned allophonic change. However, in the case of a conditioned subphonemic change, a new allophone is created in a particular phonetic environment, though the sound remains unchanged in other environments. No new phonemes are created, only a new allophone of an existing phoneme.

You should remember from chapter 2 that, in Motu, [t] has shifted to [s] before front vowels while remaining unchanged in other environments. This change is the only source of the sound [s] in Motu, as no other sound changes

have produced any [s], and there was no [s] sound at all in the protolanguage. This means that the shift of [t] to [s] did not in any way affect the phonemic structure of the language. All instances of the sound [s] in Motu today are in complementary distribution with [t]. The sound [s] only ever occurs before front vowels, while [t] never occurs before front vowels. The [s] that developed was simply a new allophone of the phoneme /t/.² This change can therefore be stated as:

$$/*t/ > /t/ : \begin{cases} [s] \text{ before front vowels} \\ [t] \text{ elsewhere} \end{cases}$$

4.2 PHONETIC CHANGE WITH PHONEMIC CHANGE

You saw in the preceding section that a phonetic change need not necessarily lead to a change in the phonemic system of a language. Often, however, phonetic change does lead to some kind of phonemic change. Generally speaking, we can say that phonetic change is a “tool” of phonemic change in the sense that most instances of phonemic change are the result of a phonetic change in that particular sound. Phonemic changes can be subcategorized into three different types: phonemic loss, phonemic addition, and rephonemicization.

4.2.1 Phonemic Loss

The term **PHONEMIC LOSS** is self-explanatory. Phoneme loss takes place when a phoneme disappears altogether between different stages of a language. All cases of unconditioned sound loss at the phonetic level necessarily imply complete phonemic loss. An example of such a loss is the disappearance of the velar nasal from the phoneme inventory of Motu, which you saw in chapter 3.

Phoneme loss often involves a conditioned sound change, occurring in some environments and not in others. While the loss of the velar nasal in Motu is an unconditioned sound change, you will frequently find that only some occurrences of a phoneme are lost, while others are retained. This situation can be referred to as **PARTIAL LOSS**, in contrast to **COMPLETE LOSS**. For an example of partial loss in Fijian, refer to the discussion of the loss of final consonants in chapter 2. This change can be represented as:

$$C > \emptyset / _ \#$$

In the Angkamuthi example that immediately followed the Fijian example in chapter 2, you can see that there has been partial consonantal loss again, this time in word-initial position (which I referred to then as apheresis), according to the following rule:

$$C > \emptyset / \# _$$

4.2.2 Phonemic Addition

This term is also self-explanatory. PHONEMIC ADDITION takes place when a phoneme is inserted into a word, in a position in which that phoneme did not originally occur. For example, in Motu again, a prothetic /l/ was added before the vowel /a/, creating a new set of words distinguished by this sound, as you saw in chapter 2.

Certainly, simple phonetic addition does not necessarily lead to phonemic addition. It is possible for a sound to be added without actually affecting the phonemic form of a word. In the Mpakwithi language of northern Queensland (in Australia), words beginning with fricatives and the rhotic flap have added an optional prothetic schwa; for example:

| | | |
|----------|------------------|--------------|
| /βaði/ : | [βaði] ~ [əβaði] | 'intestines' |
| /ðaj/ : | [ðaj] ~ [əðaj] | 'mother' |
| /ra/ : | [ra] ~ [əra] | 'stomach' |

There is no separate schwa phoneme in this language. The sound [ə] occurs only in forms such as those just given, and it is completely predictable in its occurrence. It never contrasts with anything. While the following phonetic change has taken place (i.e., a schwa is added before fricatives and /r/ at the beginning of a word), the actual phonemic form of such words has not changed:

$$/*\emptyset/ > \emptyset / _ \left\{ \begin{array}{l} \text{fricative} \\ r \end{array} \right.$$

This has therefore been an example of phonetic addition without phonemic addition.

A further type of phonemic addition occurs in loan situations. If speakers of a language borrow a lot of words without adapting them to their existing system, new phonemes can enter the language through the medium of those loans. The English phoneme [ʒ] is an example of this; it entered the language through loans from French (such as *rouge*, *measure*, *treasure*, and the like) and is now established in the system, although with a somewhat restricted distribution.

4.2.3 Rephonemicization

The most common kind of phonemic change to result from phonetic change is *REPHONEMICIZATION*. What this involves is the creation of a new pattern of oppositions in a language by simply changing around some of the existing phonemes, or by changing some of the existing phonemes into completely new phonemes. Whereas phoneme addition means adding a new phoneme in a word where there was no phoneme originally, and phoneme loss means deleting a phoneme from a word where there originally was one, rephonemicization involves changing around the phonemes that are already there in the word. There are a number of different kinds of rephonemicization: shift, merger, and split. I describe each of these next.

4.2.3.1 *Shift*

The first kind of rephonemicization that we consider goes under the name of *SHIFT*. When phonemic shift has taken place, two words that were distinguished in the protolanguage by means of a particular pair of sounds are still distinguished in the daughter language, but the distinction between the two words is marked by a different pair of sounds. That is, a minimal pair in the protolanguage will still be different in the daughter language, but the difference is not marked by the original sounds. For instance, in the history of the Banoni language of Papua New Guinea, voiceless stops became voiced fricatives (along with a number of other changes). It is quite possible to imagine a minimal pair in the protolanguage in which meanings are distinguished by the presence or absence of a voiceless stop between vowels. In the modern language, however, the same difference in meaning is marked instead by the presence or absence of a voiced fricative in the same position.

A thoughtful reader should have noticed that this description of phonemic shift does not seem to be very different from what I said earlier about purely phonetic change. When allophonic change takes place, there is also a change in the actual sounds that are used to distinguish meanings. The important difference between the two situations is that, with phonemic shift, the original sound and the new sound must actually belong to separate phonemes. In Banoni today, there are pairs to show that voiceless stops and voiced fricatives are phonemically distinct, for example:

[kasi:] 'my brother'

[yasi] 'open'

This shows that when the voiceless stops changed to voiced fricatives, there was an actual shifting around of phonemes in the language, not just a shifting around of the allophones within a phoneme.

4.2.3.2 *Merger*

The second kind of rephonemicization is phonemic MERGER. This is the process by which two separate phonemes end up as a single phoneme. Words that used to be distinguished by some difference in sound cease to be distinguished, and what were originally minimal pairs become homophones (or homonyms)—that is, words with the same form but different meanings. For instance, the Motu word /lada/ is a homophone, referring both to ‘gills of fish’ and ‘name’. In the protolanguage from which Motu was derived, there were originally two different words, distinguished by different phonemes:

- *aʝan ‘name’
 *asaʝ ‘gills of fish’

There has been a phonemic merger of /ʝ/ and /s/ as /d/ (as well as a loss of final consonants and the addition of a prothetic /l/), producing the modern homophone.

Another example of merger comes from Indo-European. Proto-Indo-European³ had three types of velar sounds: a velar /k/; a labiovelar /kʷ/; and a front (palatalized) /kʲ/, which is usually represented in books about Indo-European as /kʰ/. (There were also corresponding sets of voiced and breathy stops.) None of the descendent languages has all three types of stops, but because the mergers happened in different ways in different branches of the family, we can tell that there must have been three sets of sounds. (Examples are taken from Benjamin Fortson 2004:52–53.)

| | *k̂ *k̂erd- ‘heart’ | *k *kes- ‘comb’ | *kʷ *kʷi- , kʷo ‘who, what’ |
|---------------------|---------------------|-----------------|-----------------------------|
| Hittite | kard- | kišš | kuit |
| Greek | kard- | késkeon | tí |
| Latin | cord- | | quid |
| Old Irish | cride | cír | |
| Old English | heart | heord | hwat |
| Tocharian | kär- | | |
| Sanskrit | śrād- | | kás |
| Old Church Slavonic | sřid- | kosa | |
| Lithuanian | šrdis | kasà | kàs |

What has happened is that the three stops in Proto-Indo-European merged into two sets in just about all the attested languages (they further merged to a single set in Tocharian, and the Anatolian language Luvian seems

to keep all three distinct). However, the merger happened in two ways. In one set of languages, *k̂ and *k merged to *k, and *k^w remained distinct. This is what happened in the history of English, Latin, Greek, and a number of other languages. In the other set of languages, *k and *k^w merged as *k, and *k̂ remained distinct. This happened in Sanskrit, Lithuanian, and Old Church Slavonic. Subsequently there were other changes.

Phonemic merger can be represented as follows:

$$\left. \begin{matrix} X \\ Y \end{matrix} \right\} > Z$$

(although merger can involve more than just two sounds).

When phonemes merge in this way, there are two possible forms for the phoneme that is symbolized here as Z. First, Z could be identical to one of the original phonemes. Second, it could be different from either of the original phonemes (i.e., a completely new phoneme). An example of phonemic merger where the resulting phoneme is phonetically the same as one of the original phonemes is Uradhi, an Australian language of northern Queensland:

Uradhi

- *paṭa > waṭa 'bite'
- *pinta > winta 'arm'
- *pupu > wupu 'buttocks'
- *wapun > wapun 'head'
- *wujpu > wujpu 'old man'

The original /p/ and /w/ have merged as /w/ (though only in word-initial position):

$$\left. \begin{matrix} p \\ w \end{matrix} \right\} > w / _$$

An example of the second possibility is the following change in Fijian:

Fijian

- *tuba > tuva 'fish poison'
- *batu > vatu 'stone'
- *ubi > uvi 'yam'
- *pitu > vitu 'seven'
- *pəpu > vonu 'turtle'

The original phonemic distinction between /b/ and /p/ is lost, and the descendant of the merged phoneme is different from both of the original phonemes to /v/:

$$\left. \begin{array}{l} *b \\ *p \end{array} \right\} > v$$

I have been talking about merger, but I have not pointed out that there is a distinction to be made between partial merger and complete merger. COMPLETE MERGER means that the sound change that produces the merger is UNCONDITIONED: the change affects that particular sound in all environments in which it occurs. PARTIAL MERGER, in contrast, means that the sound change is a conditioned one: the particular phonemes merge only in certain environments and are kept distinct in others. The example that I gave above of Uradhi as an example of the merger of /p/ and /w/ is actually an example of partial rather than complete merger, as it was necessary to indicate the environment in which the change took place. The merger takes place only word-initially, while in word-medial position the original distinction between /p/ and /w/ is maintained.

4.2.3.3 Phonemic Split

The opposite of phonemic merger is PHONEMIC SPLIT. Words which originally contained the same phoneme end up having different phonemes. Phonemic split can arise when a single sound changes in different ways in different phonological environments. We can represent this kind of change in the following way:

$$*X > \begin{cases} Y / A \\ Z / B \end{cases}$$

However, if there is a conditioned sound change of this type, and the only source for the new sound is this change, then we cannot speak of phonemic split. What we have is a case of subphonemic change, as we have only produced a new allophone of an existing phoneme in a specific environment. This is exactly what we saw happening in Motu, where the original [t] has changed to [s] in some environments and remained as [t] in others. This cannot be considered as phonemic split because no new phonemes are involved.

But if two or more sound changes operate at once to produce the same sound, then we can speak of phonemic split. In the Angkamuthi language of Cape York in Queensland (Australia), the following change took place:

$$*l > \begin{cases} j / _ a, i \\ l / _ u \end{cases}$$

If there were no other changes word-initially (and if there was not already a phoneme /j/ in the language), we could say that this change simply produces a new allophone of /l/ word-initially before /a/ and /i/. If there was not an original /j/ phoneme and the following change were to take place, we could also speak of genuine phonemic split taking place:

$$*l > j$$

With this change, /j/ and /l/ could no longer be in complementary distribution, so a phonemic split would have resulted.

4.3 PHONEMIC CHANGE WITHOUT PHONETIC CHANGE

In this section, we look at a series of situations in which the phonemic status of a sound changes without any actual phonetic change taking place in the sound that has changed phonemically (though there may be phonetic changes elsewhere in the word). The way this change arises is through the loss of conditioning environment.⁴

Originally, in English, there was no velar nasal phoneme /ŋ/, though this sound did occur as an allophone of the phoneme /n/ before velar sounds. This can be represented by the following allophonic statement:

$$/n/: \begin{cases} [ŋ] \text{ before velars} \\ [n] \text{ elsewhere} \end{cases}$$

A word like *singer*, which we now write phonemically as /sɪŋə/,⁵ was originally phonemically /sɪŋə/, but phonetically the medial nasal had the same pronunciation as it has today. This word was therefore pronounced as [sɪŋgə]. This is an example of a phonemic change (i.e., /n/ shifting to /ŋ/) that does not involve any phonetic change. How did this come about?

The separate status of the phoneme /ŋ/ came about as the result of another change that caused the loss of the sound that conditioned the choice between the alveolar and the velar allophones of /n/. Look at the following earlier forms and the changes that they underwent. (These forms are given first phonemically; the second form in square brackets gives the actual phonetic form.)

| Earlier English | | Modern English | |
|-------------------|---|----------------|--------|
| * /sɪn/ : [sɪn] | > | /sɪn/ : [sɪn] | 'sin' |
| * /sɪŋg/ : [sɪŋg] | > | /sɪN/ : [sɪN] | 'sing' |
| * /læmb/ : [læmb] | > | /læm/ : [læm] | 'lamb' |

Word-finally after nasals in English, the voiced stops /b/ and /g/ (but not /d/) were lost by a rule of the form:

$$\left. \begin{array}{l} b \\ g \end{array} \right\} > \emptyset / \text{nasal} _ \#$$

This explains the presence of the so-called silent *b* in words such as *climb* and *lamb*. Now, you will remember that it was the presence of a velar phoneme earlier in English that conditioned the choice of a velar allophone of the phoneme /n/ rather than an alveolar allophone. So, phonemic /sɪŋg/ was phonetically [sɪŋg] (as it still is in some northern dialects in England).⁶ However, once the final /g/ was lost, the [ŋ] now came to be in contrastive distribution with [n], whereas before the two were in complementary distribution. As evidence of this, we find the minimal pair /sɪŋ/ 'sing' and /sɪn/ 'sin'. Here you can see that although the velar nasal itself did not change *phonetically* in English, its *phonemic* status has changed because its original conditioning environment has been lost.

Another well-known example of this kind of change is the development of umlaut in Germanic languages. Umlaut is the changing of a vowel of a root to become either more front or more high in certain morphological categories. As we see in chapter 2, the irregular plural in English of *foot/feet*, as well as other forms such as *tooth/teeth*, derive from an earlier plural suffix /-i/, which was added to the singular roots /fo:t/ and /to:θ/, respectively. Then a purely allophonic change took place, by which all back-rounded vowels became front-rounded vowels when the following syllable contained a front vowel. So although there was no phonemic change in the plural, there was a change in the phonetic form of the plural of these two words under the influence of the following plural suffix:

| | | | |
|---------------------|---|-------------------|---------|
| * /fo:ti/ : [fo:ti] | > | /fo:ti/ : [fø:ti] | 'feet' |
| * /to:θi/ : [to:θi] | > | /to:θi/ : [tø:θi] | 'teeth' |

The next change involves a change in the phonemic status of the front-rounded vowels. Although these vowels themselves did not then change phonetically in any way, there was a general rule of apocope at this stage in the history of English which deleted the final /-i/ marking the plural. Thus:

**/fo:ti/ : [fø:ti] > /fø:t/ : [fø:t]* 'feet'

**/to:θi/ : [tø:Ti] > /tø:θ/ : [tø:θ]* 'teeth'

This loss of the conditioning vowel resulted in the existence of minimal pairs between back- and front-rounded vowels, with the back-rounded form occurring in the singular and the front-rounded form occurring in the plural. It is from these two forms that the modern irregular plurals are directly derived.

I mentioned in the preceding section that although Motu has undergone a change by which /t/ developed a new allophone of the form [s] before a front vowel, this did not introduce any new phonemic contrasts into the language. Now there is a tendency among younger Motu speakers to drop word-final vowels. So we find alternative pronunciations such as the following:

/tinagu/ : [sinagu ~ sinag] 'mother'

/oiemu/ : [oiemu ~ oiem] 'your'

/namo/ : [namo ~ nam] 'good'

/mate/ : [mase ~ mas] 'die'

Let's imagine that in two generations' time this change might have become general and that all word-final vowels following consonants were lost by a rule that we could write as follows:

**V > ø / C _ #*

Let us examine what would happen to minimal pairs such as */lati/* 'no' and */lata/* 'long'. These forms are currently pronounced as follows:

/lati/ : [lasi ~ las] 'no'

/lata/ : [lata ~ lat] 'long'

If the rule of optional word-final vowel loss were to become general, this pair, which is now distinguished phonemically by the nature of the final vowel, would come to be distinguished solely by the nature of what were originally intervocalic consonants, as follows:

/las/ 'no'

/lat/ 'long'

Thus, what was originally just a phonetic difference between [t] and [s] would become a phonemic contrast between /t/ and /s/.

Reading Guide Questions

1. What is allophonic change?
2. What is phonemic loss?
3. What is the difference between partial and complete loss?
4. What is rephonemicization?
5. What is phonemic shift? How does this differ from allophonic change?
6. What is phonemic merger?
7. What is the difference between complete and partial phoneme merger?
8. What is phonemic split?
9. Explain in what ways a sound can change phonemically without changing phonetically.

Exercises

1. Examine the following forms in Tongan and Māori. Assume that the vowels of Tongan reflect the vowels of the original language and that Māori has innovated. Both Tongan and Māori today have five short vowel phonemes. Would you classify the changes to the vowels in Māori as phonetic change, phonemic shift, phonemic merger, or phonemic split?

| Tongan | Māori | |
|----------|--------|---------------|
| ʔutu | ʔutu | 'mouth' |
| au | au | 'I' |
| hoa | hoa | 'friend' |
| fulufulu | hūhūhū | 'feather' |
| ihu | ihu | 'nose' |
| inu | inu | 'drink' |
| hiʔoa | iʔoa | 'name' |
| malawē | marae | 'open ground' |
| mata | mata | 'face' |
| mate | mate | 'dead' |
| moana | moana | 'sea' |
| mutu | mutu | 'finish' |
| nifo | niho | 'tooth' |
| lau | rau | 'leaf' |
| nima | rima | 'five' |
| tolu | toru | 'three' |
| tapu | tapu | 'forbidden' |

2. Less-educated speakers of some regional dialects of Tok Pisin in Papua New Guinea change some of the sounds used by speakers of the standard dialect. Imagine somebody speaking the following extremely nonstandard regional dialect. Standard Tok Pisin has no [f] while the nonstandard dialect described here has no [p]. There is no [s] or [l] in the nonstandard dialect. Describe the changes they have made to the phonemic system of the standard language in terms of the kinds of changes that we have been looking at in this chapter.

| Standard Tok Pisin | Nonstandard Tok Pisin | |
|--------------------|-----------------------|---------------|
| ples | feret | 'village' |
| poret | foret | 'frightened' |
| mipla | mifara | 'we' |
| larim | rarim | 'leave' |
| kisim | kitim | 'take' |
| lotu | rotu | 'church' |
| sarip | tarif | 'grass knife' |
| popaia | fofaia | 'miss' |
| sori | tori | 'concerned' |
| belo | bero | 'bell' |
| sapos | tafot | 'if' |
| kirap | kiraf | 'get up' |
| gutpla | gutfara | 'good' |

3. Examine the Mbabaram forms in Dataset 3. In the original language, there were only three vowel phonemes: /i/, /u/, and /a/. Describe how the changes that have taken place have affected the phonemic system.
4. In the Lakalai forms in Dataset 5, describe the various changes that have taken place as merger, loss, or shift.
5. Describe the sound changes implied in Tiene in Dataset 13, including the tone changes.

Further Reading

Winfred P. Lehmann, *Historical Linguistics*, chapter 10, "Change in Phonological Systems," pp. 147–76.

Raimo Anttila, *An Introduction to Historical and Comparative Linguistics*, chapter 4, "Sound Change," pp. 57–87.

Robert J. Jeffers and Ilse Lehiste, *Principles and Methods for Historical Linguistics*, chapter 5, "Phonological Change," pp. 74–87.

Hans Henrich Hock, *Principles of Historical Linguistics*, chapter 4, "Sound Change and Phonological Contrast," pp. 52–60.

Lyle Campbell, *Historical Linguistics*, chapter 2 "Sound Change," pp. 19–25

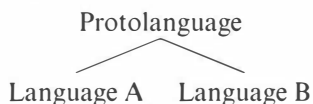
The Comparative Method (1): Procedures

Up to now, I have been giving examples of changes in languages from an earlier form (marked with the asterisk *) to a later form, but I have not said how these earlier forms have actually been worked out. So far, this has all simply been done on trust! The use of the asterisk is intended to mark the words as unrecorded, never actually seen or heard by anybody who is around now. Do linguists just guess at these forms and hope they are more or less right, or is there some special method by which we can deduce what these forms were like? How can we “undo” the changes that have taken place in languages to find out what the original forms were likely to have been? The method is not a hard and fast “algorithm” for working out what’s happened, but there are a series of heuristics or guidelines for making the hypotheses. In this chapter, we’ll talk about how to make these hypotheses.

5.1 SOUND CORRESPONDENCES AND RECONSTRUCTION

I have already discussed the idea of languages being genetically related in families, all of which are descended from a single ancestor, which

we call the protolanguage. This model of language evolution looks like this:



Even if we have no written records of the protolanguage, it is often possible to reconstruct some of the aspects of the original language from the REFLEXES in the daughter languages by using the COMPARATIVE METHOD. When I use the term “reconstruct”, I mean that we make some kind of estimation about what a protolanguage might have been like. We are in a sense “undoing” the changes that have taken place between the protolanguage and its various descendant languages. To do this, you have to examine what we call “reflexes of forms” of the original language in these daughter languages. By this, I mean that you have to look for forms in the various related languages which appear to be derived from a common original form. Two such forms are cognate with each other, and both are reflexes of the same form in the protolanguage.

In carrying out linguistic reconstruction in this way, we use the comparative method. This means that we compare cognate forms in two (or preferably more) related languages in order to work out some original form from which these cognates could reasonably be derived. In doing this, we have to keep in mind what is already known about the kinds of sound changes that are likely, and the kinds of changes that are unlikely. (Thus it is necessary to keep in mind the survey of types of sound change that are described in chapter 2 of this book when doing reconstruction of this kind.)

5.2 AN EXAMPLE OF RECONSTRUCTION: PROTO-POLYNESIAN

5.2.1 Setting Out the Data

Now that we have learned some of the basic terminology that is necessary for reconstructing languages, let us go on to look at an actual linguistic situation and see what we can make of it. We will look at some data from four Polynesian languages: Tongan, Samoan, Rarotongan (spoken in the Cook Islands, near Tahiti), and Hawaiian (table 5.1).¹ Assuming that there was once a language that we can now call Proto-Polynesian, what do we have to do in order to reconstruct this language out of this body of data in its modern descendant languages?

TABLE 5.1 Data from Four Polynesian Languages

| | Tongan | Samoaan | Rarotongan | Hawaiian | |
|-----|----------|---------|------------|----------|----------------------|
| 1. | tapu | tapu | tapu | kapu | 'forbidden' |
| 2. | pito | pute | pito | piko | 'navel' |
| 3. | puhi | feula | puʔi | puhi | 'blow' |
| 4. | tafaʔaki | tafa | taʔa | kaha | 'side' |
| 5. | taʔe | tae | tae | kae | 'feces' |
| 6. | taŋata | taŋata | taŋata | kanaka | 'man' |
| 7. | tahi | tai | tai | kai | 'sea' |
| 8. | malohi | malosi | kaʔa | ʔaha | 'strong' |
| 9. | kalo | ʔalo | karo | ʔalo | 'dodge' |
| 10. | aka | aʔa | aka | aʔa | 'root' |
| 11. | ʔahu | au | au | au | 'gall' |
| 12. | ʔulu | ulu | uru | poʔo | 'head' |
| 13. | ʔufi | ufi | uʔi | uhi | 'yam' |
| 14. | afi | afi | aʔi | ahi | 'fire' |
| 15. | faa | faa | ʔaa | haa | 'four' |
| 16. | feke | feʔe | ʔeke | heʔe | 'octopus' |
| 17. | ika | iʔa | ika | iʔa | 'fish' |
| 18. | ihu | isu | putaŋio | ihu | 'nose' |
| 19. | hau | sau | ʔau | hau | 'dew' |
| 20. | tafuafi | siʔa | ʔika | hiʔa | 'firemaking' |
| 21. | hiku | siʔu | ʔiku | hiʔu | 'tail' |
| 22. | hake | aʔe | ake | aʔe | 'up' |
| 23. | huu | ulu | uru | komo | 'enter' |
| 24. | maŋa | maŋa | maŋa | mana | 'branch' |
| 25. | maʔu | mau | mau | mau | 'constant' |
| 26. | maa | mala | mara | mala | 'fermented' |
| 27. | naʔa | faʔaŋa | maninia | naa | 'quieten' |
| 28. | nofo | nofo | noʔo | noho | 'sit' |
| 29. | ŋalu | ŋalu | ŋaru | nalu | 'wave' |
| 30. | ŋutu | ŋutu | ŋutu | nuku | 'mouth' |
| 31. | vaka | vaʔa | vaka | waʔa | 'canoe' |
| 32. | vaʔe | vae | vae | wae | 'leg' |
| 33. | laho | laso | raʔo | laho | 'scrotum' |
| 34. | lohu | lou | rou | lou | 'fruit-picking pole' |
| 35. | oŋo | loŋo | roŋo | lono | 'hear' |
| 36. | ua | lua | rua | lua | 'two' |

5.2.2 Finding the Cognates

There are a number of steps that you must follow. The first step is to sort out the forms that appear to be cognate from those that do not. If two words are not cognate, it means that they are derived from different original forms, and are not reflexes of the same original form as the others. In deciding whether or not two forms are cognate, you need to consider how similar they are, both in form and in meaning. If they are similar enough that it could be assumed they are derived from a single original form with a single original meaning, then we say that they are COGNATE.

You can begin by excluding from the list above a word such as /tafuafi/ 'firemaking' in Tongan (20). The words to express the same meaning in the other three languages are /siʔa/ in Samoan, /ʔika/ in Rarotongan, and /hiʔa/ in Hawaiian. These last three forms are all quite similar phonetically, as well as being identical in meaning, and it is easy to imagine that they might be reflexes of a single original word in Proto-Polynesian. The Tongan word /tafuafi/, although it has the same meaning, is so different in its shape that you can assume that it has a totally different source altogether.

The fact that the Tongan word /tafuafi/ contains the final element /-afi/, along with the fact that the Tongan word for 'fire' is /afi/ (14), suggest that this word may be a combination of some unknown element /tafu-/ and the word for 'fire'. Example 4 presents us with a similar case:

| | Tongan | Samoa | Rarotongan | Hawaiian | |
|----|----------|-------|------------|----------|--------|
| 4. | tafaʔaki | tafa | taʔa | kaha | 'side' |

It seems clear that the first two syllables of the longer Tongan word are cognate with the words in the remaining Polynesian languages. The second two syllables of the Tongan form, however, do not have any cognate forms in the other languages. We can therefore assume that in Tongan, at some stage in its history, an extra morpheme was added. What was originally regarded as being a morphologically complex word then came to be regarded by speakers as morphologically simple. That is, some other morpheme came to be *reanalyzed* as part of the root. In carrying out comparative reconstruction, you must also exclude examples such as these that involve reanalysis and consider only those parts of words that are actually cognate. We can therefore set out the cognate forms in these four languages in this case as follows, with the noncognate part of the Tongan word removed, and a hyphen being used to indicate that something has been left off:

| | Tongan | Samoa | Rarotongan | Hawaiian | |
|----|--------|-------|------------|----------|--------|
| 4. | tafa- | tafa | taʔa | kaha | 'side' |

From the data in table 5.1, there are several other forms expressing the same meaning that we would want to exclude as not being cognate because they are phonologically so different from the forms in the other languages. In the Samoan data, you should probably exclude the following: (2) /pute/ 'navel', (3) /feula/ 'blow', and (27) /faʔaŋa/ 'quieten'. In Rarotongan, you must exclude the following forms, which are apparently not cognate with words in other languages expressing the same meaning: (18) /putaŋjo/ 'nose' and (27) /maninia/ 'quieten'. Finally, in Hawaiian, you will need to exclude the following: (12) /poʔo/ 'head' and (23) /komo/ 'enter'. (While we are discussing which words we should consider to be cognate, I will also make the very obvious point that, although the Samoan word /iʔa/ 'fish' (17) and the Hawaiian word /hiʔa/ 'firemaking' (20) are similar in shape, they are not considered to be cognates because their meanings are totally different.)

5.2.3 Sound Correspondences

Having completed the first step, you are now ready to move on to step two. The second step is to set out the complete set of sound correspondences. When I talk about a SOUND CORRESPONDENCE, I mean that we try to find each set of sounds that appears to be descended from the same original sound. So, if you take the first word in the list that I have given, you will find the following correspondences between the sounds:

| | | | | |
|------------|---|---|---|---|
| Tongan | t | a | p | u |
| Samoan | t | a | p | u |
| Rarotongan | t | a | p | u |
| Hawaiian | k | a | p | u |

You can see that there is an initial correspondence of /t/ in Tongan to /t/ in Samoan, to /t/ in Rarotongan, and to /k/ in Hawaiian. The /a/ in Tongan corresponds to an /a/ in all of the remaining three languages. Similarly, there is a correspondence of /p/ in all four languages, and, finally, there is a correspondence of /u/ in all four languages. These correspondences can be set out like this:

| Tongan | Samoan | Rarotongan | Hawaiian |
|--------|--------|------------|----------|
| t | t | t | k |
| a | a | a | a |
| p | p | p | p |
| u | u | u | u |

What you have to do is list *all* such sound correspondences that are present in the whole of the data.

Actually, a quick examination of the vowel correspondences reveals that the vowels are identical in all four languages in all words. (Don't let this make you think that for other languages the vowels will be as straightforward as this! Sometimes it will be the vowels rather than the consonants which have the most complicated sets of sound correspondences. Usually, both consonants and vowels will exhibit variations in their correspondence sets.) In order to be completely thorough, I will set out the vowel correspondences for you, even though there are no differences between the four languages:

| Tongan | Samoaan | Rarotongan | Hawaiian |
|--------|---------|------------|----------|
| a | a | a | a |
| e | e | e | e |
| i | i | i | i |
| o | o | o | o |
| u | u | u | u |

Let us now concentrate on the consonant correspondences, which is where the differences are to be found between these languages. The correspondence sets for consonants work out to be as follows:

| Tongan | Samoaan | Rarotongan | Hawaiian |
|--------|---------|------------|----------|
| p | p | p | p |
| f | f | ? | h |
| t | t | t | k |
| k | ? | k | ? |
| h | s | ? | h |
| m | m | m | m |
| n | n | n | n |
| ŋ | ŋ | ŋ | n |
| v | v | v | w |
| l | l | r | l |
| ∅ | l | r | l |

There is one brief point that I should make before continuing, and this concerns the use of the zero symbol \emptyset . This symbol is used to express correspondences such as the following in the word for 'feces' (5):

| | | | | |
|------------|---|---|---|---|
| Tongan | t | a | ʔ | e |
| Samoan | t | a | | e |
| Rarotongan | t | a | | e |
| Hawaiian | k | a | | e |

In these forms, the /ʔ/ in Tongan corresponds to the absence of any sound in the other three languages. Thus, you will need to set this correspondence out as follows:

| | | | |
|--------|--------|------------|----------|
| Tongan | Samoan | Rarotongan | Hawaiian |
| ʔ | ∅ | ∅ | ∅ |

Similarly, in the word for 'gall' (11), you will see that there are two sounds in Tongan corresponding to nothing in the other languages:

| | | | | |
|------------|---|---|---|---|
| Tongan | ʔ | a | h | u |
| Samoan | | a | | u |
| Rarotongan | | a | | u |
| Hawaiian | | a | | u |

The word-initial correspondence of ʔ : ∅ : ∅ : ∅ is the same correspondence I have just set out for the medial consonant in the word for 'feces'. (Note that I have just used a slightly different way of expressing sound correspondences, using the : symbol. From now on, I will use both methods interchangeably.) The word-medial correspondence is a different one, which we can set out as follows:

h : ∅ : ∅ : ∅

One problem that you might face in drawing up your set of sound correspondences is that, in cases where you have had to exclude a form in one or more languages because it is not cognate, you might have some correspondences that appear to be incomplete. For instance, go back to cognate set (20) in the list. If we are to exclude the Tongan word /tafuafi/ because it is not cognate with Samoan /siʔa/, Rarotongan /ʔika/, and Hawaiian /hiʔa/, then we could be faced with gaps in the Tongan data for some sounds. In this case, however, it is not too difficult to fill in the gaps as there are plenty of other words in Tongan that contain sounds which are cognate with words in the other languages in which the same correspondences occur. Where Samoan has /s/, Rarotongan has /ʔ/, and Hawaiian has /h/, other cognate sets, such as (18), (19), and (21), indicate that Tongan has /h/. For the intervocalic consonant, the cognate sets numbered (9), (10), (16), (17), (21), (22), and (31) all indicate that Tongan has /k/. You already know that the vowels in all

four languages are identical in all words. So, while the Tongan word for 'firemaking' is /tafuafi/, if Tongan had retained the original word, we can predict that its shape would have been /hika/. Of course, we must not add this word into our data (though we might find somewhere else in the vocabulary of Tongan that /hika/ is found, but that it has shifted in its meaning so that it was not originally spotted as a possible cognate). If we did not have all of these other sets of cognate forms which indicate what sound corresponds in a particular language to the sounds in the other languages in a family, then it might be necessary simply to leave the slot for that sound in that language blank.

5.2.4 Reconstruction Principles

Having set out all of the sound correspondences that you can find in the data, you can now move on to the third step, which is to work out what original sound in the protolanguage might have produced that particular range of sounds in the various daughter languages. Your initial assumption should be that each separate set of sound correspondences goes back to a distinct original phoneme. In reconstructing the shapes of these original phonemes, you should always be guided by some general principles:

1. Any reconstruction should involve sound changes that are plausible, unless there is good evidence to the contrary. (You should be guided by the kinds of things that you learned in chapter 2 in this respect.)
2. Any reconstruction should involve as few changes as possible between the protolanguage and the daughter languages.²

It is perhaps easiest to reconstruct back from those sound correspondences in which the reflexes of the original phoneme (or protophoneme) are identical in all daughter languages. By guiding principle 2, you should normally assume that such correspondences go back to the same protophoneme as you find in the daughter languages and that there have been no sound changes. Thus, you should assume that the vowels of Proto-Polynesian are exactly the same as you find in the four daughter languages that we are looking at. So, for the correspondence *a : a : a : a* you should reconstruct an original /*a/, for *e : e : e : e* you should reconstruct /*e/, and so on.³

Turning our attention now to the consonant correspondences, it will also be easiest to deal with those correspondences in which the daughter languages all have the same reflex. Such correspondences include the following:

| Tongan | Samoa | Rarotongan | Hawaiian |
|--------|-------|------------|----------|
| p | p | p | p |
| m | m | m | m |
| n | n | n | n |

Again, you need to ask yourself the question, What protophoneme could reasonably be expected to have produced a /p/ in all of the daughter languages? The obvious answer is again /*p/. Applying the same reasoning, you can also reconstruct /*m/ and /*n/ for the other two correspondences that I have just listed.

The next thing that you should do is look at sound correspondence sets that only have slight differences between the various daughter languages and try to reconstruct original phonemes from the evidence that these provide. So, from the correspondence sets that I listed earlier, we will now go on to look at the following:

| Tongan | Samoa | Rarotongan | Hawaiian |
|--------|-------|------------|----------|
| t | t | t | k |
| ŋ | ŋ | ŋ | n |

In these two cases, only one language, Hawaiian, differs from the other three languages. Logically, in the first case, you could reconstruct either a /*t/ or a /*k/. Which would be the best solution? It would obviously be better to reconstruct /*t/ as the original form and to argue that this changed to /k/ in Hawaiian. To suggest /*k/ as the original form, you would need to say that this changed to /t/ in three separate languages. So, in keeping with guiding principle 2, you will often reconstruct as the original form the sound that has the widest distribution in the daughter languages. Using the same argument, you should reconstruct /*ŋ/ for the second correspondence set presented above.⁴

You should now go on to deal with those correspondences that have a greater amount of variation in the reflexes of the original phoneme. Where there is greater variation, it is going to require greater consideration on your part in doing the reconstruction. Let us take the following correspondence:

| Tongan | Samoa | Rarotongan | Hawaiian |
|--------|-------|------------|----------|
| k | ? | k | ? |

Here there are two instances of /k/ in the daughter languages and two of /?/, so the second guiding principle will no longer help us as there is no single reflex with a wider distribution than other reflexes among the daughter languages. We are therefore torn between reconstructing /*k/ and /*?/.

However, you should also remember that you are to be guided by principle 1. This guiding principle says that you should prefer a solution that involves "natural" sound change over an "unnatural" one. If you were to propose an original /*k/ rather than /*?/, you would need to say that the following change took place in Samoan and Hawaiian:

*k > ?

This is a well-known sound change that goes under the general heading of weakening or lenition. However, if you were to reconstruct, instead, */*ʔ/* for this correspondence, you would need to say that in Tongan and Rarotongan, the following change took place:

*/*ʔ > k*

While this is not an impossible change, it is certainly a rarer kind of change than the change of */k/* to *ʔ/*. Thus, according to guiding principle 1, you should probably reconstruct */*k/* in this case.

At this point, I will add a third guiding principle:

3. Reconstructions should fill gaps in phonological systems rather than creating unbalanced systems.

Although there will always be exceptions among the world's languages, there is a tendency for languages to have "balanced" phonological systems. By this I mean that where there is a set of sounds distinguished by a particular feature, this feature is also likely to be used to distinguish a different series of sounds in the language. For example, if a language has two back-rounded vowels (i.e., */u/* and */o/*), we would expect it also to have two front-unrounded vowels (i.e., */i/* and */e/*).⁵ Thus, the following represent balanced phoneme inventories, and these kinds of inventories tend to recur in the world's languages:

| | Front | Back |
|------|-------|------|
| High | i | u |
| Low | | a |

| | Front | Back |
|------|-------|------|
| High | i | u |
| Mid | e | o |
| Low | | a |

| | Front | | Back |
|------|---------|-------|------|
| | Unround | Round | |
| High | i | y | u |
| Mid | e | ø | o |
| Low | | a | |

The following, however, are "unbalanced" systems and are less likely to occur than systems such as those I have just given, as they contain gaps (indicated by dashes):

| | Front | Back |
|------|-------|------|
| High | i | — |
| Mid | e | o |
| Low | a | |

| | Front | | Back |
|------|---------|-------|------|
| | Unround | Round | |
| High | i | — | u |
| Mid | e | ø | o |
| Low | a | | |

You can also use guiding principle 3 to help in reconstructing the original phoneme from which the $k : ? : k : ?$ correspondence is derived. The correspondences that you have already looked at provide evidence for the reconstruction of the following original consonant phonemes in Proto-Polynesian:

| | Bilabial | Alveolar | Velar |
|-------|----------|----------|-------|
| Stop | *p | *t | |
| Nasal | *m | *n | *ŋ |

If you assume that languages operate in terms of balanced phonological systems, you would not expect to find a gap at the velar stop position (i.e., /k/) in the protolanguage, since you already have evidence for the existence of a velar nasal. As you are, in a sense, “looking for” a /*k/, you can use this fact as evidence in support of your reconstruction of /*k/ rather than /*?/ for this particular sound correspondence.

Let us now take the next problematic correspondence:

| Tongan | Samoan | Rarotongan | Hawaiian |
|--------|--------|------------|----------|
| f | f | ʔ | h |

This correspondence, in fact, is less problematic than the one you just looked at. Because there are a greater number of /f/ reflexes of this original phoneme than other sounds, by our second guiding principle again, you should reconstruct an /*f/ wherever this correspondence occurs. Furthermore, /*f/ > /h/ (and /*f/ > /ʔ/ is a fairly common change, whereas /*h/ or /ʔ/ > /f/ would require a more complex explanation. Unless you have evidence to the contrary (such as having already reconstructed /*f/ for another correspondence set), /*f/ is the sound to reconstruct here.

Now let us consider the correspondences involving the liquids:

| Tongan | Samoaan | Rarotongan | Hawaiian |
|--------|---------|------------|----------|
| l | l | r | l |
| ø | l | r | l |

We appear to face real problems here. We have to reconstruct two different phonemes in order to account for the two different sets of correspondences, but there is very little difference between the reflexes of these two protophonemes in the daughter languages. Three of the four languages are identical in their reflexes of these sounds, and in both sets of correspondences, /l/ is the most common reflex. Since we have to reconstruct two phonemes, we will presumably have to choose /*l/ for one and /*r/ for the other. But which will we assign to which correspondence set? The choice is fairly arbitrary. However, we could argue that loss of /*r/ is possibly slightly more likely to occur than a change of /*r/ to /*l/, so we could suggest that /*l/ is the source of the first correspondence set, and /*r/ is the source for the second correspondence set. If this is correct, we would need to say that Rarotongan underwent a change of /*l/ to /r/, while Samoaan and Hawaiian underwent a change of /*r/ to /l/, and Tongan simply lost the original /*r/ phoneme.

With this pair of reconstructions, we really are on shaky ground, and we are operating with little more than guesswork. One way of checking the accuracy of our reconstruction would be to broaden the data upon which the reconstruction is based by introducing forms from a wider range of related languages. If it turns out that by considering a larger number of Polynesian languages, we find greater numbers of lateral reflexes of our suggested /*l/ reconstruction and a greater number of rhotic reflexes of our suggested /*r/ reconstruction, this would be evidence in support of our conclusion.

We have reconstructed the following consonant inventory for Proto-Polynesian so far:

*p *t *k
 *m *n *ŋ
 *f
 *l
 *r

Now we turn to the correspondences involving the glottal sounds:

| Tongan | Samoaan | Rarotongan | Hawaiian |
|--------|---------|------------|----------|
| ʔ | ø | ø | ø |
| h | ø | ø | ø |

When we set the correspondences out like this, it is clear that Tongan is the only language to have any reflexes of these two phonemes. All of the other languages have lost them altogether. It is not too difficult to argue, therefore, that we should reconstruct */*ʔ/* and */*h/*, respectively, wherever we find these correspondences, especially since */ʔ/* and */h/* are sounds that are very commonly lost in languages. We might also note that this correspondence gives us more evidence that our reconstruction for */*k/* (rather than */*ʔ/*) above is correct.

We have yet to consider the following correspondence, however:

| Tongan | Samoa | Rarotongan | Hawaiian |
|--------|-------|------------|----------|
| h | s | ʔ | h |

Here, all of the languages except Samoan reflect a glottal sound, and */h/* is the most common reflex. However, we have already reconstructed */*h/* for the correspondence *h : ʔ : ʔ : ʔ* that I just presented. Similarly, */*ʔ/* is not a possible reconstruction, because we have already reconstructed this to account for the *ʔ : ʔ : ʔ : ʔ* correspondence set. The only possibility left seems to be to reconstruct this correspondence as deriving from */*s/*. This is actually quite reasonable. Changes of the following type are quite common in languages of the world:

**s > h > ʔ*

It is also relatively uncommon for languages to have no */s/* phoneme, especially when they have other fricatives, so this is a sound that we would normally expect to find evidence for in any protolanguage.

Both the change of */s/* to */h/* and the change of */h/* to */ʔ/* can be regarded as weakening, or lenition. Furthermore, if you did not reconstruct an */*s/* in Proto-Polynesian, you would end up with a gap in the phoneme inventory. By reconstructing an */*s/*, you would be filling the voiceless alveolar fricative slot, so that you have an inventory that looks like this:

*p *t *k *ʔ
 *m *n *ŋ
 *f *s *h
 *ʃ
 *r

(Note that a glottal nasal is a physical impossibility, and it is more common for languages to have the phoneme */h/* than */x/*, so the lack of a sound in the velar fricative slot is not a real problem, either.)

It could perhaps be argued, instead, that the two correspondences involving /h/ discussed earlier need to be completely reexamined:

| Tongan | Samoaan | Rarotongan | Hawaiian |
|--------|---------|------------|----------|
| h | ø | ø | ø |
| h | s | ʔ | h |

According to guiding principle 2, you should reconstruct as the phoneme in the protolanguage the form that has the widest distribution in the daughter languages. You might, therefore, want to reconstruct the phoneme /*h/ instead of /*s/ for the second of these correspondence sets. This would be phonetically quite reasonable according to guiding principle 1, but doing this would create problems for your handling of the first correspondence for which you have already reconstructed /*h/.

This problem could be overcome by suggesting a separate original phoneme to account for this correspondence, perhaps the voiceless velar fricative /*x/. Although this would be phonetically reasonable as well, I would argue against this solution on the grounds that it would violate a fourth guiding principle that can be set out as follows:

4. A phoneme should not be reconstructed in a protolanguage unless it is shown to be absolutely necessary from the evidence of the daughter languages.

None of the daughter languages anywhere has an /x/, so you should be automatically suspicious of a solution that suggests an /x/ in the protolanguage. Keeping this in mind, then, you should reject the revised solution and stick with the original solution.

Finally, we have the following correspondence:

| Tongan | Samoaan | Rarotongan | Hawaiian |
|--------|---------|------------|----------|
| v | v | v | w |

While we would predict /*v/ as the most likely original form for this correspondence on the basis of the distribution of its reflexes, by doing so we would create an uneven phonemic inventory. As it stands, there is no voiced/voiceless contrast in the stop or fricative series of Proto-Polynesian (all are voiceless), and to introduce a single voiced sound here would seem odd. Another odd thing about the phoneme inventory so far reconstructed for Proto-Polynesian is the lack of semivowels. We would therefore probably be more justified in reconstructing /*w/ than /*v/ in this case. The complete original phoneme inventory that we have reconstructed for Proto-Polynesian now looks something like this:

| | | | |
|----|----|----|----|
| *p | *t | *k | *ʔ |
| *m | *n | *ŋ | |
| *f | *s | *h | |
| | *l | | |
| | *r | | |
| *w | | | |
| | *i | *u | |
| | *e | *o | |
| | a | | |

Having arrived at the phoneme inventory of Proto-Polynesian by comparing the daughter languages, you can now move on to the comparatively simple task of reconstructing the forms of the individual words. To do this, you need to list the sound correspondences and set out the original phoneme that each of these goes back to (table 5.2).

TABLE 5.2 Table of Correspondences

| | Tongan | Samoan | Rarotongan | Hawaiian |
|------------|--------|--------|------------|----------|
| Vowels | | | | |
| *a | a | a | a | a |
| *e | e | e | e | e |
| *i | i | i | i | i |
| *o | o | o | o | o |
| *u | u | u | u | u |
| Consonants | | | | |
| *p | p | p | p | p |
| *f | f | f | ʔ | h |
| *t | t | t | t | k |
| *k | k | ʔ | k | ʔ |
| *s | h | s | ʔ | h |
| *ʔ | ʔ | ∅ | ∅ | ∅ |
| *h | h | ∅ | ∅ | ∅ |
| *m | m | m | m | m |
| *n | n | n | n | n |
| *ŋ | ŋ | ŋ | ŋ | n |
| *w | v | v | v | w |
| *l | l | l | r | l |
| *r | ∅ | l | r | l |

5.2.5 Residual Issues

Using the information that is set out in table 5.2, let us try to reconstruct the word for 'four', which is item 15 in the original list of cognates in table 5.1. The reflexes in the daughter languages of the original word that you are trying to reconstruct are as follows:

| | | | |
|------------|---|---|---|
| Tongan | f | a | a |
| Samoan | f | a | a |
| Rarotongan | ? | a | a |
| Hawaiian | h | a | a |

As you have a word containing three sound correspondences, this indicates that the original word must have had three original phonemes. What were those original phonemes? The $f : f : ? : h$ correspondence, if you check from table 5.2, goes back to an original $/*f/$. The two $a : a : a : a$ correspondences point to an original $/*a/$. So the Proto-Polynesian word for 'four' can be reconstructed as $/*faa/$.

Now take item 9 in the list of cognates in table 5.1, which gives the various words for 'dodge'. This involves the following correspondences between the four languages:

| | | | | |
|------------|---|---|---|---|
| Tongan | k | a | l | o |
| Samoan | ? | a | l | o |
| Rarotongan | k | a | r | o |
| Hawaiian | ? | a | l | o |

Again, referring to the list of correspondences in table 5.2, you will find that the $k : ? : k : ?$ correspondence points to an original $/*k/$. The $a : a : a : a$ correspondence, of course, goes back to $/*a/$. Table 5.2 reveals that $l : l : r : l$ goes back to $/*l/$, and finally $o : o : o : o$ goes back to $/*o/$. So, you can reconstruct the original word for 'dodge' in Proto-Polynesian as $/*kalo/$.

Although reconstruction of the vocabulary is relatively simple and straightforward, there are some situations where you cannot be sure of the original form. If you consider the following example from table 5.1, it should be clear why this is so:

| | | | | | |
|----|--------|--------|------------|----------|----------|
| | Tongan | Samoan | Rarotongan | Hawaiian | |
| 8. | malohi | malosi | kaʔa | ʔaha | 'strong' |

Here you have *two* clear cognate sets, and both could equally well be reconstructed back to the protolanguage. On the basis of the Tongan and Samoan

forms, you would be tempted to reconstruct an original word of the form */*malosi/*, while on the basis of the Rarotongan and Hawaiian data, you would need to reconstruct either */*kasa/* or */*kafa/*. All you can do in such cases is reconstruct *both* forms and indicate that one of them probably meant something different, but similar ('hard', for instance). But which of the two was the original word for 'strong' is impossible to say, on the basis of the evidence that you have. The only way to solve this problem would be to look at the word for 'strong' in a larger number of Polynesian languages.

Another problem that you will sometimes face in reconstructing vocabulary comes when you have incomplete sound correspondences that you are unable to fill from other correspondence sets in the languages that you are examining. For instance, imagine that you had only the following forms:

| | Tongan | Samoa | Rarotongan | Hawaiian | |
|----|--------|-------|------------|----------|---------|
| 9. | — | ?alo | karo | ?alo | 'dodge' |

If you did not have a cognate in Tongan (either because the meaning 'dodge' is expressed by a completely different form or because the data itself may be lacking the appropriate form), then you would not be able to reconstruct a single original form to express this meaning. This is because the correspondence of Samoan */l/* to Rarotongan */r/* and Hawaiian */l/* could point equally well to the reconstruction of both */*l/* and */*r/*. In order to be able to decide whether the form should be reconstructed as having */*r/* or */*l/*, a Tongan cognate is essential, as this is the only daughter language that still makes a distinction between the two original phonemes. If we are faced with a genuine ambiguity in our reconstructions, we can indicate this by showing that we aren't sure what the original phoneme was. So, we could give */*ka(l/r)o/* or */*kaLo/*, which would be alternative ways of saying that the evidence points to either */*kalo/* or */*karo/*, and there is no way of making a choice between the two. Similarly, on the basis of the forms */ka?a/* 'strong' in Rarotongan and *?aha/* in Hawaiian, all we can do is reconstruct */*ka(s/f)a/* or */*ka(S)a/*.⁶

Of course, if you refer back to item 9 in the original list of cognate sets (table 5.1), the Tongan form actually is cognate, and the Tongan word for this meaning is */kalo/*. This indicates that the reconstructed form is unambiguously */*kalo/* rather than */*karo/*.

5.3 RECONSTRUCTION OF CONDITIONED SOUND CHANGES

When you write the rules for the changes from Proto-Polynesian into the various daughter languages, you will find that all of the changes that have

taken place are unconditioned sound changes. That is, an original /*s/ always becomes /ʔ/ in Rarotongan, or an original /*r/ always becomes // in Hawaiian. There are no conditioned changes that have taken place only in certain environments and not in others. How does it affect our technique of reconstruction if there are conditioned sound changes involved, as well as unconditioned sound changes?

Let us look at some additional data from Tongan and Samoan:

| | Tongan | Samoan | |
|-----|-----------|-----------|---------------------|
| 37. | fefine | fafine | 'woman' |
| 38. | fiefia | fiafia | 'happy' |
| 39. | moʔuŋa | mauŋa | 'mountain' |
| 40. | tuŋaʔane | tuŋane | '(woman's) brother' |
| 41. | tuofefine | tuafafine | '(man's) sister' |

The vowel correspondences that we noted before were completely uniform through all of the languages that we looked at. Thus, on the basis of the correspondence *a* : *a* we reconstructed /*a/, while *e* : *e* points to /*e/, and *o* : *o* points to /*o/. However, these new examples point to two new sets of vowel correspondences:

| Tongan | Samoan |
|--------|--------|
| e | a |
| o | a |

Must you therefore reconstruct two separate phonemes for these two correspondence sets? If you do, they will certainly need to be phonetically similar to the vowels /e/, /o/, and /a/, yet at the same time they would need to be different from these three vowels. If you retain these three vowels, then you could cater for these additional correspondence sets by reconstructing something like /ɛ/ for the *e* : *a* correspondence, and /ɔ/ for the *o* : *a* correspondence. Your reconstructions for these additional words would end up looking like this:

| | | |
|-----|--------------|---------------------|
| 37. | *fɛfine | 'woman' |
| 38. | *fiɛfia | 'happy' |
| 39. | *mɔʔuŋa | 'mountain' |
| 40. | *tuɔŋa(ʔa)ne | '(woman's) brother' |
| 41. | *tuɔfɛfine | '(man's) sister' |

One problem with this reconstruction is that you will have violated the general principle that we should not normally reconstruct a phoneme if that

phoneme does not occur in any of the descendant languages. Since none of the Polynesian languages that we have been looking at has a contrast between /e/ and /ɛ/, or between /o/ and /ɔ/, we should be suspicious of a reconstruction that suggests such a distinction in the protolanguage.

If you examine the distribution of the suggested reconstructed sounds /*ɛ/ and /*ɔ/ with respect to /*a/, you will find that there is, in fact, complementary distribution. The reconstructed sound /*ɛ/ only ever occurs in the third syllable from the end of a word when the following syllable contains the high front vowel /i/, while /*ɔ/ only occurs in the third syllable from the end of a word when the following syllable contains the high front vowel /u/. The vowel /*a/, however, appears in all other environments. To see this, compare the forms that you have just examined with the following:

| | Tongan | Samoaan | |
|-----|--------|---------|-------------|
| 1. | tapu | tapu | 'forbidden' |
| 5. | taʔe | tae | 'feces' |
| 6. | taʔata | taʔata | 'man' |
| 7. | tahi | tai | 'sea' |
| 8. | malohi | malosi | 'strong' |
| 9. | kalo | ʔalo | 'dodge' |
| 10. | aka | aʔa | 'root' |
| 11. | ʔahu | au | 'gall' |
| 14. | afi | afi | 'fire' |

This list does not include all of the examples from the original set of cognates between the two languages, but if you carefully go through the entire list, you will find that there are no examples in Samoan which end in either /-aCuCV/ or /-aCiCV/.⁷

What you must do is look for evidence of COMPLEMENTARY DISTRIBUTION between phonetically similar correspondence sets before you do your final reconstruction. The correspondence set *e* : *a* occurs only in the third syllable from the end of a word when the following vowel correspondence involves the high front vowel /i/, while the correspondence *o* : *a* occurs only in the third syllable from the end of a word when the following vowel correspondence involves /u/. The correspondence set *a* : *a* appears in all other environments. You therefore need to reconstruct only a single phoneme for these three correspondence sets. You will not need to modify your reconstruction of /*a/, and there is certainly no need to reconstruct /*ɛ/ or /*ɔ/, as Tongan has undergone a conditioned change of the following form:

$$a > \begin{cases} o / _ CuCV \\ e / _ CiCV \end{cases}$$

Therefore, after you have set out your sound correspondences between the daughter languages, you must also do the following, as the fifth and sixth steps in applying the comparative method:

5. Look for sound correspondences that involve phonetically similar sounds.
6. For each of these phonetically "suspicious" pairs of sound correspondences, you should try to see whether they are in complementary or contrastive distribution.

This is very similar to what we do in a synchronic analysis of the phonemes of a language, except that here we are trying to analyze the phonemes of the protolanguage by using the sound correspondences as the "phonetic" raw data. We then have to decide which sound correspondences are phonemically distinctive in the original language and which are just positional variants (or "allo-correspondences" of "correspondence-emes").

Let us look at another very simple situation that we are already familiar with to see how to proceed when it comes to reconstructing conditioned sound changes. We have already seen that in the Motu language of Papua New Guinea, there has been change of */*t/* to *[s]* before the front vowels, while in all other environments it remained as *[t]*. We wrote this rule formally as follows:

$$*t > s / _ _ V_{\text{front}}$$

Rather than working from the protolanguage to the modern language, let us instead work back from Motu and one of its sister languages, applying the comparative method that we have been discussing in this chapter. The sister language that we will look at is Sinaugoro, and the data from these two languages that we will consider is as follows:⁸

| Sinaugoro | Motu | |
|-----------|------|---------------|
| tama | tama | 'father' |
| tina | sina | 'mother' |
| tayi | tai | 'cry' |
| tui | tui | 'elbow, knee' |
| yita | ita | 'see' |
| yate | ase | 'liver' |
| mate | mase | 'die' |
| natu | natu | 'child' |
| toi | toi | 'three' |

Let us apply the technique that I have just shown you. First, remember that you have to sort out the cognate forms from the noncognate forms. In this case, I have already done this, and all of the forms that are given are cognate. The second step, then, is to set out the sound correspondences. Since

you are only interested at this stage in the history of [t] and [s], you should restrict yourself only to correspondences involving these two sounds. (There are many other correspondences in the two languages where the two sounds are identical, of course, and there is also a correspondence of Sinaugoro /ɣ/ and /ŋ/ to Motu /ø/ (that is, Sinaugoro has a sound where Motu has nothing). The correspondences that we can find are:

| Sinaugoro | Motu |
|-----------|------|
| t | t |
| t | s |

There are therefore two sound correspondences here. Does this mean that you should reconstruct two separate phonemes in the original language? If you did, these would presumably be /*t/ for the first correspondence and /*s/ for the second correspondence.

However, since the *t : t* and the *t : s* correspondences both involve similar sets of sounds, you should first look for any evidence that there might be complementary distribution involved. If you cannot find any evidence of complementary distribution, then you should also look for direct evidence of contrastive distribution. What you will find when you examine the data is that the *t : s* correspondence occurs only when there is a following correspondence of front vowels (i.e., *i : i* or *e : e*), whereas the *t : t* correspondence occurs before all other vowel correspondences. If two (or more) correspondence sets are in complementary distribution in this way, then you should reconstruct only a single original phoneme for both correspondences, and we again say that a conditioned sound change must have taken place.

In this case, you would want to reconstruct a /*t/ using the principle that you should normally reconstruct the form that has the widest distribution in the daughter languages. You then need to say that a conditioned sound change took place in Motu whereby /*t/ became [s] before front vowels, as you saw earlier. The protoforms from which the Sinaugoro and Motu forms were derived can therefore be reconstructed as follows (with the *ŋ : ø* correspondence presumably coming from /*ŋ/ and the *ɣ : ø* correspondence coming from *ɣ):

| | |
|-------|---------------|
| *tama | 'father' |
| *tina | 'mother' |
| *taji | 'cry' |
| *tui | 'elbow, knee' |
| *ɣita | 'see' |
| *ɣate | 'liver' |
| *mate | 'die' |
| *natu | 'child' |
| *toi | 'three' |

With these reconstructed forms, it is obvious that Sinaugoro directly reflects the original forms without change, with Motu being the only innovating language.

Now that you know that you must check phonetically similar sets of sound correspondences for complementary or contrastive distribution, you should go back and check your Polynesian correspondences as well. Which correspondences should you check for complementary distribution because of their phonetic similarity? The first obvious pair of correspondences that you should test are those involving the liquids, for which our earlier reconstructions were as follows:

| | Tongan | Samoa | Rarotongan | Hawaiian |
|----|--------|-------|------------|----------|
| *l | l | l | r | l |
| *r | ø | l | r | l |

Has there been a conditioned sound change in Tongan in which a single original phoneme was lost in some environments and retained in others? Or were there indeed two separate proto-phonemes which have merged in Samoan, Rarotongan, and Hawaiian?

To test these two possibilities, I will list the full cognate sets in which these forms occur:

| | Tongan | Samoa | Rarotongan | Hawaiian | |
|-----|----------|----------|------------|----------|----------------------|
| | <i>l</i> | <i>l</i> | <i>r</i> | <i>l</i> | |
| 9. | kalo | ?alo | karo | ?alo | 'dodge' |
| 12. | ?ulu | ulu | uru | — | 'head' |
| 29. | ŋalu | ŋalu | ŋaru | nalu | 'wave' |
| 33. | laho | laso | ra?o | laho | 'scrotum' |
| 34. | lohu | lou | rou | lou | 'fruit-picking pole' |
| | <i>ø</i> | <i>l</i> | <i>r</i> | <i>l</i> | |
| 23. | huu | ulu | uru | — | 'enter' |
| 26. | maa | mala | mara | mala | 'fermented' |
| 35. | oŋo | loŋo | roŋo | lono | 'hear' |
| 36. | ua | lua | rua | lua | 'two' |

You will need to test all possible conditioning environments. You should remember from your study of phonology that when you are looking for possible conditioning factors for allophones of phonemes, you need to consider the following:

1. The nature of the sound (or sounds) which follow
2. The nature of the sound (or sounds) which precede

3. The nature of the syllable (i.e., whether open or closed)
4. The position in the word (i.e., whether initial, medial, or final)
5. Any possible combination of such conditioning factors

Let us consider these possible conditioning factors to see if these two sets of correspondences are in complementary or contrastive distribution.

First, let us look at the nature of the following sound. Immediately after the first set of correspondences (i.e., $l : l : r : l$), you will find the following correspondence sets:

u : u : u : u
 a : a : a : a
 o : o : o : o

After the second set of correspondences (i.e., $\emptyset : l : r : l$), you will find the following vowel correspondences:

u : u : u : u
 a : a : a : a
 o : o : o : o

In fact, you have exactly the same sets of sound correspondences occurring after both liquid correspondences. In order to demonstrate the fact that there is no complementary distribution, you only need overlap in the two sets of environments with respect to a single correspondence, and here you have all three sets of following environments being the same.

Of course, you also have to check all other possible conditioning factors now that you have checked the following sound, so let us try to find out if it is the nature of the preceding correspondence which acts as a conditioning factor. Before the $l : l : r : l$ correspondence, you will find the following vowel correspondence sets:

u : u : u : u
 a : a : a : a

Before the second correspondence, you will find the following:

u : u : u : u
 a : a : a : a

Again, exactly the same two sets of vowel correspondences appear before the two correspondence sets that you are checking, so there is no complementary distribution with respect to this environment, either. The third possibility (i.e., whether the syllable is open or closed) is of little use to you here because

all of the syllables in these languages are open. You should check the position in the word. When you do this, you will find that the two sets of correspondences occur both initially and medially. Finally, you should consider the possibility of there being some more complex conditioning factors. However, none is apparent.

This evidence means that you are forced to conclude that the two correspondence sets involving liquids are in contrastive distribution and that you were correct in the first place in reconstructing two separate phonemes. In fact, you can even find a subminimal pair of words from the data that I have presented in order to back up this conclusion. (No complete minimal pairs are available, but perhaps if more data were available we would be able to find one.) Compare the forms for 'head' and 'enter'; remember that a dash means that there is no comparable cognate form in the language:

| | Tongan | Samoaan | Rarotongan | Hawaiian | |
|-----|--------|---------|------------|----------|---------|
| 12. | ʔulu | ulu | uru | — | 'head' |
| 23. | huu | ulu | uru | — | 'enter' |

Between the correspondences in which all of the languages have /u/, we find that both correspondence sets occur. Thus, Tongan has the sequence /ulu/ contrasting with /uu/. So, you can conclude that there was a phonemic distinction in the original language that goes back to an original subminimal pair: that is, /*ʔulu/ 'head' versus /*huru/ 'enter'. Although Samoaan, Rarotongan, and Hawaiian have all unconditionally merged the original distinction between /*l/ and /*r/, the original opposition is still reflected in Tongan, which has retained the /*l/ and unconditionally lost the /*r/.

Here I have described a means of reconstructing the phonological system of a protolanguage and also its lexicon. We call this method of reconstruction the "comparative method." The comparative method involves carefully carrying out all of the following steps:⁹

1. Sort out those forms that appear to be cognate and set aside the noncognate forms.
2. Write out the full set of correspondences between the languages you are looking at (including correspondences where the sounds are identical all the way through). Be careful to note correspondences where a sound in one language corresponds to \emptyset (or the absence of a sound) in another language.
3. Group together all correspondences that have reflexes that are phonetically similar.
4. Look for evidence of complementary and contrastive distribution between these suspicious pairs of correspondences.
5. For each correspondence set that is not in complementary distribution with another correspondence set, assume that it goes back to a separate original phoneme.

6. Make an estimation about the original form of the phoneme using the following criteria:
 - (a) The proposed original phoneme must be plausible, meaning that the changes from it to the reflexes in the descendant languages must fit our knowledge about what kinds of sound changes are common in the world's languages.
 - (b) The sound that has the widest distribution in the daughter languages is most likely to be the original phoneme.
 - (c) A sound corresponding to a gap in the reconstructed phoneme inventory of the protolanguage is also likely to be a possible reconstruction for one of the correspondence sets.
 - (d) A sound that does not occur in any of the daughter languages should not be reconstructed unless there are very good reasons for doing so.
7. For each group of correspondence sets that are in complementary distribution, assume that they all go back to a single protophoneme, and use the same criteria given in item 6 in this list to reconstruct its shape.

5.4 THE REALITY OF PROTOLANGUAGES

At the beginning of this chapter on the comparative method, I said that the method involved a certain amount of guesswork but that this guesswork was intelligent rather than a shab in the dark. But what do our reconstructions actually represent? Do they represent a real language as it was actually spoken at some earlier time, or do our reconstructions only give an approximation of some earlier language?

One point of view is that we are not actually trying to reconstruct the facts of a language as it was actually spoken when we are applying the comparative method—nor should we even try to do this. Some linguists argue that we should not try to suggest any phonetic form of reconstructed original phonemes deduced from the evidence of sound correspondences between daughter languages. Rather, what we should do is simply deduce that, in a particular word, there was a phoneme that was distinct from all other sounds but we do not know exactly what its phonetic form was. According to this point of view, a “protolanguage” as it is reconstructed is not a “language” in the same sense as any of its descendant languages or as the “real” protolanguage itself. It is merely an abstract statement of correspondences.

Other linguists, while not going as far as this, have stated that, while languages that are related through common descent are derived from a single ancestor language, we should not necessarily assume that this language really existed as such. This method allows us to derive a set of hypotheses about

the protolanguage, but there are numerous ways in which we might run into problems. We might not be able to reconstruct all the changes (if a change happened in all the languages, we might not be able to uncover it). We might not be able to recover allophony in the protolanguage. There are times when we simply cannot be sure what the original phonetics of the forms were. A good case is the difference between the Polynesian $l : l : r : l$ and $o : l : r : l$ correspondences that we looked at earlier. We reconstructed $/*l/$ for the first of these correspondences and $/*r/$ for the second. However, it is quite possible that we are wrong. In such cases as these, it would be wiser to regard $/*l/$ and $/*r/$ not so much as reliable phonetic indications of the original forms but simply as indications that there was a phonemic distinction of some sort (probably involving liquids).

Sometimes linguists prefer to avoid making a commitment to a particular phonetic shape for a protophoneme but at the same time want to avoid assigning totally arbitrary symbols to account for a set of sound correspondences in the daughter languages. One frequently employed device in these sorts of situations is to distinguish the protophonemes by which two phonetically similar correspondence sets are derived by using lowercase and uppercase forms of the same symbol. In the example that I have just given, for instance, you could avoid making a detailed claim about the phonetic form of the protolanguage by arbitrarily reconstructing the correspondence $l : l : r : l$ as going back to $/*l/$, while suggesting $/*L/$ as the source for the correspondence $o : l : r : l$. By using the capital letter here, you are saying that this was probably some kind of liquid, but you are not sure exactly what it was. Another option in these kinds of situations is to use subscript or superscript numerals, as in $/*l_1/$ and $/*l_2/$.

Reading Guide Questions

1. What do we mean when we say that one form is a reflex of another form?
2. What are cognate forms?
3. What is the comparative method?
4. What is linguistic reconstruction?
5. What do we mean by "sound correspondences" when applying the comparative method?
6. What kinds of factors must we consider when reconstructing the phonemes of a protolanguage from the sound correspondences in the daughter languages?
7. How can we reconstruct a phoneme if a conditioned sound change has taken place?
8. In what situations is the comparative method unable to reconstruct a protolanguage correctly?

Exercises

1. Write formal rules expressing the changes that have taken place in Tongan, Samoan, Rarotongan, and Hawaiian, using the explanation in this chapter. Also state, for each of these changes, whether it is a conditioned or an unconditioned change, and say whether it is an example of phonemic loss, addition, shift, split, or merger.
2. Look at the Yimas and Karawari forms in Dataset 4. How do you think the original forms given on the left for the protolanguage were arrived at? Do you think they are reasonable reconstructions to make on the basis of the evidence that you have?
3. Look at the Suena and Zia forms in Dataset 6. Did the ancestral language have contrastive nasals? Why?
4. Look again at the Suena and Zia forms in Dataset 6. There are some correspondences between Suena /a/ and Zia /o/. Do these correspondences require us to reconstruct and additional vowel phoneme? Why?
5. Look at the information in Dataset 7 from the Korafe, Notu, and Binandere languages and reconstruct the original forms.
6. Examine the data from the northern and southern dialects of Paamese in Dataset 8 and reconstruct the original language. (It will help if you look at the rules described in sec. 9.3 in conjunction with this exercise.)
7. Examine the forms in Dataset 10 from the Sepa, Manam, Kairiru, and Sera languages. Take the language pair Sepa and Manam and say which sets of forms you think are cognate and which you think are not cognate. Now do the same for the pair Sepa and Kairiru.
8. Examine the following pairs of cognate forms in Abau and Idam, which are both spoken in the West Sepik Province of Papua New Guinea. Make an attempt to reconstruct the form in the protolanguage from which these forms are descended, and state what changes have taken place.

| Abau | Idam | |
|--------|--------|--------------|
| anan | anan | 'centipede' |
| am | am | 'place' |
| ak | ak | 'talk' |
| sak | sak | 'snake' |
| hauk | ɸauk | 'lake' |
| sauk | sauk | 'sago jelly' |
| kwal | kwal | 'bangle' |
| nanak | nanak | 'get' |
| naukan | naukan | 'branch' |
| hau | ɸau | 'taro' |
| auk | auk | 'string bag' |
| nausam | nausam | 'dry tree' |

9. Try to reconstruct the original forms from which the Ndao and the Sawu forms (from eastern Indonesia) are derived, and state what changes have taken place in both languages.

| Ndao | Sawu | |
|------|------|---------------------------|
| haha | wawa | 'pig' |
| silu | hilu | 'wear cloth around waist' |
| ceo | heo | 'nine' |
| əci | əhi | 'one' |
| heʔo | weʔo | 'tongue' |
| saʔu | haʔu | 'breast' |
| caʔe | haʔe | 'climb' |
| həru | wəru | 'moon' |
| dəsi | dəhi | 'sea' |
| hei | wei | 'give' |
| səmi | həmi | 'receive' |
| hela | wela | 'axe' |

10. Examine the list of cognate forms in Dataset 16 from the Aroma, Hula, and Sinaugoro languages of the Central Province of Papua New Guinea. Use the comparative method to reconstruct what you think to be the forms for all of these words in the protolanguage. Do not forget to look for complementary distribution among phonetically similar sets of correspondences, to avoid reconstructing too many protophonemes. (Note that the data have been slightly regularized to make the problem more workable.) Are there any words for which you are unable to reconstruct the original form? Why can you not do this?
11. Examine the following original forms in the Gamilaraay and Yuwaaliyaay languages of New South Wales (in Australia). Assume that the original language had /*ɹ/. Under what circumstances and in what ways did this change in Yuwaaliyaay?

| Gamilaraay | Yuwaaliyaay | |
|------------|-------------|------------|
| bɪɹ: | biju: | 'hole' |
| bwa | buja | 'bone' |
| ɟwa:j | ɟuja:j | 'flame' |
| gwa:r | guja:r | 'tall' |
| jiɹa | jija | 'tooth' |
| mwa:j | muja:j | 'cockatoo' |
| biɹi | bi: | 'chest' |
| maɹa | ma: | 'hand' |

| | | |
|----------|----------|-------------|
| ɟaɟaɟ | ja:j | 'sun' |
| ɟaɟaɟ | ga:j | 'language' |
| ŋuɟu | ŋu: | 'he/she' |
| juɟu | ju: | 'dust' |
| ɟɟigaɟa: | ɟɟigaɟa: | 'bird' |
| waɟaba | wajaba | 'turtle' |
| waɟaɟa:l | wajaga:l | 'left hand' |
| waɟawaɟa | wajawaja | 'crooked' |

12. Examine the following original forms in the Gamilaraay and Wiradjuri languages of New South Wales (in Australia). Reconstruct the original forms of these words and write rules that account for the changes:

| Wiradjuri | Gamilaraay | |
|-----------|------------|------------|
| ɟalaɟ | ɟalaɟ | 'tongue' |
| ɟuwaɟ | ɟuwaɟ | 'blood' |
| ɟulaɟ | ɟulaɟ | 'skin' |
| muɟaɟ | muɟaɟ | 'cockatoo' |
| ɟulip | ɟuli | 'goanna' |
| ɟip | ɟi: | 'meat' |
| wip | wi: | 'fire' |
| ɟip | ɟi: | 'heart' |
| ɟinaɟ | ɟina | 'foot' |
| ɟuja | ɟuja | 'fish' |
| ɟuraɟ | ɟura | 'bark' |
| ɟanaɟ | ɟana | 'liver' |
| ɟuwaɟ | ɟuwa | 'fog' |
| mipɟaɟ | mipa | 'what' |
| ŋamuɟ | ŋamu | 'breast' |
| ɟiliɟ | ɟili | 'lip' |
| ŋuruɟ | ŋuru | 'night' |
| ɟiaɟ | ɟia | 'tooth' |
| ɟaliɟ | ɟali | 'water' |

13. Examine the data for Nyulnyulan (Western Australia) in Dataset 15. Reconstruct the original forms and provide a list of the sound changes that you hypothesize, expressing them in rule form in as general a form as possible.

14. Modern French has the words *écoute* [ekut] 'listen to', *étranger* [etrãže] 'foreign', and *état* [eta] 'state', which were copied into English in the past as *scout* (i.e., one who listens), *strange* (i.e., something that is foreign), and *state*. At a later stage, English recopied the last two words as *estranged* (as in estranged wife) and *estate*. From the form of these lexical copies in English, what can you suggest about the history of the three French words given above?

Further Reading

- Ronald W. Langacker, *Language and Its Structure*, chapter 8, "Genetic Relationships," pp. 207–19.
- Robert J. Jeffers and Ilse Lehiste, *Principles and Methods for Historical Linguistics*, chapter 1, "Comparative Reconstruction," pp. 17–36.
- Theodora Bynon, *Historical Linguistics*, chapter 1.7, "Phonological Reconstruction (The Comparative Method)," pp. 45–57.
- Hans Henrich Hock, *Principles of Historical Linguistics*, chapter 18, "Comparative Method: Establishing Linguistic Relationship," pp. 556–80; chapter 19, "Comparative Reconstruction," pp. 581–627.
- Mark Hale, *Historical Linguistics*, part I, "Language" and "Language Change": Preliminaries, pp. 1–48.
- Lyle Campbell, "Beyond the Comparative Method."
- Joseph and Janda provide a good compilation in *The Handbook of Historical Linguistic* of many of the issues discussed briefly here. See especially Robert Rankin's chapter "The Comparative Method" (pp. 183–212).
- The papers in Durie and Ross, *The Comparative Method Reviewed*, are further advanced reading on some of the issues discussed here.

Determining Relatedness

Up until now, we have assumed that the languages we are discussing are related, but we have not talked about how to work out whether the languages are related in the first place. There are two main situations where we want to investigate the relatedness between languages. One is where we do not know what relatives a language has, and we want to work out which languages our study language is related to. The other is where we know something about the relatives of the language, but we want to find out which languages our study language is *more closely related to*. This is called SUBGROUPING.

Determining relatedness and determining subgrouping are not the same process, although similar types of evidence can be used in each case. Making a case that two languages are genetically related primarily involves showing that they share material which is extremely unlikely to have arisen by chance. In contrast, showing subgrouping requires showing that the languages have undergone the same changes. We talk more about these two techniques in this chapter.

6.1 FINDING FAMILIES

Most of the time, the initial finding of language families is a matter of being in the right place at the right time. A linguist might notice that some features of a language resemble something that they know from another language. There

might be some similar words or some similar affixes. In some cases, it is a single peculiarity which is so unlikely to be due to chance that it sends the linguist looking for other similarities.

It will be possible to find some similarities between any pair of languages, whether or not they are related. This is because there is a finite number of sounds that each language makes use of; moreover (as we saw in chapter 1), some types of words are more likely than others to have the same form across languages. It is not surprising that a very large number of languages have baby talk words for 'mother' that sound something like [məmə] or [mama], since labial nasals are some of the first sounds that babies produce. In other cases, words might have a similar form and similar meaning purely by chance. The Mbabaram word for 'dog' is *dək*, almost identical to the English word. This is not because it is a loan from English; it is a chance resemblance produced by sound change (the Mbabaram word goes back to something like *kutaka).

So, given that there may be similarities between languages due to chance or to universals, what similarities constitute good evidence of linguistic relatedness? The best similarities to use are the same types of evidence that we would use for reconstruction: that is, *systematic* meaningful correspondences in lexical items, morphology, and grammar. Specifically:

1. There should be regular correspondences in lexical items. These correspondences do not necessarily have to involve phonetically similar sounds. As we saw in sec. 2.9, over time cognate words can look rather different from one another. But the correspondences do need to be regular.
2. Correspondences should not be confined to a single area of the language (or to a single area of grammar). Similarities that are restricted to just one area of the languages are difficult to interpret. For example, it might be that two languages have similar pronouns, but there is little else they share. On the one hand, we could argue that there has been sufficient lexical change that other similarities have been eroded, leaving only pronouns as the identifiable cognate items. On the other hand, that explanation begs the question of why the pronouns alone should remain similar, apparently immune from the sound and other changes that affected all the other cognate material.
3. Shared suppletive forms are more indicative of a relationship than random shared items. This is because shared suppletive forms such as "good, better, best" tend not to be borrowed and are less likely to arise by chance.

These three pieces of evidence together constitute demonstration of genetic relatedness. We should also consider what similarities *do not* provide evidence for relatedness. Nonlinguistic features of speech communities, such as religion, race, genetics, or cultural practices, provide no evidence for language classification. Speech communities can and do shift languages, cultural practices, and religion. Another set of similarities which are not evidence

TABLE 6.1 Numbers in Assorted European languages

| English | Dutch | German | French | Italian | Russian |
|---------|-------|--------|--------|---------|----------|
| wan | en | ains | œ | uno | adʲin |
| tu: | twe: | tsvai | dø | due | dva |
| θri: | dri: | dʁai | tʁwa | tre | trʲi |
| fɔ: | fi:r | fi:ʁ | katʁ | kwatro | tʃitirʲi |
| farv | feif | fynf | søk | tʃiŋkwe | pʲatʲ |

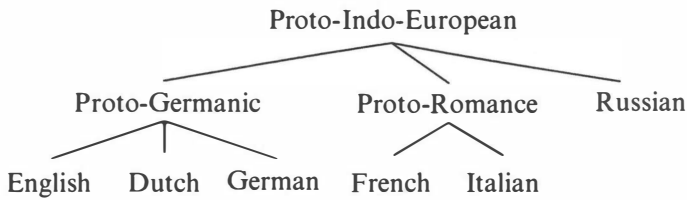
are typological features such as basic word order, the number of phonemes, the number of cases, or whether the language has ergative alignment. All of these features show considerable diversity within known families. They are therefore not stable enough over time to reveal evidence of deep genetic relatedness.

Finally, it should be pointed out that we can never prove that two languages are not related. We can show that there is no evidence of a convincing sort that any given pair of languages are related, but that doesn't mean that the languages are not related at some point in the past, only that we can't show it with our methods.

6.2 SUBGROUPING

In chapter 5, you learned how to reconstruct earlier stages of a family and to describe the sound changes that the languages had undergone. By using the comparative method, we can reconstruct a protolanguage and we can use the results that it provides to determine which languages are more closely related to other languages in a family. Compare the words from six Indo-European languages in table 6.1:¹ There are enough similarities even here, in the words for 'two' and 'three', for example, to suggest that we could justify putting these six languages into a single language family. However, some other similarities seem to suggest that English, Dutch, and German are closer to each other than they are to the other three languages. Similarly, French and Italian seem to be fairly closely related to each other, while being less closely related to the others. Finally, Russian seems to stand out on its own. What we can say here is that we have three SUBGROUPS of the one language family—one containing the first three languages, one containing the next two, and a final subgroup with only a single member.

We can represent subgrouping in a FAMILY TREE by a series of branches coming from a single point. The family tree for the six languages discussed above would look something like this:



This diagram can be interpreted as meaning that English, Dutch, and German are all derived from a common protolanguage (which we can call Proto-Germanic) that is itself descended from the protolanguage that is ancestral to all of the other languages (which we can call Proto-Indo-European). We can therefore offer a tentative definition of a subgroup by saying that it comprises a number of languages that are all descended from a common protolanguage that is intermediate between the ultimate (or highest-level) protolanguage and the modern language, and which, as a result, are more similar to each other than to other languages in the family.

In summary, here is a set of procedures for doing subgrouping:

1. Gather data from languages known to be related. (Subgrouping tells you *how* various languages are related, not whether or not they are related.)
2. Reconstruct the protolanguage using the comparative method.
3. Note the sound changes which have occurred in the history of each language.
4. Make careful note of the relative chronology inherent in your reconstructions.
5. Group together the languages which have undergone shared changes (a period of common development).
6. Remember that the best diagnostic evidence for subgrouping is unusual change.
7. Draw a family tree which reflects the subgrouping you have worked out.
8. Don't forget to check your rules.

6.3 SHARED INNOVATION AND SHARED RETENTION

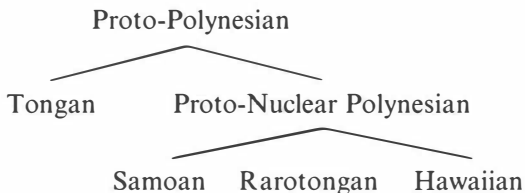
Clearly, languages that belong to the same subgroup must share some similarities that distinguish them from other languages in the family that do not belong to this subgroup. However, the simple fact that there are similarities does not necessarily mean that two languages belong in the same subgroup. If we say that two languages belong in the same subgroup, we imply that they have gone through a period of common descent and that they did not diverge until a later stage in their development.

Similarities between languages can be explained as being due to either shared retention from the protolanguage or shared innovations since the time

of the protolanguage. If two languages are similar because they both have some feature that has been retained from the protolanguage, you cannot use this similarity as evidence that they have gone through a period of common descent. The retention of a particular feature in this way is not significant because you should expect a large number of features to be retained anyway.

However, if two languages are similar because they have both undergone the same innovation or change, then you can say that this is evidence that they have had a period of common descent and that they therefore do belong to the same subgroup. You can say that a shared innovation in two languages is evidence that those two languages belong in the same subgroup because exactly the same change is unlikely to take place independently in two separate languages. By suggesting that the languages have undergone a period of common descent, you are saying that the particular change took place only once between the higher-level protolanguage and the intermediate protolanguage which is between this and the various modern languages that belong in the subgroup. Other changes then took place later in the individual languages to differentiate one language from another within the subgroup.

If you look back to the reconstructions that you made for Proto-Polynesian in chapter 5, you will see that Samoan, Rarotongan, and Hawaiian have all undergone unconditional loss of the original phonemes /*h/ and /*ʔ/. This suggests that Samoan, Rarotongan, and Hawaiian all belong together in a subgroup of Polynesian from which Tongan is excluded. Between Proto-Polynesian and the intermediate ancestor language from which these three languages are derived (but not Tongan), there was an intermediate protolanguage, which we can call Proto-Nuclear Polynesian:



While it is shared innovations that we use as evidence for establishing subgroups, certain kinds of innovations are likely to be stronger evidence for subgrouping than other kinds. As I have just said, subgrouping rests on the assumption that shared similarities are unlikely to be due to chance. However, some kinds of similarities between languages are in fact due to chance; that is, the same changes do sometimes take place quite independently in different languages. This kind of situation is often referred to as PARALLEL DEVELOPMENT OR DRIFT. One good example of drift is in the Oceanic subgroup of the Austronesian family of languages (which includes

all of the Polynesian languages, as well as Fijian, and the Austronesian languages of Fiji, Vanuatu, New Caledonia, Solomon Islands, and Papua New Guinea). In Proto-Oceanic, word-final consonants were apparently retained from Proto-Austronesian. However, many present-day Oceanic languages have since apparently lost word-final consonants by a general rule of the form

*C > ø / ___ #

The fact that many Oceanic languages share this innovation is not sufficient evidence to establish subgroups. Loss of final consonants is a very common sort of sound change that could easily be shared by to chance, and the same sound change occurs in Oceanic, as well as in some languages that we would not otherwise want to call Oceanic languages. In the Enggano language, spoken on an island off the coast of southern Sumatra, final consonants were also lost, but we would not necessarily want to say that this language belongs in the Oceanic subgroup as this language shares no other features of Oceanic languages.

In classifying languages into subgroups, you therefore need to avoid the possibility that innovations in two languages might be due to drift or parallel development. You can do this by looking for the following in linguistic changes:

1. Changes that are particularly unusual.
2. Sets of several phonological changes, especially unusual changes that would not ordinarily be expected to have taken place together.
3. Phonological changes that correspond to unconnected grammatical or semantic changes.

For example, if Samoan, Rarotongan, and Hawaiian only shared the single change whereby /*h/ was lost, it might be possible to argue that this is purely coincidental, especially as the loss of /h/ is a fairly common sort of change anyway. However, as these three languages also share the change, we can argue that coincidence is less likely to be the explanation and that these three languages are indeed members of a single subgroup.

If two languages share a common sporadic or irregular phonological change, this provides even better evidence for subgrouping those two languages together, as the same irregular change is unlikely to take place twice independently. One piece of evidence that can be quoted for the grouping of Oceanic languages into a single subgroup of Austronesian is the irregular loss of /*r/ that has taken place in the Proto-Austronesian word /*mari/ 'come'. On the basis of evidence from the present-day Oceanic languages, we can reconstruct the form /*mai/ 'come' in Proto-Oceanic. On the basis of the reconstructed Proto-Austronesian form /*mari/, however, we would have expected the Proto-Oceanic form to be /*mari/ instead of

*/*mai/*. Proto-Oceanic appears to have lost this sound in just this single word to produce an irregular reflex of */*mari/*. It is highly unlikely that every single Oceanic language would have independently shifted */*mari/* to */*mai/*, so we conclude instead that this irregular change happened just once, between Proto-Austronesian and Proto-Oceanic, and that the modern Oceanic languages reflect this irregularity as a retention from Proto-Oceanic.

The Oceanic subgroup of the Austronesian family has not been established on the basis of just this single innovation, even though it is an irregular one. Several other regular phonological changes have also taken place at the same time. These include the following:

*ə > o

*b > p

*g > k

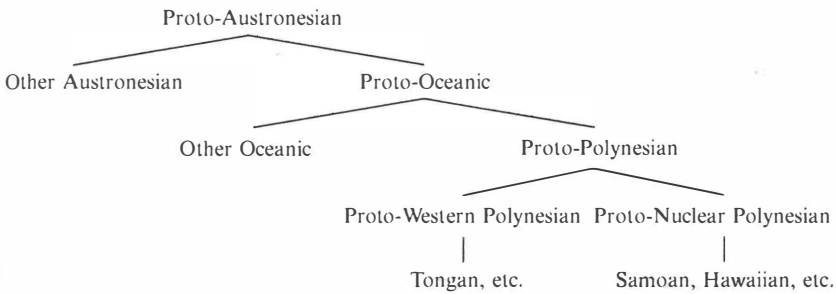
These involve a change of schwa to /o/, as well as the devoicing of stops, so parallel development is unlikely to be the explanation. We can therefore conclude that any Austronesian language that shares all of these innovations is a member of the Oceanic subgroup.

The pair of shared innovations that I gave earlier in Samoan, Rarotongan, and Hawaiian are also better evidence for subgrouping than just a single change. For instance, both Tongan and Hawaiian have undergone a shift of */*s/* to */h/*. It would contradict the conclusion that I just reached to say that Tongan and Hawaiian belong to a single subgroup on the basis of this shared innovation. Where there is information that is consistent with competing subgrouping interpretations, we should evaluate this and see which solution is the most reasonable one. The fact that the first conclusion was reached on the basis of a pair of shared innovations whereas the second conclusion would have to be based on just a single innovation makes the first conclusion a more reliable one. We must simply conclude that both Tongan and Hawaiian independently changed */*s/* to */h/* at separate times in history after the two had diverged.

Finally, if we can match phonological innovations with shared grammatical or semantic innovations, then we can argue that we have good evidence for putting the languages that share these features into the same subgroup. Although the grammatical reconstruction of Proto-Austronesian is much less well developed than its phonological reconstruction, some linguists argue that many aspects of the basic clause structure of Oceanic languages are different from the structure of Proto-Austronesian. If this turns out to be confirmed, then this would be further evidence for the existence of an Oceanic subgroup.

When we speak of subgroups of languages, it is possible to speak of higher-level subgroups and lower-level subgroups. As you have seen,

languages that belong to a subgroup within a single language family have experienced a period of common descent. However, it is possible for languages within a single subgroup of a larger language family also to be subgrouped together on the basis of shared innovations. This means that we can speak of subgroups within subgroups. For instance, there are strong arguments for saying that the Polynesian languages represent a separate subgroup within the Oceanic subgroup, on the basis of their shared phonological, lexical, and grammatical innovations. In this kind of situation, we can speak of Oceanic being a higher-level subgroup, while the Polynesian languages constitute a lower-level subgroup. Languages that belong together in higher-level subgroups therefore diverged relatively early, while lower-level subgroups involve later developments. Of course, the Polynesian languages can be further subgrouped into even lower level subgroups again, and I have already indicated that we can justify a subgroup consisting of Samoan, Rarotongan, and Hawaiian, as well as a Western Polynesian subgroup, of which Tongan is a member. We could represent the different levels of subgrouping as follows:



6.4 LONG-DISTANCE RELATIONSHIPS

One area of historical linguistics which makes the news from time to time are the long-distance proposals for very archaic relationships between families. Work by Greenberg, Ruhlen, and others is often mentioned here, but there is quite a lot of work in this vein.

It is intuitively very appealing to wish to trace the linguistic ancestry of as many groups as possible. After all, wouldn't it be great if we could not only show that we once all spoke the same language but also reconstruct aspects of it? And if we could work out what the major splits and intermediate protolanguages were, just as we can for more recent families like Indo-European, Austronesian, or Algonquian? Such information would be really useful for

prehistory. It would also be great to have language family data going back tens of thousands of years, because then we could correlate the linguistic results with genetic data.

Several methods are commonly used for long-range comparisons. One is to compare protolanguages. That is, if we want to find out the properties of a putative ancestor language of Indo-European and Finno-Ugric languages, wouldn't we be saving time if we just compare our reconstructions for Proto-Indo-European and Proto-Finno-Ugric? However, remember that protolanguages are hypotheses for prior forms; they aren't "real" languages: they are fragments of hypothesized languages: Therefore, comparing them as though they were real languages greatly increases the chances of flawed comparisons.

Another method for finding long-distance relationships is called "mass comparison." It was developed by Joseph Greenberg (see, for example, Greenberg [1963, 1987]). It relies on finding similarities between languages under a much less strict basis for comparison than the comparative method. The same problems that we identified in comparing protolanguages are the basis for mass comparison. If the signal is very weak at great time depth, the idea is that we should relax semantic or phonological identity constraints in order to get more data and to "boost" the signal.

There have been many criticisms of mass comparison and related methods. A fair number of the criticisms reduce to data problems. Many etymologies for long-range relationships are CVC or CV syllables. Combined with the relaxation in semantic identity, this allows for lots of potential "cognates" but no way for identifying better potential cognates than others. Relaxing strictness for the initial comparison makes it difficult to tell what are real cognates and what are chance resemblances. A good off-the-cuff test for a long-range proposal is to see if you can add "cognates" from your favorite language which is not proposed to be part of the phylum. For example, there are about as many Amerind "cognates" in the Australian language Bardi as there are for the average Amerind language. One of the advantages of the comparative method is that it can rule in and rule out languages from a particular hypothesis of relatedness. If a long-range proposal is sufficiently permissive that any language can be potentially part of the family, there is no evidence for the proposed family itself.

Tempting as it might be to look tens of thousands of years back into the past, we can't do that with linguistic data at this stage. There are two reasons for this. First, reliable methods where we can build a good case for relationship require a certain strength of signal in the data. We need regular correspondences to make the case (as we saw in sec. 6.1), and after a certain period of time, and in certain language-contact conditions, the signal decays very quickly. That is, after a certain amount of time, enough changes build up that there are too few recurrent correspondences. Reliable methods cannot reach far back enough in time to let us reconstruct proto-world.

The second reason is that none of the methods advanced as alternatives to the comparative method can distinguish between random fluctuations and chance resemblances on the one hand and genuine remote cognates on the other.²

Reading Guide Questions

1. Why can't we prove conclusively that two languages aren't related?
2. What is a subgroup?
3. What is the difference between a shared retention and a shared innovation?
4. Why can similarities between languages that are due to shared retentions not be used as evidence for subgrouping?
5. What is drift or parallel development? How does this affect the way we go about deciding on subgroups?
6. What sorts of innovations are the best kind of evidence for subgrouping?

Exercises

1. Pick two dictionaries of other languages at random from your university library. Look up 30 words of basic vocabulary in each and compare them. Are the languages likely to be related? Why or why not? What are the problems in using a method like this?
2. Look at the Korafe, Notu, and Binandere forms in Dataset 7. On the basis of the reconstruction of the changes from the protolanguage that you worked out in the exercises at the end of chapter 5, would you say that Notu belongs to the same subgroup as Korafe or Binandere? Why?
3. Look back at the reconstruction of the protolanguage for Aroma, Hula, and Sinaugoro that you did in the exercises for chapter 5 (Dataset 16). What subgrouping hypothesis can you make for these three languages on the basis of shared innovations?
4. Look at the Nyulnyulan data in Dataset 15. What subgrouping is suggested from the data? What are your reasons for your hypothesis?
5. Look at the forms in Proto-Gazelle Peninsula (New Britain, Papua New Guinea) in Dataset 17. What is the subgrouping of the four speech communities that are represented? Give the justification for the answer that you propose. (Note that the superscript vowels represent phonetically reduced sounds that are nearly voiceless, and not stressable.)
6. Look at the following data from six different languages and answer the questions:
 - a. How many language families are represented in these data?
 - b. What are your reasons for saying this?

- c. What factors can you suggest to account for the similarities between languages that you say do not belong to a single family?

| A | B | C | D | E | F | |
|----------|----------|--------------|----------|---------|----------|-----------|
| mwana | mwana | umwana | baceh | anak | bata | 'child' |
| lia | dila | lila | girjeh | triak | ijak | 'cry' |
| pwa | nua | nwa | nufidan | minum | inum | 'drink' |
| moto | tija | umulio | atef | api | apoj | 'fire' |
| nne | ia | ne | cæhær | æmpat | ampat | 'four' |
| kilima | mongo | ulupili | tel | bukit | bukid | 'hill' |
| ceka | seva | seka | xændidan | tærtawa | tawa | 'laugh' |
| mguu | kulu | ukuulu | saq | kaki | pa | 'leg' |
| mdo | mokoba | umulomo | læb | bibir | bibig | 'lip' |
| mtu | muntu | umuntu | mærd | oraj | tau | 'man' |
| habari | nsangu | iceevo | xæbær | kabar | balita | 'news' |
| moja | mosi | mo | jek | satu | isa | 'one' |
| nabii | mbikudi | umusimicisi | næbij | nabi | propetas | 'prophet' |
| mvua | mvula | imfula | baran | hujan | ulan | 'rain' |
| merikebu | maswa | ubwato | mærkæb | kapal | bapor | 'ship' |
| dhambi | masumu | icakuvifja | zamb | dosa | kasilan | 'sin' |
| askari | kinwani | icita | æskær | askar | suldado | 'soldier' |
| kidonda | mputa | icilonda | zæxm | sakit | sakit | 'sick' |
| lutoba | maloji | isiwi | xutbæh | xutbah | salita | 'speech' |
| hadhithi | jana | icisimicisjo | hædis | cerita | istoria | 'story' |
| hekalu | kinlongo | itempuli | hæjkil | rumah | templo | 'temple' |
| tatu | tatu | tatu | seh | tiga | tatlo | 'three' |
| mti | nti | umuti | dæræxt | pohon | puno | 'tree' |
| bili | zole | vili | do | dua | dalawa | 'two' |

7. The following data come from four languages spoken in the area of Cape York in northern Queensland in Australia. Examine the reconstructed protolanguage and the descendant forms, and suggest a subgrouping hypothesis on the basis of the shared innovations. There is one set of changes which is problematic for an otherwise strong subgrouping hypothesis. What original sound is involved? Can you suggest one or more solutions to this problem (in the abstract)?

| Proto-Cape York | Atampaya | Angkamuthi | Yadhaykenu | Wudhadhi | |
|-----------------|---------------------|---------------------|--------------------|----------|------------|
| *kaca | ya _ḥ ḥa | a _ḥ ḥa | a _ḥ ḥa | — | 'rotten' |
| *kantu | ya _ḥ ntu | antu | antu | antu | 'canoe' |
| *puṅku | wuṅku | wuṅku | wuṅku | — | 'knee' |
| *paṅka | ṅaṅka | aṅka | aṅka | aṅka | 'mouth' |
| *juku | juku | juku | juku | — | 'tree' |
| *pinta | winta | winta | winta | inta | 'arm' |
| *puṅa | wuṅa | wuṅa | wuṅa | uṅa | 'sun' |
| *cipa | lipa | jipa | jipa | — | 'liver' |
| *wapun | wapun | apun | apu | apun | 'head' |
| *wujpu | wujpu | ujpu | ujpu | ujpu | 'bad' |
| *ujpuṅ | ujpuṅ | ujpuṅ | ujpuṅ | ujpuṅ | 'fly' |
| *ajpaṅ | ajpaṅ | ajpaṅ | ajpaṅ | ajpaṅ | 'stone' |
| *calan | lalan | jalan | jala | alan | 'tongue' |
| *paṅṭal | waṅṭaw | waṅṭa: | waṅṭa: | — | 'yam' |
| *jaṅṭal | jaṅṭaw | jaṅṭa: | jaṅṭa: | — | 'road' |
| *pili | wili | wili | wili | — | 'run' |
| *ṅuṅka | ṅuṅka | juṅka | juṅka | uṅka | 'cry' |
| *ja | ja | ja | ja | — | 'throw' |
| *rupal | rupaw | jupa: | jupa: | — | 'white' |
| *rucu | ruṭu | juṭu | juṭu | uṭu | 'dead' |
| *pilu | wilu | wilu | wilu | ilu | 'hip' |
| *pupu | wupu | wupu | wupu | upu | 'buttocks' |
| *ṅampu | ṅampu | ampu | ampu | ampu | 'tooth' |
| *maji | maji | aji | aji | aji | 'food' |
| *ṅukal | ṅukaw | uka: | uka: | ukal | 'foot' |
| *miṅa | miṅa | iṅa | iṅa | iṅa | 'meat' |
| *iwuṅ | — | — | iwuṅ | iwuṅ | 'ear' |
| *japan | ja _ḥ pan | ja _ḥ pan | ja _ḥ pa | — | 'strong' |

8. You have seen that subgrouping depends on being able to distinguish shared innovations from shared retentions in the protolanguage. Features are reconstructed in the protolanguage partly on the basis of the extent of their distribution in the daughter languages, as you learned in chapter 5. What methodological problem do we face here?

9. Consider the following pieces of evidence for a putative language relationship. Is it convincing? Why or why not? Does all the evidence point in the same direction?
- There are 10 languages in the putative family. Seven are spoken in the same valley, while the other three are spoken several hundred miles away, to the north and east.
 - The proposal for relatedness among these languages was initially made by a famous full professor.
 - All the languages have a plural marker, *-ap*.
 - Eight of the ten languages have verb serialization.
 - All the languages spoken in the valley have the word [k^wæt'ʒəŋ] for 'saltbush' and [sɪpk^wəʒət] for 'peat'.
 - Half the languages show a peculiarity in morphology where the order of subject and object agreement markers is reversed in the past causative. It is usually subject-tense-mood-root-aspect-object, but in the causative the order is object-tense-mood-causative-root-aspect-subject.
 - Two widely separated languages show resemblances in approximately 35 percent of their basic vocabulary.
10. The following families are some recent proposals for long-distance relationships. Pick one of these hypotheses and find out something about it. Who originally proposed it? What was the basis for the proposal? Has there been any debate about it? Has the proposal been stable, or have different language families been included at various times?

| | |
|--|-----------------------|
| Altaic | Japanese-Austronesian |
| Australian | Macro-Jê |
| Austic (Austro-Asiatic and Austronesian) | Na-Dené |
| Basque-Caucasian | Nostratic |
| Eskimo-Aleut-Austronesian | Penutian |
| Hokan | Trans-New Guinea |
| Indo-Uralic | |

Further Reading

- Lyle Campbell, *American Indian Languages*, chapter 2, "The History of American Indian (Historical) Linguistics," pp. 26–89.
- Stefan Georg and Alexandr Vovin, "From Mass Comparison to Mess Comparison." James A. Matisoff, "On Megalocomparison."
- Claire Bowern and Harold Koch, "Introduction" to *Australian Languages*, pp. 1–16.
- Lyle Campbell and William Poser, *Language Classification*, chapters 9 ("Assessment of Proposed Distant Genetic Relationships") and 10 ("Beyond the Comparative Method?"), pp. 234–329.



Internal Reconstruction

In chapter 5, you learned how to apply the comparative method to reconstruct an earlier form of an unrecorded language by comparing the forms in the various daughter languages that are descended from it. However, the comparative method is not the only method that you can use to reconstruct linguistic history. A second method of reconstruction known as INTERNAL RECONSTRUCTION allows you to make guesses about the history of a language as well. The basic difference between the two methods is that in the case of internal reconstruction, you reconstruct only on the basis of evidence from within a single language, whereas in the comparative method, you reconstruct on the basis of evidence from several different languages (or dialects). With the comparative method you arrive at a protolanguage from which two or more languages (or dialects) are derived, while with the internal method of reconstruction, you simply end up with an earlier stage of a language. We can call this stage of a language that you have reached by internal reconstruction a PRELANGUAGE.

Internal reconstruction is often used in morphological reconstruction for making inferences about prior morphological stages. However, it is also used in reconstruction in syntax, and the results from grammaticalization theory are often used in conjunction with arguments from internal reconstruction in syntax. In this chapter, though, we talk only about internal reconstruction in morphology. I cover internal reconstruction in syntax in sec. 12.3.

7.1 USING SYNCHRONIC ALTERNATIONS

The Dutch linguist van der Tuuk (1971:xlīiii) once said: "Every language is more or less a ruin." What he meant was that as a result of changes having taken place, some "residual" forms are often left to suggest what the original state of affairs might have been. Applying the method of internal reconstruction is in some sense similar to the science of archeology. In archeology we use the evidence of the present (i.e., the covered remains of earlier times) to reconstruct something of the past. Archeology does not enable us to reconstruct everything about the past—only those facts that are suggested by the present-day "ruins" from the past.

Let us now look at an example of a linguistic change that has taken place in a language and see what sorts of ruins it leaves in the modern language. The language that we consider here is Samoan. This language has verbs that appear in both intransitive and transitive forms. The intransitive form is used when there is no following object noun phrase, and verbs in this construction involve the bare root with no suffixes of any kind. In the case of transitive verbs (which are used when there is a following object noun phrase), there is a special suffix that is added to the verb. In Samoan, different transitive verbs take different suffixes, as shown by the following examples:

| Intransitive | | Transitive | |
|--------------|----------|------------|---------------------|
| inu | 'drink' | inu-mia | 'drink (something)' |
| ŋau | 'break' | ŋau-sia | 'break (something)' |
| mataʔu | 'afraid' | mataʔu-tia | 'fear (something)' |
| taji | 'weep' | taji-sia | 'weep for' |
| alofa | 'love' | alofa-ŋia | 'love (somebody)' |
| fua | 'weigh' | fua-tia | 'weigh (something)' |
| ole | 'cheat' | ole-ŋia | 'cheat at' |
| silā | 'look' | silā-fia | 'see' |

Samoan has a variety of suffixes to mark exactly the same function, including the following: */-mia/*, */-sia/*, */-tia/*, */-ŋia/*, and */-fia/*. This variety in the transitive suffixes is the result of a sound change that took place at some time before the emergence of modern Samoan.

From comparative evidence, we know that the verb roots of the language that Samoan is descended from originally ended in both vowels and consonants. For instance, compare the following forms in Samoan and the distantly related language Bahasa Indonesia:

Bahasa Indonesia Samoan

| | | |
|-------|--------|----------|
| minum | inu | 'drink' |
| takut | mataʔu | 'afraid' |
| tajis | taji | 'weep' |

There is also comparative evidence to suggest that transitive verbs were once marked by adding the special suffix /-ia/ to the verb. Then there was a general change in the history of Samoan by which final consonants were lost. When the final consonants were lost, they disappeared in the intransitive forms of the verb but were retained in the transitive forms because when the suffix /-ia/ was added, the consonants were no longer at the end of the word but in the middle.

Now that, in Samoan, there were no longer any consonants at the ends of words, the consonants that were retained in the transitive forms of the verb came to be reanalyzed as part of the following suffix instead of being part of the root. So, what was originally a suffix with a single form has now developed a wide range of different forms, or ALLOMORPHS, as a result of a single sound change having taken place. These allomorphs are morphologically conditioned, which means that each verb must be learned with its particular transitive suffix, and there is nothing in the phonological shape of the verb that gives any clue as to which form of the suffix the verb will take. These changes are set out as follows:

| Pre-Samoan | | Samoan | | |
|--------------|-------------|--------------|------------|---------|
| Intransitive | Transitive | Intransitive | Transitive | |
| *inum | *inum-ia | inu | inu-mia | 'drink' |
| *ŋaus | *ŋaus-ia | ŋau | ŋau-sia | 'break' |
| *mataʔut | *mataʔut-ia | mataʔu | mataʔu-tia | 'fear' |
| *tajis | *tajis-ia | taji | taji-sia | 'weep' |
| *alofaŋ | *alofaŋ-ia | alofa | alofa-ŋia | 'love' |
| *fuat | *fuat-ia | fua | fua-tia | 'weigh' |
| *oleŋ | *oleŋ-ia | ole | ole-ŋia | 'cheat' |
| *silaf | *silaf-ia | sila | sila-fia | 'see' |

In talking about this problem, I have used the knowledge that I already have about the history of Samoan from comparative evidence to help you understand what has happened in the development of the transitive suffixes in the language. However, it would have been possible to make the same reconstruction on purely internal evidence. What you do when you apply

the internal method of reconstruction is to look at cases of morphological alternation (or allomorphs of morphemes), and you work on the assumption that unusual or complex distributions of allomorphs may well go back to a simpler state of affairs than you find in the modern language.

The distribution of the different forms of the transitive suffix is complex, in that each verb has to be learned along with its transitive counterpart, and there are no general rules that can be learned to help a speaker of the language. It is relatively unusual for languages to leave so much for the learner to have to remember, so you could assume that in Pre-Samoan the language was somehow more “learnable” and that this earlier, simpler system has broken down because of some sound change having taken place. The unpredictability in the Samoan data does not lie in the vowels as these are consistently *-ia*. What needs explanation is the existence of the preceding consonants. If you assume that the consonants were originally part of the root and there was a later loss of word-final consonants, then this gives a very simple picture of Pre-Samoan morphology, and it involves a natural sound change (i.e., the loss of final consonants).

Let us look at some data from a different language—German. The change that we will be dealing with is the devoicing of stops word finally that we looked at in chapter 2. In modern German, the plural of certain nouns is formed by adding the plural suffix */-ə/*, while in other nouns, the plural is formed by adding the suffix */-ə/* and at the same time changing the final voiceless consonant to the corresponding voiced consonant. So, compare the following singular and plural nouns in German:

| Singular | Plural | |
|----------|--------|---------|
| laut | lautə | ‘sound’ |
| bo:t | bo:tə | ‘boat’ |
| ta:k | ta:gə | ‘day’ |
| hunt | hundə | ‘dog’ |

Here again, you can see that there is complexity in the morphological alternations of the language, and you should ask yourself if this complexity could reasonably be derived from an earlier, more simple way of forming the plural. The suffix */-ə/* is common to all forms, so you can assume this to be original. You should note, however, that some plurals have preceding voiced consonants and some have preceding voiceless consonants, whereas the singular forms all have final voiceless consonants. By assuming that the plural roots represent the original forms of the roots, you can say that the singular forms have undergone a change of final devoicing according to the following rule:

$$*C_{[voiced]} > C_{[voiceless]} / _ \#$$

Clearly, the consonants in the plural would have been “protected” from this rule by the presence of the following plural suffix, and this is why they did not undergo devoicing.

To be sure, not all cases of morphological alternation can be reconstructed as going back to a single original form that “split” as a result of a sound change taking place. The important point to keep in mind is that the modern alternations must be derivable from an original form by means of reasonable kinds of sound changes. So, while you might want to reconstruct the /-s/, /-z/, and /-əz/ markers of the plural of English nouns as going back to something simpler in the past because of their phonetic similarity, you would be unlikely to reconstruct irregular plurals such as the following as being derived from the same source (however we might want to reconstruct it):

| Singular | Plural |
|----------|----------|
| foot | feet |
| goose | geese |
| man | men |
| woman | women |
| child | children |
| louse | lice |

Forms that are as divergent as this must clearly go back to irregular forms even in Pre-English.

7.2 INTERNAL RECONSTRUCTION AND INDO-EUROPEAN LARYNGEALS

One famous example of internal reconstruction involves reconstructed consonants in Indo-European known as LARYNGEALS. In this case, these consonants were reconstructed on the basis of internal patterns in ancient Greek, as well as comparative evidence with other Indo-European languages. Here we consider only the internal Greek evidence.

Many Ancient Greek words show alternations in their roots, as well as inflection for prefixes and suffixes. Consider the following words.¹ I have put the alternating part of the word in bold face.

| | | | |
|-----------|-----------|---------|------------|
| leipō | leloipa | elipon | ‘leave’ |
| eidō | oīda | īdmen | ‘know/see’ |
| eleusomai | elēloutha | ēluthon | ‘come’ |

| | | | |
|----------------|-------------------|--------------------|----------|
| rēūma | (ro-os) (< rouos) | perirrutos | 'stream' |
| menos | memona | memasan (< memṣan) | 'think' |
| patera | eupatora | patrasi | 'father' |
| petomai | potānos (winged) | eptomēn | 'fly' |

In these words you can see that there is a regular pattern. In the first column, the words have an [e] vowel (e.g., *leip*). In the second column, they all have an [o] vowel (e.g., *loip*). In the third column, however, there is nothing. You can see this particularly in the last word, by contrasting *petomai* 'I fly' and *potānos* 'winged' with *eptomēn* 'I flew'. This pattern can be summarized as follows:

e : o : ∅

Now, consider the alternations in these words:

| | | | |
|---------|--------|--------|---------|
| tithēmi | thōmos | thetos | 'put' |
| histāmi | — | statos | 'stand' |
| phāmi | phōna | phatos | 'say' |
| didōmi | dōron | dotos | 'give' |
| pōma | — | potos | 'drink' |

In these data, there is a similar but not identical pattern. Here, instead of the e : o : ∅ vowel alternation patterns, we have the following:

ē : ō : e

ā : ā : a

ō : ō : o

Now, is it possible to reconcile these two patterns? Yes. There is one other difference to note about the e : o : ∅ pattern versus the second pattern with long vowels. In the first case, the stems that alternate are basically all CVC (*leip* : *loip* : *lip*; *men* : *mon* : *mṣ* [with subsequent *n > a]; *pet* : *pot* : *pt*; and so on). But in the second set of data, the stems look like they have the form CV: (*thē* : *thō* : *the*, etc). Therefore, we could propose that the words in the second set used to also have the structure CVC and exhibit the e : o : ∅ pattern, but they also contained another segment which has subsequently been lost. We might want to represent this as CeX : CoX : CX. We would also have to assume that the segments in question caused the vowels to undergo compensatory lengthening and that the realization of the vowels was affected by the following segment. One type seems only to have lengthened the preceding vowel, but the second type seems to turn an **e into an *a, and the third to turn an **o into an *o. (The double asterisks here mean that the reconstruction is an internal reconstruction within the protolanguage.)

The second set of data leads us to reconstruct three subcases of the alternation, as follows:

eE : oE : E
 eA : oA : A
 eO : oO : O

That is, in words like the root for 'put', we have something that gives us an [e] vowel when it appeared alone. In words like the root 'stand', we have something that gives an [a] vowel, and in the case of the 'give' root, something that gives [o]. The "something" has come to be known as a laryngeal because the missing segments were most likely glottal, laryngeal, or pharyngeal fricatives. These days, the laryngeals are most commonly written as h_1 for the *E*-producing laryngeal, h_2 for the *A*-producing laryngeal, and h_3 for the *O*-producing laryngeal.

This reconstruction is originally due to Ferdinand de Saussure, and it was very controversial at the time. However, the decipherment of Hittite texts provided external support for the reconstruction. Anatolian languages (the subgroup of Indo-European to which Hittite belongs) have a consonant written as *h* in many of the places where Saussure predicted a laryngeal. Two examples are given here:

| Greek | Latin | Hittite | | English | Proto-Indo-European |
|-------|-------|---------|------------|---------|---------------------|
| anti | ante | ḫant- | (forehead) | before | * h_2 enti |
| | ovis | ḫawi | (Luwian) | sheep | * h_3 ewi- |

7.3 LIMITATIONS OF INTERNAL RECONSTRUCTION

The internal method of reconstruction has a number of inherent limitations, and it is for this reason that it is not used nearly as much as the comparative method in reconstructing the history of languages. For one thing, it clearly does not take us back as far in time as does the comparative method. It's also much less reliable, as we'll see below. For these reasons, you would normally consider using the internal method only in the following circumstances:

1. Sometimes, the language that you are investigating might be a linguistic isolate; that is, it may not be related to any other language (and is therefore in a family of its own). In such a case, there is no possibility of applying the comparative method as there is nothing to compare this language with. Internal reconstruction is therefore the only possibility that is available.

2. Sometimes, similarly, the language you are studying is so distantly related to its sister languages that the comparative method is unable to reveal very much about its history. This would be because there are so few cognate words between the language that you are working on and its sister languages that it would be difficult to set out the systematic sound correspondences.
3. You may want to know something about changes that have taken place between a reconstructed protolanguage and its descendant languages.
4. Finally, you may want to try to reconstruct further back still from a protolanguage that you have arrived at by means of the comparative method. The earliest language from which a number of languages is derived is, of course, itself a linguistic isolate in the sense that we are unable to show that any other languages are descended from it. There is no reason why you cannot apply the internal method of reconstruction to a protolanguage, just as you could with any other linguistic isolate, if you wanted to go back still further in time.

Apart from the fact that the internal method is restricted in how far back in time it can take us, there are some other limitations that are inherent to the method. As I showed you in the previous section, this method can usually only be used when a sound change has resulted in some kind of morphological alternation in a language. Morphological alternations that arise as a result of sound change always involve conditioned sound changes. If an unconditioned sound change has taken place in a language, there will be no synchronic residue of the original situation in the form of morphological alternations, so the internal method will be unable to produce results in these kinds of situations. (We might be able to infer, from gaps in the phoneme system of the language or from frequency distributions of phonemes, that a change had taken place, but we would not necessarily be able to work out which sounds it affected. For example, if a language has no /b/ phoneme, and there are twice as many instances of the phoneme /w/ as there are of /p/, we'd be justified in assuming a sound change of /*b/ > w in the history of the language. But we wouldn't be able to recover which words had /*b/ and which had /*w/.

Another kind of situation in which the internal method may be inapplicable—or worse, where it may even lead to false reconstructions—is when intermediate changes are affected by other later changes, with the first changes leaving no traces in the modern language. For example, in modern French, there are morphological alternations of the following kinds:

| Noun | | Verb | |
|------|------------------------|-------|-----------------------|
| nō | 'name' | nōme | 'to name' |
| fē | 'end' | finiŕ | 'to finish' |
| œ | 'one (<i>masc.</i>)' | yn | 'one (<i>fem.</i>)' |

On the basis of these alternations, you would be justified in reconstructing the Pre-French of the forms of the words in the left-hand column as having had the following original shapes:

- *nɔm 'name'
- *fin 'end'
- *yn 'one'

In order to account for the forms of the modern nouns in French, you would need to reconstruct a number of sound changes. (Although I have not given a large number of examples here, you can assume that these changes will account for a large number of other forms in the language which undergo the same kinds of alternations.) The first change would be that vowels preceding word-final nasal consonants underwent assimilatory nasalization. Following this would be a change whereby word-final nasal consonants were lost. Finally, you would need to reconstruct a rule that lowered nasalized vowels from high to mid. Thus, we could reconstruct the following sequence of events in the history of French:

| Pre-French | *nɔm | *fin | *yn |
|---------------------|------|------|-----|
| Vowel nasalization | nɔ̃m | fɪ̃n | ɹ̃n |
| Nasal deletion | nɔ̃ | fɪ̃ | ɹ̃ |
| High-vowel lowering | — | fɛ̃ | œ̃ |
| Modern French | nɔ̃ | fɛ̃ | œ̃ |

While these changes all seem to be perfectly plausible, they are in fact not supported by the written evidence that we have for the development of the French language. Written evidence indicates that the changes that actually took place were somewhat more complicated than this. First, the vowel nasalization rule did not apply as I have just suggested. What actually happened was that word-final [m] shifted to [n]:

*m > n / __ #

It was only then that the vowel nasalization change took place. However, this did not apply before [n] in just word-final position, as vowels before nonfinal [n] also nasalized. By a yet later series of changes, [n] in coda positions was deleted, while vowels before noncoda nasals lost their nasalization, resulting in the present forms. You can see that there is considerable detail on which the internal method of reconstruction has proved to be inaccurate in this case. It failed to reconstruct the change of final [m] to [n], and it also got the details wrong as to how the vowels came to be nasalized.

Other problems are involved in interpreting the results of internal reconstruction. By this method, for example, we may be tempted to reconstruct

an earlier stage of Samoan (which I referred to above as Pre-Samoan) in which there are final consonants on verbs. However, the method does not give any indication of how much earlier than modern Samoan it was that verbs actually had these final consonants. It is often assumed that a reconstructed prelanguage arrived at in this way represents a form of the language that was spoken somewhere between the present and the time that it split off from its nearest ancestor. But it would be quite incorrect to equate Pre-Samoan with a stage of the language somewhere between modern Samoan and the time that this split off from its closest Polynesian relatives, as these languages also exhibit similar kinds of variations. What we came up with in the preceding exercise involved a mixture of root-final consonants belonging to a protolanguage that goes back considerably earlier than Proto-Polynesian and the shapes of the remainder of the words belonging to modern Samoan. Although we do not have written evidence in this case to show that this reconstruction is in error, we are fortunate in having comparative data on related languages. In the case of a genuine language isolate, we would not be so lucky, and our reconstruction would therefore be that much less reliable.

Finally, internal reconstruction essentially projects synchronic alternations into the past. But such alternations do not always have the origin we would be tempted to reconstruct. For example, in some Australian languages (such as Bardi), clusters are simplified across morphological boundaries. It would be tempting to reconstruct this as a sound change; the evidence from internal reconstruction points in that direction. However, what actually seems to have happened is that intervocalic stops were deleted, and later on other vowels dropped out in some environments. Comparative evidence shows us that the consonant clusters that we'd want to reconstruct from internal evidence probably never existed.

Not all current straightforward morphological alternations have a straightforward history. Elsewhere in this book, we talk about the alternation in the English prefixes /un-/ and /in-/. If we wished to perform internal reconstruction on this alternation, we would want to say that at some time in the history of English, these prefixes were invariably [in-] and [un-], and at some point, there was a change where the [n] assimilated to the following consonant. However, that would not be right. The assimilation had already happened in Latin, the language from which these words were borrowed. The assimilation patterns were borrowed along with with the words, and the change was not part of English at all.

7.4 SUMMARY: PROCEDURES FOR INTERNAL RECONSTRUCTION

While we have noted many problems with the application of internal reconstruction, it's also a powerful method and useful in several circumstances.

Here is a summary of the steps to use in applying the method. This set of instructions is from Harold Koch.

1. Assemble a set of tentative cognate alternating morphs in a single language. To qualify as tentative cognates, the morphs must exhibit similarities in both their semantic and their phonological makeup that could possibly be accounted for by phonological changes (avoid suppletive allomorphs such as *go* and *went*).
2. Match the tentative cognate morphs segment by phonological segment.
3. Isolate the matched phonological segments which distinguish the (allo)morphs. These are called alternating phonemes, and are in a relationship of morphophonemic alternation.
4. For each set of alternating phonemes, identify the phonological (and morphological) environment of each alternant. Be prepared to posit a phonological environment which is different from that of the attested language stage.
5. For each set of alternating phonemes, posit (a) a prephoneme in an earlier stage of the language and (b) a chronologically ordered set of changes which will transform the pre-phoneme into the attested phoneme in each of the morphs.
6. Prefer the most plausible solution for the sequence of changes. The evidence for plausibility comes from typology; the plausibility of the changes is judged by the evidence of diachronic typology. Prefer the most economical solution—that is, the solution that involves the fewest changes between the prelanguage and the attested language.

Reading Guide Questions

1. The comparative method and the method of internal reconstruction appear to be quite different. Can you find any similarities between them?
2. When might you want to use internal reconstruction instead of the comparative method?
3. What is a language isolate?
4. What sort of data do we take as the basis for applying the method of internal reconstruction?
5. What assumptions do we operate under when we apply the internal method of reconstruction?
6. Can all cases of morphological alternation be reconstructed as resulting from sound changes having taken place?
7. What are some of the problems in using the internal method of reconstruction?
8. When does internal reconstruction provide the wrong answer?

Exercises

1. Examine the following forms in southern Paamese (spoken in Vanuatu) and use the method of internal reconstruction to re-create the original root forms of the words, then state what changes have taken place.

| | | | | | |
|-------|-----------|---------|----------------|---------|--------------------|
| aim | 'house' | aimok | 'this house' | aimos | 'only the house' |
| ahat | 'stone' | ahatuk | 'this stone' | ahatus | 'only the stone' |
| ahin | 'woman' | ahinek | 'this woman' | ahines | 'only the woman' |
| atin | 'cabbage' | atinuk | 'this cabbage' | atinus | 'only the cabbage' |
| atas | 'sea' | atasik | 'this sea' | atahis | 'only the sea' |
| metas | 'spear' | metasok | 'this spear' | metasos | 'only the spear' |
| ahis | 'banana' | ahisik | 'this banana' | ahisis | 'only the banana' |
| ahis | 'rifle' | ahisuk | 'this rifle' | ahisus | 'only the rifle' |

2. Examine the following data from Bislama (spoken in Vanuatu) in which the roots and the transitive verbs derived from these are presented. State what you think the original form of the transitive suffix might have been, and state what changes have taken place.

| Root | | Transitive Verb | |
|-------|------------|-----------------|-------------------------|
| rit | 'read' | ritim | 'read' |
| bon | 'burnt' | bonem | 'burn' |
| smok | 'smoke' | smokem | 'smoke' |
| skras | 'itch' | skrasem | 'scratch' |
| slak | 'loose' | slakem | 'loosen' |
| stil | 'steal' | stilim | 'steal' |
| rus | 'barbecue' | rusum | 'barbecue' |
| tait | 'tight' | taitem | 'tighten' |
| boil | 'boil' | boilem | 'boil' |
| draun | 'sink' | draunem | 'push underwater' |
| ciki | 'cheeky' | cikim | 'give cheek to' |
| pe | 'payment' | pem | 'pay for' |
| rere | 'ready' | rerem | 'prepare' |
| drai | 'dry' | draim | 'dry' |
| melek | 'milk' | melekem | 'squeeze liquid out of' |
| level | 'level' | levelem | 'level out' |

3. Examine the following Huli (Southern Highlands, Papua New Guinea) numerals, which are given in their basic forms used in counting, as well as

their ordinal forms (first, second, third, etc). Reconstruct the original ordinal suffix, and state what changes have taken place.

Counting Ordinal

| | | |
|--------|----------|----------|
| tebo | tebone | 'three' |
| ma | mane | 'four' |
| dau | dauni | 'five' |
| waraga | waragane | 'six' |
| ka | kane | 'seven' |
| hali | halini | 'eight' |
| di | dini | 'nine' |
| pi | pini | 'ten' |
| hombe | hombene | 'eleven' |

4. Examine the following forms, again from Huli. Reconstruct the original verb roots and the original pronominal suffixes, and state what changes have taken place.

| | | | |
|---------|----------------------|----------|---------------------|
| ebero | 'I am coming' | wiru | 'I put' |
| ebere | 'you are coming' | wija | 'he/she put' |
| ibira | 'he/she is coming' | widaba | 'put everyone!' |
| ibiru | 'I came' | homaro | 'I am dying' |
| ibiri | 'you came' | homare | 'you are dying' |
| ibija | 'he/she came' | homara | 'he/she is dying' |
| ibidaba | 'come everyone!' | homaru | 'I died' |
| laro | 'I am speaking' | homari | 'you died' |
| lare | 'you are speaking' | homaja | 'he/she died' |
| lara | 'he/she is speaking' | homadaba | 'everyone die!' |
| laru | 'I spoke' | biraro | 'I am sitting' |
| lari | 'you spoke' | birare | 'you are sitting' |
| laja | 'he/she spoke' | birara | 'he/she is sitting' |
| ladaba | 'speak everyone!' | biraru | 'I sat' |
| wero | 'I am putting' | birari | 'you sat' |
| were | 'you are putting' | biraja | 'he/she sat' |
| wira | 'he/she is putting' | biradaba | 'sit everyone!' |

5. Linguists sometimes use the evidence provided in rhyming poetry to justify their conclusions about the pronunciations of words in the past. The

following nursery rhymes contain nonrhyming words. What do you think they can tell us about the history of English?

- a. Ride a cock-horse
To Banbury Cross
To see a fine lady
Upon a white horse
Rings on her fingers
And bells on her toes
She shall have music
Wherever she goes
- b. Jack and Jill
Went up the hill
To fetch a pail of water
Jack fell down
And broke his crown
And Jill came tumbling after
- c. Old Mother Hubbard
Went to the cupboard
To get her poor doggie a bone
But when she got there
The cupboard was bare
So the poor doggie had none
- d. Hickory dickory dock
The mouse ran up the clock
The clock struck one
The mouse ran down
Hickory dickory dock

Further Reading

- Robert J. Jeffers and Ilse Lehiste, *Principles and Methods for Historical Linguistics*, chapter 3, "Internal Reconstruction," pp. 37–53.
- Raimo Anttila, *An Introduction to Historical and Comparative Linguistics*, chapter 12, "Internal Reconstruction," pp. 264–73.
- Winfred P. Lehmann, *Historical Linguistics*, chapter 6, "The Method of Internal Reconstruction," pp. 99–106.

- Hans Henrich Hock, *Principles of Historical Linguistics*, chapter 17, "Internal Reconstruction," pp. 532–55.
- Talmy Givón, "Internal Reconstruction: As Method, as Theory."
- Alice Harris, "Reconstruction in Syntax: Reconstruction of Patterns"
- Anthony Fox, *Linguistic Reconstruction*, chapter 7, "Internal Reconstruction," pp. 145–216.
- Spike Gildea, *Reconstructing Grammar*.
- Don Ringe, "Internal Reconstruction."

Computational and Statistical Methods

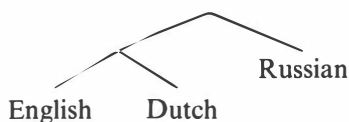
Since 2000, there has been increasing use of computational techniques in historical linguistics, particularly programs adapted from computational biology. One statistical method for computing language relationships has been around for more than 50 years, and that is LEXICOSTATISTICS. When people think of quantitative methods in linguistics, they tend to think of lexicostatistics and glottochronology. Lexicostatistics and glottochronology do not have a good reputation in standard historical linguistics, but they are not the only quantitative methods we can use. In this chapter, we look at lexicostatistics, glottochronology, and several more recent phylogenetic methods that have their origins in evolutionary biology.¹

8.1 DISTANCE-BASED VERSUS INNOVATION-BASED METHODS

First, it is useful to distinguish two types of methods for making hypotheses about the past. Up to now, we have been doing reconstruction by comparing attested languages, working out the correspondences, and figuring out a set of changes which are likely to have happened between the protolanguage and the modern languages. This is called an INNOVATION-BASED METHOD. That is,

the groupings among the languages emerge from the common changes that are reconstructed.

Another type of method is used in subgrouping (but *not* in reconstruction). This method exploits the fact that closely related languages usually have more material in common with each other than they do with the languages to which they are less closely related. For example, English and Dutch share many items of vocabulary and morphology with each other, far more than they share with Russian. This is obvious from even a very rapid and superficial comparison of the languages. On this basis, we can draw the following partial family tree:



This method of inferring relationships does not require reconstruction because it relies on what items languages have in common, not where the commonalities came from. Methods that hypothesize relationships in this way are called **DISTANCE-BASED METHODS** because they infer the historical relationships from the linguistic distance between languages. Lexicostatistics is a commonly used distance-based method. As we will see below, both of these methods have advantages and disadvantages.

8.2 LEXICOSTATISTICS

Lexicostatistics is often used with languages for which there are limited amounts of data available. Since Melanesia and Australia are areas of great linguistic diversity, and because comparatively few of these languages are well known to linguists, this technique has been used very frequently in trying to determine the nature of interrelationships in that part of the world (though this technique is not frequently used when comparing better-known languages). We therefore need to have a good understanding of how linguists have applied this technique, as well as the strengths and weaknesses of the technique as it has been applied.

8.2.1 Basic Vocabulary

Lexicostatistics allows us to determine the degree of relationship between two languages, simply by comparing the vocabularies and determining the degree of similarity between them. This method operates under two basic assumptions. The first assumption is that some vocabulary items are much less subject to lexical change than others: that is, certain parts of the lexicon include words that are less likely to be completely replaced by noncognate

forms. The area of the lexicon that is assumed to be more resistant to lexical change is referred to as CORE VOCABULARY or BASIC VOCABULARY. Underlying the lexicostatistical method is the fact that this core of relatively change-resistant vocabulary is the same for all languages. The universal core vocabulary includes items such as pronouns, numerals, body parts, geographical features, basic actions, and basic states. Items like these are unlikely to be replaced by words copied from other languages because all people, whatever their cultural differences, have eyes, mouths, and legs and know about the sky and clouds, the sun and moon, stones and trees, and so on. Other concepts, however, may be culture-specific, or known only to people of certain cultures. The word 'canoe', for example, is culture-specific because somebody who grew up in the desert of central Australia would be unlikely to have a word to express this meaning in their language. Similarly, the word for 'boomerang' would also be culture-specific because not all cultures have such implements. Such words are generally found much more likely to have been copied. In fact, the English word 'boomerang' was borrowed from an Australian language about 200 years ago; it is probably from the Dharawal word *bumaraj*.

The contrast between the amount of lexical change that takes place in the core vocabulary as against the peripheral vocabulary (or the general vocabulary) can be seen by looking at the vocabulary of English. If you take the dictionary of English as a whole, you will find that about 50 percent of the words have been copied from other languages. Most of these have been copied directly from French, as there has been massive lexical influence from French on English since the Norman Invasion of 1066. Many other words have been copied from forms that were found in ancient Latin and Greek. French has also taken many words from the same languages, which makes the lexicons of English and French appear even more similar, even with words that were not directly copied from French into English. However, if we restrict ourselves just to the core vocabularies of French and English, we find that there is much less sharing of cognate forms, and the figure for words copied from French into English in this area of the lexicon drops to as low as 6 percent (depending on what is counted as "basic vocabulary").

The second assumption of the lexicostatistical method is that the actual rate of lexical replacement in the core vocabulary is more or less stable and is therefore about the same for all languages over time. In peripheral vocabulary, of course, the rate of lexical replacement is not stable at all and may be relatively fast or slow, depending on the nature of cultural contact between speakers of different languages. This second assumption has been tested in 13 languages for which there are written records going back over long periods of time. It has been found that there has been an average vocabulary retention of 80.5 percent every 1,000 years. That is to say, after 1,000 years, a language will have lost about one-fifth of its original basic vocabulary and replaced it with new forms.²

If these assumptions are correct, then it should be possible to work out the degree of relationship between two languages by calculating the degree

of similarity between their core vocabularies. If the core vocabularies of two languages are relatively similar, then we can assume that they have diverged quite recently and therefore belong to a lower-level subgroup. If, on the other hand, their core vocabularies are relatively dissimilar, then we can assume that they must have diverged at a much earlier time and therefore belong to a much higher level of subgrouping.

8.2.2 Subgrouping Levels

Different levels of subgrouping have been given specific names by lexicostatisticians, as follows:

| Level of Subgrouping | Shared Cognate Percentage in Core Vocabulary |
|----------------------------|---|
| Dialects of a language | 81–100 |
| Languages of a family | 36–81 |
| Families of a stock | 12–36 |
| Stocks of a microphylum | 4–12 |
| Microphyla of a mesophylum | 1–4 |
| Mesophyla of a macrophylum | 0–1 |

You should note immediately that lexicostatisticians are using the term “family” in a completely different way from the way we have been using it in this textbook. I (and most other historical linguists) take “family” to refer to all languages that are descended from a common ancestor language, no matter how closely or distantly related they are to each other within that family. According to a lexicostatistical classification, however, a family is simply a particular level of subgrouping in which the members of that subgroup share more than 36 percent of their core vocabularies. Languages that are in lesser degrees of relationship (but still presumably descended from a common ancestor) are not considered to be in the same family but in the same stock or phylum.

8.2.3 Applying the Method

Having outlined the assumptions behind lexicostatistics and the theory behind its application, I now go on to show how lexicostatisticians have followed this method. The first problem is to distinguish the so-called core vocabulary from the peripheral vocabulary. I gave some indication earlier about the kinds of words that would need to go into such a list. But how

long should it be? Some have argued that we should use a 1,000-word list, others a 200-word list, and others a 100-word list. (Notice how the lengths of these lists all involve numbers that can easily be divided by 100 to produce a percentage. One suspects that these lists are not being drawn up according to any firm linguistic criterion about what can be shown to be “basic” as against “peripheral” vocabulary but merely to make the lexicostatisticians’ task of calculation easier.) It would be awkward to insist on a 1,000-word list for the languages of Australia and Melanesia where many languages are only very sketchily recorded and linguists do not have access to word lists of this length. Many people think that a 100-word list is too short and the risk of error is too great, so most lexicostatisticians tend to operate with 200-word lists.³ The most popular list of this length is known as the Swadesh list (table 8.1), which is named after the linguist Morris Swadesh who drew it up in the early 1950s.

Even with this list, there are problems in applying it to some of the languages of Melanesia, Australia, and the South Pacific. First, it contains words like *and* and *in*, which in some of these languages are not expressed as separate words but as affixes. It contains the separate words *woman* and *wife*, even though in many languages both of these meanings are expressed by the same word. It contains words such as *freeze* and *ice*, which are clearly not applicable in languages spoken in tropical areas. Other words could be included in a basic vocabulary for Pacific languages but would not be suitable for other languages, for example: *canoe*, *bow* and *arrow*, *chicken*, *pig*, and so on. A basic vocabulary for Australian languages could, of course, include items such as *grey kangaroo* and *digging stick*. These days, when linguists do lexicostatistics, they often use locally adapted versions of the Swadesh list. *Snow* is not a basic vocabulary item in tropical areas, for example, and kinship terms may be frequently borrowed in areas where the cultural norm is to marry out of one’s clan group.

Let us leave aside the problem of exactly what should be considered basic vocabulary and go on to see how we use a basic word list of this kind in a language in order to determine its relationship to another language. The first thing that you have to do is to examine each pair of words for the same meaning in the two languages, to see which ones are cognate and which ones are not. Ideally, whether or not a pair of words are cognate should be decided only after you have worked out the systematic sound correspondences between the two languages. If there are two forms that are phonetically similar but show an exceptional sound correspondence, you should assume that there has been lexical copying, and the pair of words should be excluded from consideration. It is important that you exclude copied (or borrowed) vocabulary when you are working out lexicostatistical figures, as these can make two languages appear to be more closely related to each other than they really are.

Let us now look at an actual problem. I use the lexicostatistical method to try to subgroup the following three languages from Central Province in Papua New Guinea: Koita, Koiari, and Mountain Koiari (table 8.2). Rather

TABLE 8.1 The Swadesh List

| | | | | | |
|----------|---------|------------|---------------|-------------|--------|
| all | dull | heart | neck | skin | turn |
| and | dust | heavy | new | sky | twenty |
| animal | | here | night | sleep | two |
| ashes | ear | hit | nose | small | |
| at | earth | hold/take | not | smell | vomit |
| | eat | horn | | smoke | |
| back | egg | how | old | smooth | walk |
| bad | eight | hundred | one | snake | warm |
| bark | eye | hunt | other | snow | wash |
| because | | husband | | some | water |
| belly | fall | | person | spear | we |
| big | far | I | play | spit | wet |
| bird | fat | ice | pull | split | what |
| bite | father | if | push | squeeze | when |
| black | fear | in | | stab/pierce | where |
| blood | feather | | rain | stand | white |
| blow | few | kill | red | star | who |
| bone | fight | knee | right/correct | stick | wide |
| breast | fire | know | right side | stone | wife |
| breathe | five | | river | straight | wind |
| brother | float | lake | road | suck | wing |
| burn | flow | laugh | root | sun | wipe |
| | flower | leaf | rope | swell | with |
| child | fog | left side | rotten | swim | woman |
| claw | foot | leg | rub | | woods |
| clothing | four | live | | tail | work |
| cloud | freeze | liver | salt | ten | worm |
| cold | fruit | long | sand | that | |
| come | full | louse | say | there | ye |
| cook | | | scratch | they | year |
| count | give | man/male | sea | thick | yellow |
| cut | good | many | see | thin | |
| | grass | meat/flesh | seed | think | |
| dance | green | moon | seven | this | |
| day | guts | mother | sew | thou | |
| die | | mountain | sharp | three | |
| dig | hair | mouth | shoot | throw | |
| dirty | hand | | short | tie | |
| dog | he | name | sing | tongue | |
| drink | head | narrow | sister | tooth | |
| dry | hear | near | sit | tree | |

TABLE 8.2 Data from Central Province Languages, Papua New Guinea

| | Koita | Koiari | Mountain Koiari | |
|-----|--------|--------|-----------------|------------------------|
| 1. | yata | ata | maraha | 'man' |
| 2. | mayi | mavi | keate | 'woman' |
| 3. | moe | moe | mo | 'child' |
| 4. | yamika | vami | mo ese | 'boy' |
| 5. | mobora | mobora | koria | 'husband' |
| 6. | mabara | mabara | keate | 'wife' |
| 7. | mama | mama | mama | 'father' |
| 8. | neina | neina | neina | 'mother' |
| 9. | da | da | da | 'I' |
| 10. | a | a | a | 'you (<i>sing.</i>)' |
| 11. | au | au | ahu | 'he, she, it' |
| 12. | omoto | kina | kina | 'head' |
| 13. | hana | homo | numu | 'hair' |
| 14. | uri | uri | uri | 'nose' |
| 15. | ihiko | ihiko | gorema | 'ear' |
| 16. | meina | neme | neme | 'tongue' |
| 17. | hata | auki | aura | 'chin' |
| 18. | ava | ava | aka | 'mouth' |
| 19. | dehi | gadiva | inu | 'back' |
| 20. | vasa | vahi | geina | 'leg' |
| 21. | vani | vani | fani | 'sun' |
| 22. | vanumo | koro | didi | 'star' |
| 23. | gousa | yuva | goe | 'cloud' |
| 24. | veni | veni | feni | 'rain' |
| 25. | nono | hihi | heburu | 'wind' |

than use a full 200-word list, I will make things simpler by using a shorter 25-word list and assume that it is representative of the fuller list.

The first thing that you have to do is distinguish cognate forms from forms that are not cognate. One way to do this is to mark how many cognate sets there are to express each meaning. For instance, in the word for 'man' (1), there are two cognate sets, as Koita and Koiari have forms that are clearly cognate (*yata* and *ata*, respectively), whereas Mountain Koiari has *maraha*. You can therefore label the first set as belonging to Set A and the second as belonging to Set B:

| | Koita | Koiari | Mountain Koiari | |
|----|-------|--------|-----------------|-------|
| 1. | A | A | B | 'man' |

TABLE 8.3 Cognate Sets for Data in Table 8.2

| | Koita | Koiari | Mountain Koiari | |
|-----|-------|--------|-----------------|------------------------|
| 1. | A | A | B | 'man' |
| 2. | A | A | B | 'woman' |
| 3. | A | A | A | 'child' |
| 4. | A | A | B | 'boy' |
| 5. | A | A | B | 'husband' |
| 6. | A | A | B | 'wife' |
| 7. | A | A | A | 'father' |
| 8. | A | A | A | 'mother' |
| 9. | A | A | A | 'I' |
| 10. | A | A | A | 'you (<i>sing.</i>)' |
| 11. | A | A | A | 'he/she/it' |
| 12. | A | B | B | 'head' |
| 13. | A | B | C | 'hair' |
| 14. | A | A | A | 'nose' |
| 15. | A | A | B | 'ear' |
| 16. | A | B | B | 'tongue' |
| 17. | A | B | C | 'chin' |
| 18. | A | A | B | 'mouth' |
| 19. | A | B | C | 'back' |
| 20. | A | B | C | 'leg' |
| 21. | A | A | A | 'sun' |
| 22. | A | B | C | 'star' |
| 23. | A | B | C | 'cloud' |
| 24. | A | A | A | 'rain' |
| 25. | A | B | C | 'wind' |

On the other hand, the word for 'chin' (17) is quite different in all three languages, so we would need to recognize three different cognate sets:

| | Koita | Koiari | Mountain Koiari | |
|-----|-------|--------|-----------------|--------|
| 17. | A | B | C | 'chin' |

Finally, the word for 'sun' (21) is clearly cognate in all three languages, so you would need to recognize only a single cognate set:

| | Koita | Koiari | Mountain Koiari | |
|-----|-------|--------|-----------------|-------|
| 21. | A | A | A | 'sun' |

Table 8.3 sets out the cognate sets for each of these three languages on the basis of the information that I have just given you.

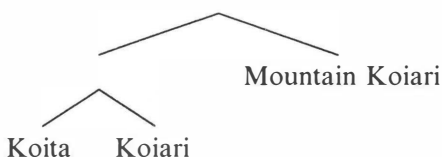
Now you need to work out the degree to which each pair of languages among the three shares cognates. First, examine the pair Koita and Koiari. If you count the number of pairs in these two languages marked as cognate (i.e., which are both marked A) and those marked as noncognate (i.e., in which one is marked A and the other is marked B), you will find that there are 16 forms that are shared between the two languages and 9 which are not. From this, you can say that 16/25 of the core vocabulary of these two languages is cognate. If you do this for the remaining pairs of languages from the three languages that we are considering, you will end up with three fractions, which can be set out in the following way:

| | | |
|-------|--------|-----------------|
| Koita | | |
| 16/25 | Koiari | |
| 9/25 | 12/25 | Mountain Koiari |

You should now convert these figures to percentages:

| | | |
|-------|--------|-----------------|
| Koita | | |
| 64% | Koiari | |
| 36% | 48% | Mountain Koiari |

Now that you have the cognate percentage figures, you need to know how to interpret them. Clearly, Koita and Koiari are more closely related to each other than either is to Mountain Koiari. On the basis of these figures, you could therefore draw a family tree of the following kind:



In terms of the degrees of relationship that I talked about earlier, these languages would all be contained within a single family; they share between 36 percent and 81 percent of their core vocabularies.

This was a simple example because we considered only three languages. Although the same principles apply when we are considering cognate percentages for larger numbers of languages, the procedures for working out the degrees of relationship can become more complex. Let us take the following lexicostatistical figures for 10 hypothetical languages and interpret the data according to these same principles:

| | | | | | | | | | |
|-----|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| A | | | | | | | | | |
| 91% | B | | | | | | | | |
| 88 | 86% | C | | | | | | | |
| 68 | 62 | 64% | D | | | | | | |
| 67 | 65 | 66 | 63% | E | | | | | |
| 55 | 51 | 56 | 53 | 55% | F | | | | |
| 57 | 53 | 54 | 57 | 56 | 89% | G | | | |
| 23 | 27 | 36 | 31 | 32 | 30 | 29% | H | | |
| 25 | 28 | 33 | 29 | 27 | 34 | 22 | 88% | I | |
| 31 | 22 | 30 | 27 | 28 | 26 | 28 | 86 | 89% | J |

Where do you start from in a more complicated case like this? The first step is to try to find out which languages in the data are most closely related to each other. To do this, you should look for figures that are significantly higher than any other figures in the table, which is an indication that these particular pairs of languages are relatively closely related to each other. In this chart, therefore, the sets of figures that are set in bold type are noticeable in this respect:

| | | | | | | | | | |
|------------|------------|----------|----------|----------|------------|----------|------------|------------|----------|
| A | | | | | | | | | |
| 91% | B | | | | | | | | |
| 88 | 86% | C | | | | | | | |
| 68 | 62 | 64% | D | | | | | | |
| 67 | 65 | 66 | 63% | E | | | | | |
| 55 | 51 | 56 | 53 | 55% | F | | | | |
| 57 | 53 | 54 | 57 | 56 | 89% | G | | | |
| 23 | 27 | 36 | 31 | 32 | 30 | 29% | H | | |
| 25 | 28 | 33 | 29 | 27 | 34 | 22 | 88% | I | |
| 31 | 22 | 30 | 27 | 28 | 26 | 28 | 86 | 89% | J |

Communities A, B, and C are clearly closely related to each other. Communities F and G also belong together, and so do the three communities H, I, and J.

Now you need to find out what is the next level of relationship. To make this task easier, you can treat the subgroups that you have just arrived at as single units for the purpose of interpretation. To do this, you should relabel the units so that it is clear to you that you are operating with units at a different level of subgrouping. You can use the following labels:

A, B, C I
 D II
 E III
 F, G IV
 H, I, J V

Now work out the shared cognate percentages between these five different lower-level units, in order to fill in the information in the chart below:

I
 — II
 — III
 — IV
 — V

Where the new label corresponds to a single language in the original chart, you can simply transfer the old figures across to the appropriate places in the new chart:

I
 — II
 63% III
 — IV
 — V

However, where the new labels correspond to a number of different communities in the original chart, you will need to get the averages of the shared cognate figures in each block and enter them in the appropriate place in the new one. So, in comparing I and II, you will need to get the figures for the shared cognates of A with D, of B with D, and C with D and enter the average of those figures under the intersection of I and II. Since A and D have 68 percent cognate sharing, B and D have 62 percent, and C and D share 64 percent of their cognates, the average level of cognate sharing between I and II works out at 65 percent. So, you can now add one more figure to the chart:

— I
 65% II
 63% III
 — IV
 — V

If you do this methodically for every pair of groupings, you will end up with the following chart:

| | | | | |
|-----|-----|-----|-----|---|
| I | | | | |
| 65% | II | | | |
| 66 | 63% | III | | |
| 54 | 55 | 55% | IV | |
| 28 | 29 | 29 | 27% | V |

You should now treat this chart in the same way as you treated the first one—simply look for the highest cognate figures as an indication of the next level of linguistic relationship. From these figures, it seems that I, II, and III are more closely related to each other than to either IV or V, as the shared cognate percentages range above 60 percent, whereas they are in the 20–60 percent range for the other groups. For the next step, you should group together I, II, and III in the same way and relabel them (this time as, say, X, Y, and Z) as follows:

| | |
|------------|---|
| I, II, III | X |
| IV | Y |
| V | Z |

Once again, calculate the averages of the cognate figures, which will work out to be as follows:

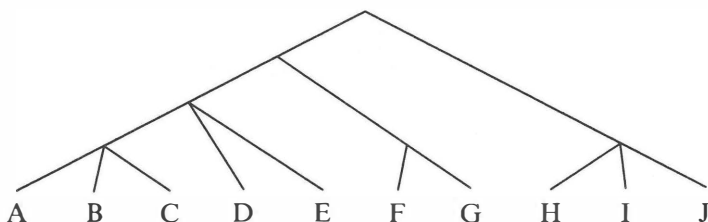
| | | |
|-----|-----|---|
| X | | |
| 55% | Y | |
| 29% | 27% | Z |

It is clear from this final chart that X and Y are more closely related to each other than either is to Z.

The final step in the procedure is to gather all of these facts together and represent the conclusions on a family tree that will clearly indicate the degrees of relationship among the 10 original speech communities. At the lowest level of relationship, you will discover that the following units belong together, while D and E are on their own:

| |
|---------|
| A, B, C |
| F, G |
| H, I, J |

At the next level of relationship, you find that D and E belong to the same group as A, B, and C, while the others are all on their own. At the next level, F and G could be related to the same subgroup as the subgroup consisting of A, B, C, D, and E, with H, I, and J being a separate subgroup of their own. This situation can be represented in a family tree diagram in the following way:



Having dealt in some detail with the claim that lexicostatistics enables us to work out degrees of relationship within a language family, I now go on to discuss a second claim that lexicostatisticians sometimes make, though most linguists are now very cautious about this. If we accept the basic assumption that languages change their core vocabulary at a relatively constant rate, we should be able to work out not only the degree of relationship between two languages but also the actual period of time that two languages have been separated from each other. Once the percentage of cognate forms has been worked out, we could use a formula to work out the TIME DEPTH, or the period of separation, of two languages.⁴ This is known as GLOTTOCHRONOLOGY.

The claim of glottochronology is that languages replace approximately 20 percent of their core lexicon over a 1,000-year period. Therefore, we would expect languages that share approximately 80 percent of their vocabulary to have diverged somewhere around 1,000 years ago. Glottochronology is to a certain extent the linguistic equivalent of the DNA clock. We know that mutations in DNA occur at a roughly constant rate, and we know what that rate is. Therefore by comparing the genetic difference between two species, we know approximately how long ago their lineages diverged. By comparing lots of different species, we can build a DNA tree of species.

The mathematical formula used for calculating time depth using glottochronology is a simple decay formula of the form

$$t = \frac{\log C}{2 \log r}$$

where t stands for the number of thousands of years that two languages have been separated, C stands for the percentage of cognates as worked out by comparing basic vocabulary, and r stands for the constant rate of change mentioned in sec. 8.2.1 (i.e., 0.805). Going back to the earlier

problem involving Koita, Koiari, and Mountain Koiari, if you wanted to know how long it has been since Koiari split off from Koita, it would take the cognate percentage of 64 percent (the figure given in the table for these two languages) and convert it to a factor of one (0.64) and apply the formula:

$$\begin{aligned}
 t &= \frac{\log C}{2 \log r} \\
 &= \frac{\log .64}{2 \log .805} \\
 &= \frac{.446}{2 \times .217} \\
 &= 1.028
 \end{aligned}$$

This means that Koita and Koiari are calculated to have diverged 1.028 thousand years ago, or 1,028 years ago.

8.3 CRITICISMS OF LEXICOSTATISTICS AND GLOTTOCHRONOLOGY

The techniques of lexicostatistics and glottochronology have not been without their critics. I have already hinted at a number of practical problems that are associated with these methods. First, there is the problem of deciding which words should be regarded as core vocabulary and which should not. Obviously, it may be possible for different sets of vocabulary to produce differing results. Another difficulty involves the actual counting of forms that are cognate against those that are not cognate in basic vocabulary lists from two different languages. As I said earlier, ideally, copied vocabulary should be excluded from cognate counts, but to do this you need to know what the regular sound correspondences are between the two languages in order to exclude exceptional forms, which are probably copied. However, since we are working with fairly short word lists, there may not be enough data to make generalizations about sound correspondences. Also, we are not likely to know much about the protolanguage if we are dealing with languages for which we have only limited amounts of data, and this will make it even more difficult to distinguish genuine cognates from copied vocabulary.

Lexicostatisticians, in fact, tend to rely heavily on what is often euphemistically called the "inspection method" of determining whether or not two forms are cognate in a pair of languages. What this amounts to is that you are more or less free to apply intelligent guesswork as to whether you think two forms are cognate. If two forms look cognate, then they can be

given a "yes" score, but if they are judged not to look cognate, then they are given a "no" score.

Of course, two different linguists can take the same lists from two different languages, and since there is no objective way of determining what should be ticked yes and what should be ticked no, it is possible that both will come up with significantly different cognate figures at the end of the exercise. For example, I have done counts on the basis of word lists calculated by other people and have ended up with figures between 10 percent and 20 percent higher or lower than their count. If two different scholars compare the same pair of languages and one comes up with a figure of 35 percent cognate sharing and the other concludes that there is 45 percent cognate sharing, then one is going to have to say that the two represent different families within the same stock, while the other will end up saying that they are from two languages within the same family. In glottochronological terms, this could mean a difference in time depth of up to 600 years.

A further problem that arises in the use of lexicostatistical figures to indicate degrees of linguistic relationship is that different linguists sometimes use different cutoff points for different levels of subgrouping, and there is not even agreement on what sets of terminology should be used to refer to different subgroups of languages. While I have used the term in this textbook to refer to all languages descended from a protolanguage, according to one system, a language family refers only to languages that share more than 36 percent of their core vocabulary, while according to another system, languages in the same family must share more than 55 percent of their core vocabularies.

Apart from these practical problems, there are some more basic theoretical objections to these methods, which tend to destroy the validity of the underlying assumptions that I presented earlier. First, we need to question the validity of the assumption that there is a constant rate of lexical replacement in core vocabulary for all languages over time and that this rate of replacement is 19.5 percent every 1,000 years. This figure was arrived at by testing only 13 of the world's languages, and these were languages with long histories of writing, 11 of which were Indo-European languages. However, differing cultural factors can affect the speed at which lexical replacement can take place. In chapter 11, I describe how lexical replacement can be accelerated in languages in which even basic vocabulary can become proscribed by taboo. The result of lexical replacement because of taboo is that even basic vocabulary, if given sufficient time, will be subject to replacement. If languages copy words from neighboring languages in order to avoid a forbidden word, two languages that were originally very different from each other will end up sharing a high proportion of even their core vocabularies. Moreover, certain cataclysmic changes can radically speed up the replacement rate or drastically reduce the amount of identifiably shared vocabulary. For example, if a language simplifies all of its consonant clusters, loses a voicing contrast, and develops tone, it may well be that many cognates are not readily apparent by the inspection

method. In this scenario, the changes would be linked, but the result would be a falsely old glottochronological figure. Second, when speakers of a language come into contact with a lot of new items, they need to innovate lexical items quickly. They do not borrow a word here or a word there. In fact, it has recently been suggested that lexical changes occur in bursts, with languages being quite stable for a long period of time and then undergoing a number of innovations relatively quickly. This would still produce an average figure per 1,000 years, but the average would not be very meaningful and could differ widely, depending on whether or not the languages had undergone the burst.

A second theoretical problem with lexicostatistics involves the interpretation of the data. Given that change is random within the core vocabulary, it is logically possible for two languages to change the same 19.5 percent of their core vocabulary every 1,000 years and to retain the remaining 80.5 percent intact over succeeding periods. It is also possible at the other extreme for two languages that in the beginning shared the same proportions of their core vocabulary to replace 19.5 percent of their core vocabularies every 1,000 years, yet for the 19.5 percent to be different in each successive period. The result of this will be that two pairs of languages, while separated by the same period of time, might have dramatically different vocabulary retention figures, depending on which items were actually replaced. Some languages will be accidentally conservative, while others will accidentally exhibit a high degree of change. Although the time depth would be the same, we would be forced to recognize two very different degrees of linguistic relationship.

8.4 SUBGROUPING COMPUTATIONAL METHODS FROM BIOLOGY

I mentioned at the start of this chapter that lexicostatistics and glottochronology are not the only quantitative methods in historical linguistics. One of the biggest problems with lexicostatistics is that it is often practiced as an alternative to the comparative method. That is, it is done in situations where preliminary comparison is needed or where the linguist, for whatever reason, does not wish to use the comparative method. However, there is potential for more sophisticated quantitative methods used in conjunction with the comparative method. In recent years, several new methods have come into linguistics, and quantitative methods are now being used on families in several different parts of the world.

8.4.1 Inferring Correspondence Sets

There are a couple of different ways in which computational methods are used in historical linguistics. One is in the discovery of cognate sets. Some

computer programs can take transcribed data and attempt to identify the cognate sets and reconstruct the changes that the languages have undergone. Other programs require the cognate sets to be aligned (as we did in chapter 5) before the computer software tries to compute the changes. I am skeptical about the utility of these programs. In my experience, the correspondence set identification programs do much worse than a human at identifying the correspondences, even worse than someone with little experience in historical linguistics. Furthermore, the process of identifying correspondence sets (as required in the second type of program) usually leads to the linguist identifying the changes in the data. That is, by the time the linguist has tagged the correspondence sets, he or she has worked out the answer without the need of the computer program!

8.4.2 Inferring Subgrouping

Most computational work in linguistics, however, does not involve the compilation of cognate sets but, rather, the analysis of relatedness. That is, the programs use data coded by the linguist to work out subgrouping within a family, to search for previously unidentified relationships, or to evaluate potential competing hypotheses. The type of program I discuss here takes data about properties of languages and uses that data to represent linguistic relationships and to draw hypotheses about when different languages split. That is, they do not explicitly try to reconstruct the sound changes or to mimic what historical linguists do; rather, these programs work in different ways to draw trees.

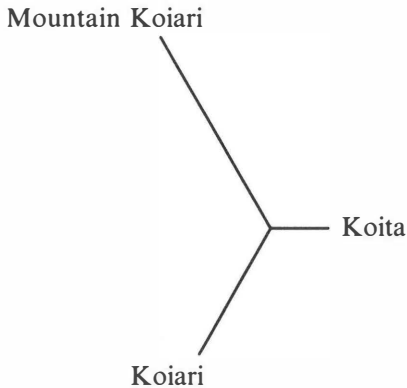
Computational methods can augment the comparative method in several ways. We should not be using these tools in areas where the comparative method does a better job. Rather, we should be exploiting the main advantage of such methods, which is to reduce the risk of researchers unintentionally biasing their results by only paying attention to certain items. This is a particular problem when families are worked on by small numbers of people and where there are few people with in-depth knowledge of languages and it is therefore harder to catch such mistakes. Another advantage of these methods is that we can use them to look for patterns in the data. For example, it is possible to look for traces of borrowing in different semantic fields quite easily using these methods. We can also compare and quantify the evidence for competing subgrouping theories. Furthermore, quantifying the results of the comparative method makes it easier for scholars in quantitative fields to interpret the results. Linguists are constantly complaining that their results are misrepresented or overlooked in the prehistory literature; here is a way to make them more accessible. Finally, not all computational methods produce trees: some produce networks. Using these methods can therefore give us an alternative model for thinking about linguistic relationships.

8.4.3 Some Definitions

The terminology of the computational historical linguistics literature is somewhat similar to mainstream historical linguistics, but some important terms originate in the field of evolutionary biology. There are also some differences in the way that the data are set up. Throughout this section, I continue to use the Koiari and Koita data from the lexicostatistics illustration.

The basic unit of comparison is the **TAXON**. Koita, Koiari, and Mountain Koiari are all taxa. Taxa can form a **CLADE**. In my example of a family tree, Koita and Koiari form a clade; Koita, Koiari, and Mountain Koiari form a clade; but Koiari and Mountain Koiari on their own do not. For our purposes here, a clade is equivalent to a subgroup.

Taxa are grouped into trees, as in the methods you've seen already. Linguistic family trees represent hypotheses about changes that have happened through time, and as such they have a protolanguage at the top and the modern languages at the bottom. This is known as a **ROOTED TREE**, because we know in this case what the root (or the base) of the tree is and what the descendants are. The family tree I gave for Koari, Moita, and Mountain Koiari is a rooted tree. Computational trees, however, are often represented as **UNROOTED TREES**. Unrooted trees show which languages are closer to one another, but they do not show a single protolanguage from which all of the subsequent languages developed. This is what an unrooted tree of our example languages looks like:



Finally, I mentioned earlier that trees are not the only way of representing relationships. Relationships can also be expressed using **NETWORKS**. The main difference between trees and networks is that while nodes in a tree can only have one parent, that is not true for networks. Figure 8.1 shows a network of Quechua and Aymara varieties. Networks are useful for showing how much unambiguous support there is for a particular type of branching. On the one hand, a network that looks rather like a tree implies that there is

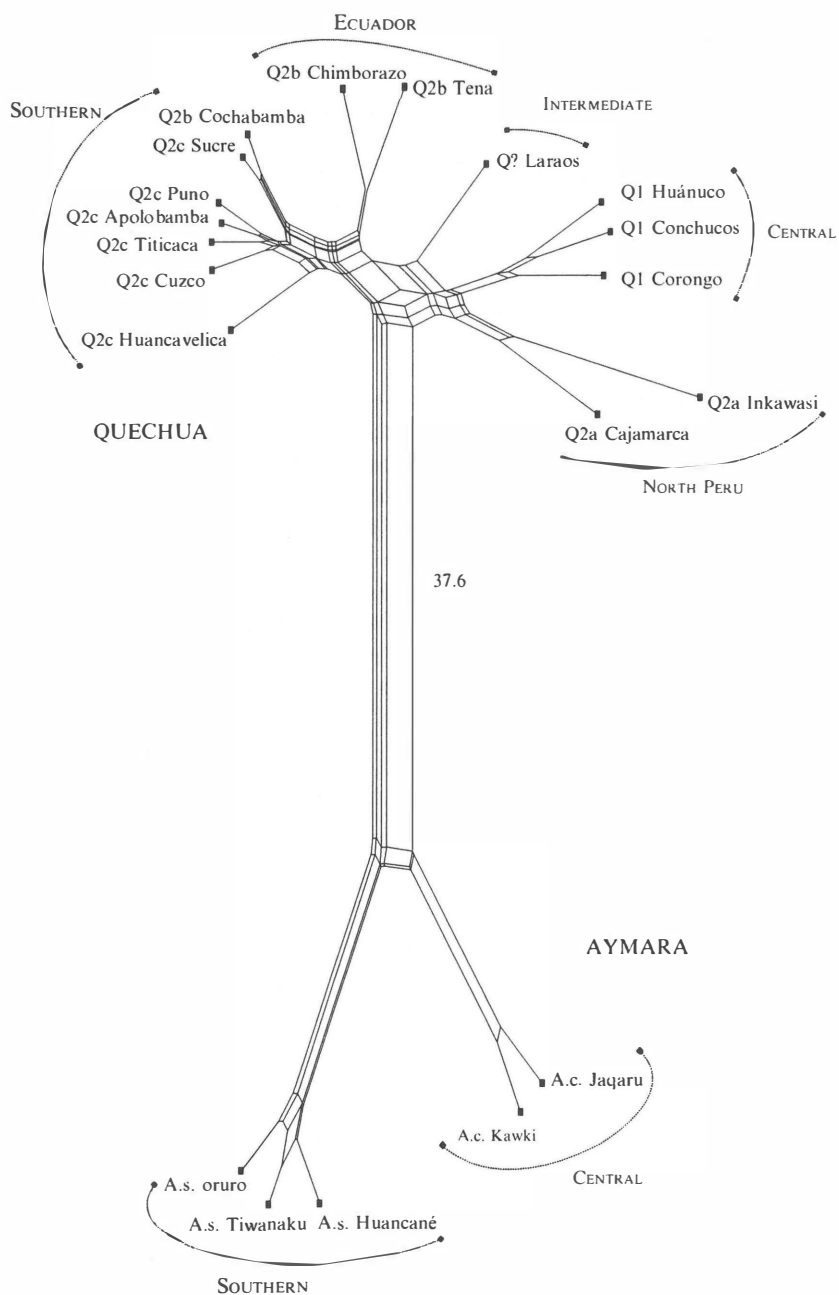


FIGURE 8.1 Network of Quechua and Aymara varieties (Heggerty 2008:44)

a good deal of consensus in the data for the splits. In figure 8.1, the split between Quechua on the one hand and the two branches of Aymara on the other illustrate this. On the other hand, a network that is very “weblike” implies that there are multiple ways to hypothesize the relationships between taxa.

The terminology used to talk about the data itself is also different. Up until now in this book, we have talked about correspondence sets. In the computational historical linguistic literature, the items being compared are called CHARACTER SETS. The characters are the features of the languages used for comparison.

8.4.4 Selecting Data and Coding Characters

In order to use these methods, you need to prepare the data so the information can be processed. Twenty words is not a large-enough data sample to work on this method properly; you want as many data points as possible.⁵ This will do for an illustration, however.

The first consideration is what you are coding for. Coding can be BINARY or MULTISTATE. In BINARY CODING, you are coding for the presence or absence of a particular feature. The language is coded as 1 if the feature is present and 0 otherwise. So, in our example data, one character set might be “has a cognate of *yata* in the meaning ‘man’”. Koita and Koiari would get a 1 for this, and Mountain Koiari would get a 0. We could also have a character set meaning something like ‘has a cognate of *maraha* in the meaning ‘man’’. In that case, Mountain Koiari would get a 1, and the other languages would get a 0.

| | Koita | Koiari | Mountain Koiari | |
|-----|-------|--------|-----------------|-------|
| 1a. | 1 | 1 | 0 | ‘man’ |
| 1b. | 0 | 0 | 1 | |

In MULTISTATE CODING, the character sets are the translations, and each putative cognate set receives a different coding. The coding is similar to that used in lexicostatistics (and is therefore subject to the same problems).

| | Koita | Koiari | Mountain Koiari | |
|----|-------|--------|-----------------|-------|
| 1. | 1 | 1 | 2 | ‘man’ |

Each possible coding for a character set is called a STATE. In this example, Koita and Koiari have state 1 for the character, while Mountain Koiari shows state 2.

You'll notice that in this coding scheme, some of our data are not very informative about relationship. Items 3, 7–11, 14, 21, 24, and perhaps 18 are all cognate: they'd all get the same coding. Nine of our 25 cognates don't give us any information about subgrouping from the lexical point of view. However, there are phonological differences in the data. In item 11, Mountain Koiari has a form with *h* while the other languages have nothing; it has *f* in item 21, where the other languages have *v* (and this is repeated in 24); and Koita and Koiari show a correspondence of $\gamma : v$ in items 2 and 4. This is also information about language divergence and can be coded as a phonological character set. Here are some phonological character sets in the data:

| | Koita | Koiari | Mountain Koiari | Description |
|----|-------|--------|-----------------|-----------------------------|
| a. | 1 | 2 | — | $\gamma : \emptyset$ |
| b. | 1 | 2 | — | $\gamma : v$ |
| c. | 1 | 1 | 2 | $\emptyset : \emptyset : h$ |
| d. | 1 | 1 | 2 | $v : v : f$ |
| e. | 1 | 1 | 2 | $e : e : \emptyset$ |

If you do not have information on a particular correspondence, you should code the data as missing, as I have done here with the dashes.

In the process of coding the data, you will need to make some decisions about what is cognate and what is not. In my example, I did not include as cognate the items in 18 or all the items in 23, even though I could tell a story about how these words related to one another (which may or may not be confirmed by more data from the languages). Keep notes on your judgments, and be consistent in the application of your decisions. Coding decisions are important. For example, some data codings take reconstructions as the character sets. Therefore a set might be labeled "participates in the split of **a* into /*a*/ and /*ɛ*/." While there might not be anything wrong with this at first sight (especially in well-studied families), it runs the risk of introducing circularity. This is because when linguists posit a sound change and a reconstruction, they usually have an intuition about the subgrouping. It is therefore not surprising in such cases that the computationally generated tree supports the linguists' hypothesis! To be on the safe side, it is better to code character sets as correspondences (as we did above) rather than as participants in changes.

In some versions of lexicostatistics, obvious loan words are excluded. In this method, however, excluding loans would potentially bias the results: (we only know for certain that a word is a loan once we do the reconstructions and work out the sound changes; if we eliminate some loans, we still potentially have confused data (through unidentified loans).

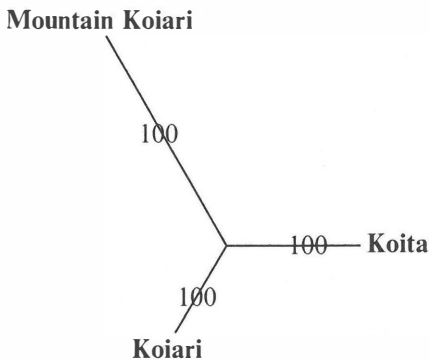
8.4.5 Methods for Inferring Phylogenies

Now that you have coded your data, the next stage is to calculate the relationships. There are a couple of ways to do this. One is to use a method such as NEIGHBORNET.⁶

First, we take the coded data and compute a distance matrix (the lexicostatistical percentage tables shown earlier in this chapter are a type of distance matrix, although that is not the only way to calculate distances). From the distance matrix, we infer a collection of tree splits. These are the possible ways of grouping together the languages in the family. For example, in the data we have been considering in this chapter, we saw that the best hypothesis was one that groupstogether Koita and Koiari. We could have a hypothetical tree that grouped together Koiari and Mountain Koiari in a subgroup, but intuitively that is a bad hypothesis. We can see that the data simply do not support it.

The second stage is therefore to use the calculated similarities between languages to generate a collection of possible subgroupings. The subgroupings are given a score; the best splits are those which minimize the distance between immediate neighbors and maximize the distance between the clustered neighbors and the rest of the tree. That is, we want to find the languages that are closest to each other (and therefore most likely to be most recently descended from a common ancestor) and to minimize the possibility that there is another language which is more close to one of them.

The NEIGHBORNET program calculates potential ways of splitting the data and scores them. The splits are then aggregated and represented in a network, like the one we saw for Aymara and Quechua in Figure 8.1. Calculating the distance matrix and splits for our Koita and Koiari data results in the following tree. The 100's on the branches indicate the confidence level. A figure of 100 means that the split is well supported. The length of the branch is also meaningful in these diagrams. The length of the branch is roughly proportional to the degree of difference (and therefore to the amount of change) between the two languages. In this example, the lengths imply that Mountain Koiari has undergone more changes than Koiari or Koita.



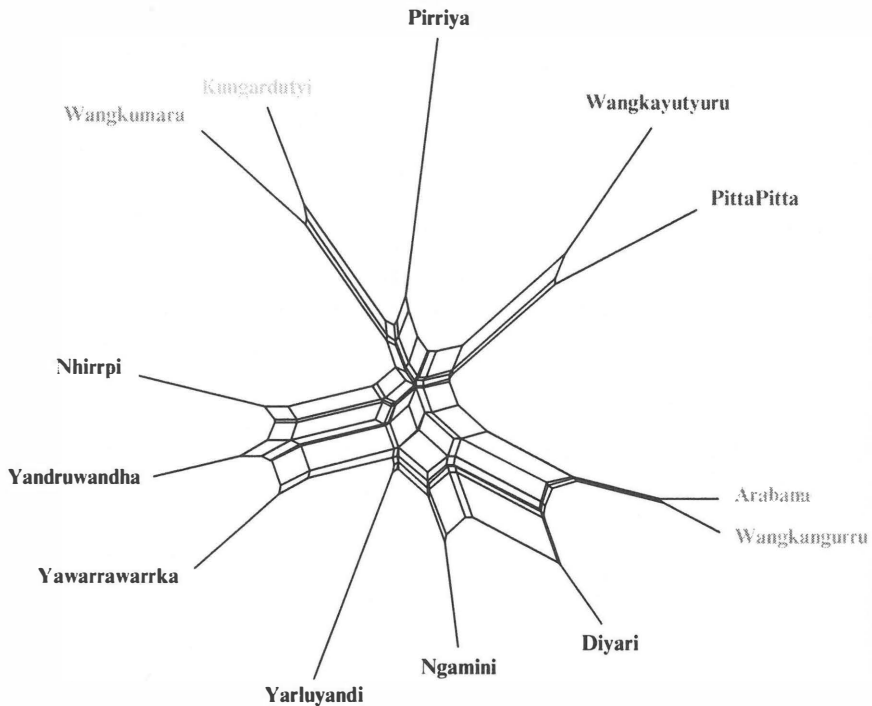


FIGURE 8.2 Karnic network

This example is not too spectacular because the data are relatively clear here, but there are other cases where the signals for subgrouping are much more difficult to interpret. An example of a subgroup with many conflicting signals for subgrouping is given in Figure 8.2. This network shows the Karnic subgroup of Pama-Nyungan; it has been the subject of seven proposals since 1919. The conflicting signals arise because of borrowing, some parallel development, and early dialect split, which was then clouded by subsequent descent patterns. The network shown in the figure is a collection of more than 600 lexical, phonological, and morphological character sets. Subsequent analysis reveals that some parts of the data are much more treelike than others and that the messy splitting (and therefore the conflicting subgrouping hypotheses) is concentrated in certain parts of the data.

NeighborNet is one of three common methods in use for computational historical linguistics. The second type of methods are “parsimony methods.” These methods work rather differently from NeighborNets. While NeighborNets calculate splits on the basis of distance between languages, PARSIMONY METHODS try to fit as much of the data as possible onto a tree with the smallest number of branches. This is similar to the process of inference, which linguists work through when they are computing subgrouping in their heads. That is, we wish to group together the languages which are the

largest number of changes without positing needless extra steps. The “PERFECT PHYLOGENIES” method developed by Nakhleh et al. (2005) is of this type. Several articles in the further reading section describe this method in more detail.

Finally, let us now briefly look at one other computational method, used in linguistic work chiefly by Gray and colleagues (2007) and known as BAYESIAN PHYLOGENETICS. This method uses binary-coded data. The input data used is lexical replacement, and the character sets are defined as the presence or absence of a particular cognate (with a particular meaning). This presence or absence presents a change of state at some point in the devolution of the very history of the family. That is, at some point, either one or more languages lost the trait or some languages acquired it. This change is assigned to an intermediate node in the tree, and then possible trees are evaluated for how likely they are to represent the best model for the evolution of those traits. That way, a set of hypothetical trees is generated. The trees can be scored for how good a fit they are for the data. The next step in the model is to evaluate generated trees against other possible trees. This is done by randomly changing part of the tree and then evaluating the new tree produced by the change. If it has a better score than the old tree, the old tree is discarded and the procedure is repeated with the new tree (the new one is better than the old one, but that does not mean it’s the best one). The search continues until we have a set of “most probable” trees. This is because exhaustively searching for all possible “best” trees is computationally extremely intensive; therefore, searching in many random parts of the tree space allows us to strike a balance between avoiding locally (but not globally) optimal trees and being able to compute the best tree in a reasonable amount of time.

We must remember that the quality of the results is heavily dependent on the quality of the data and coding. Here I have suggested that the data coding should mimic correspondence sets as much as possible. Other work in this area has used typological features as well as or instead of lexical data. This should not be preferred because of the high likelihood of false similarities. As we see in chapter 12, certain types of grammatical changes are common in languages. Other changes can be dependent on one another. For example, the order of the object and the verb in the clause is strongly correlated with the order of nouns and adjectives and noun phrases. Therefore, if a change occurs that swaps the order of object and verb, this increases the likelihood that the order of noun and adjective also change. These dependencies between characters can introduce conflicting signals if they form the core of the input data.

Reading Guide Questions

1. What is the difference between distance-based and innovation-based methods?
2. Give an example of an innovation-based method.

3. What is lexicostatistics?
4. What basic assumptions underlie the method of determining linguistic relationships by lexicostatistics?
5. What is the inspection method of determining whether or not two forms are cognate?
6. What is the difference between core and peripheral vocabulary?
7. What is glottochronology?
8. What are some problems associated with lexicostatistics and glottochronology?
9. What is the difference between rooted and unrooted trees?
10. What is a NeighborNet?
11. What are some of the advantages of computational methods?

Exercises

1. Refer to Dataset and see if you can make any judgments about the subgrouping of Sepa, Manam, Kairiru, and Sera from lexicostatistical data.
2. Use the same dataset to get one of the free computational programs running (e.g., Splitstree, available from splitstree.org).
3. Using the data in Dataset, try to calculate lexicostatistical percentages for the family. What problems do you encounter?
4. Use your codings for lexicostatistics in the previous question to code the data for a phylogenetic method. Use binary coding.
5. Use the following dataset from made-up languages Xish, Yish, and Zish. Construct as many character sets as you can from the data.

| | Xish | Yish | Zish | |
|-----|--------|----------|--------|----------|
| 1. | waj | waj | lalp | 'arm' |
| 2. | larp | wajaθsat | lalp | 'armpit' |
| 3. | θris | θris | lis | 'blood' |
| 4. | wamwu | wamu | wamwu | 'brain' |
| 5. | lamar | lamar | lamal | 'ear' |
| 6. | maraj | lum | malaj | 'elbow' |
| 7. | ηiyal | misal | ηiyal | 'eye' |
| 8. | sid | sidsid | fjid | 'finger' |
| 9. | sanpar | sanpar | sanpal | 'foot' |
| 10. | aθsat | aθsat | asat | 'hair' |
| 11. | napu | nap | napu | 'head' |
| 12. | ratnis | ratnis | latnis | 'heart' |
| 13. | lumsi | maraj | lumfi | 'knee' |

- | | | | | |
|-----|--------|-----------|--------|-----------|
| 14. | θusa | θus | lumfi | 'leg' |
| 15. | kimsi | lipa | lipa | 'liver' |
| 16. | pulku | pulk | lanlan | 'lung' |
| 17. | θapma | θapm | apma | 'neck' |
| 18. | ruθam | ruθam | luam | 'nose' |
| 19. | tiwŋa | tiwŋ | tiwŋa | 'stomach' |
| 20. | lisθin | lisθinθin | lifin | 'toe' |

Further Reading

- Russell Gray, S.J. Greenhill, and R.M. Ross, "The Pleasures and Perils of Darwinizing Culture."
- David Bryant, F. Filimon, and Russell Gray, "Untangling Our Past: Languages, Trees, Splits and Networks."
- Luay Nakhleh, D. Ringe, and T. Warnow, "Perfect Phylogenetic Networks: A New Methodology for Reconstructing the Evolutionary History of Natural Languages."
- Theodora Bynon, *Historical Linguistics*, chapter 7, "Glottochronology (or Lexicostatistics)," pp. 266–72.
- Robert J. Jeffers and Ilse Lehiste, *Principles and Methods for Historical Linguistics*, chapter 8, "Lexicostatistics," pp. 133–37.
- Winfred P. Lehmann, *Historical Linguistics*, chapter 7, "Study of Loss in Language: Lexicostatistics," pp. 107–14.
- Sarah Gudschinsky, "The ABC's of Lexicostatistics (Glottochronology)."
- Keith Johnson, *Quantitative Methods in Linguistics*, chapter 8, "Historical Linguistics," pp. 182–215.

The Comparative Method (2): History and Challenges

The comparative method (discussed in chapter 5) was developed mainly in the 1800s, largely by German scholars. This method may seem straightforward if you carefully apply it, following the steps that I set out in that chapter. However, it can sometimes be difficult to apply the method in particular situations, and we need to take into consideration some more issues in reconstruction. Remember that I said in chapter 8 that the comparative method is not an algorithm for “discovering” protolanguages; rather, it is a set of heuristics (guiding tools) for you to use in making hypotheses about the past history of languages.

In this chapter, I look at some of the problems that linguists have come across in applying the method. I begin by looking at the historical development of the comparative method and its refinement by the Neogrammarians of the twentieth century, along with some of the difficulties in the method that they recognized from the very beginning. I then go on to look at some further reasons why treating the comparative method as algorithmic leads to problems.

9.1 BACKGROUND: THE NEOGRAMMARIANS

The comparative method was first developed in Europe, and it was first applied to the languages of the Indo-European and Finno-Ugric language families. It was perhaps natural that European scholars should investigate the history of their own languages first, as these were languages with a very long history of writing. This made it possible to start their reconstructions further back in time than they could have done with languages that were unwritten or which had only recently been written. A long history of writing also made it possible to check on the accuracy of reconstructions that had been made from the present.

After the period of European voyaging and exploration between the 1400s and the 1700s, scholars came into contact with a wide range of languages that were previously unknown in Europe. Word lists were compiled in “exotic” languages for people to see the similarities and differences between them. Before the nineteenth century, a field of inquiry called ETYMOLOGY had become quite well established. This term is currently used to refer simply to the study of the history of words, though in earlier times the history of “words” and the history of “languages” were often confused.¹

Many of the early attempts at etymology would be regarded as childish by modern standards. One French scholar, Étienne Guichard, in 1606 compiled a comparative word list in Hebrew, Chaldaic, Syrian, Greek, Latin, French, Italian, Spanish, German, Flemish, and English; he tried to show that all languages could be traced back to Hebrew! The kind of evidence that he presented to support his hypothesis was the existence of similarities between words such as Hebrew *dabar*, English *word*, and Latin *verbum*. Some scholars who followed Guichard were more skeptical of these methods, and Voltaire, a famous French writer, described etymology as the science in which “the vowels count for little and the consonants for nothing” (quoted in Muller 1864: 238). Unkind words, but true, at least as Guichard had applied it.

Late-eighteenth-century work on relationships between Sanskrit and other Indo-European languages profoundly altered the perception of the nature of linguistic relationships among serious scholars. However, this did little to stop those less concerned with these more modern views from continuing on the path of earlier commentators—I hesitate now to use the word “scholars”—in making random observations about similarities between languages as evidence of linguistic relationships. Books published in the late 1800s attempted to demonstrate the relationship between the languages of Vanuatu and those of the Middle East; this is a relationship that no modern linguist would take the slightest bit seriously, I should point out. Other scholars have taken random similarities in language and cultural artifacts as evidence that Hawaii was populated from Greenland, that parts of Polynesia were populated from South America, and that different peoples on Earth were

provided with aspects of their culture by beings from outer space. I wouldn't want to rule out these interpretations as impossible, but the linguistic evidence is certainly far from compelling, and modern linguists tend to assign these kinds of view to the lunatic fringe.

Work from this time began to place reliance on similarities in the *structure* of the Indo-European languages, rather than the individual similarities between *words*, as important for determining language relationships. This observation led to a new intellectual climate in the study of language relationships, as scholars started looking for grammatical similarities between languages to determine whether or not they should be considered to be related. Lexical similarities, it was argued, were poor evidence of genetic relationship, as similarities between practically any word in any two languages can be established with enough effort.

Rasmus Rask in 1818 investigated the history of the Icelandic language on the basis of its grammatical similarities to other Germanic languages (such as Norwegian, German, and English) and largely ignored the lexicon. Rask also argued, however, that while individual lexical similarities were not good evidence of linguistic relationship, repeated occurrences of sound correspondences between words could not be due to chance, so these were good evidence of genetic relationship. By recognizing only repeated occurrences of sound correspondences as valid evidence in the study of language, it was possible to exclude chance lexical similarities such as those noted by Guichard for Hebrew, English, and Latin.

In 1822, Jakob Grimm described a series of sound correspondences that he had noted between Sanskrit, Greek, Latin, and the Germanic languages (which also include the now-extinct Gothic language, as well as English). For instance, he noticed that, very often, where Sanskrit, Greek, and Latin had a /p/, the Germanic languages had an /f/; where Sanskrit, Greek, and Latin had a /b/, the Germanic languages had a /p/; and finally, where Sanskrit had a /bh/,² the Germanic languages had a /b/—for example:

| Sanskrit | Greek | Latin | Gothic | English |
|------------------|------------------|-----------------|---------------|----------------------------|
| pa:da | pous | pes | fotus | foot |
| — | tur be : | tur ba | þaur p | thor p ³ |
| bhra:ta : | phra:te r | fra:te r | brôþar | brother |

(You should note that we are considering only the sounds written in bold type at this point. The remaining sounds have less obvious correspondences than these, so perhaps you can appreciate the advantage in having learned to apply the comparative method using the much more straightforward correspondences that are found in the Polynesian languages. There are also some secondary changes in the words for 'brother'.) The full set of sound correspondences that Grimm noted are set out here, along with the reconstructed protophonemes:

| Proto-Indo-European | Sanskrit | Greek | Latin | Germanic |
|---------------------|----------|-------|-------|----------|
| *p | p | p | p | f |
| *t | t | t | t | θ |
| *k | c | k | k | x |
| *b | b | b | b | p |
| *d | d | d | d | t |
| *g | j | g | g | k |
| *bh | bh | ph | f | b |
| *dh | dh | th | f | d |
| *gh | jh | kh | h | g |

Germanic voiceless fricatives correspond mostly to voiceless stops in the other languages, and Germanic voiceless stops correspond to voiced stops. Germanic voiced stops have a more complicated set of correspondences, as they correspond to voiced aspirated stops in Sanskrit and voiceless aspirated stops in Greek (with the Latin correspondences being somewhat less predictable in this case).

According to the methodology that I set out in chapter 5, the forms in the left-hand column can be reconstructed for the language from which all of these languages were descended. That is, we reconstruct in the protolanguage the form that is most widely distributed in the daughter languages, and we reconstruct original forms that involve “natural” rather than “unnatural” changes. You can see that of the four descendant languages, Sanskrit is clearly the most conservative as it has undergone fewer changes in these consonants from the protolanguage (though there are plenty more changes in other aspects of the language). The Germanic languages are clearly the ones that have changed the most since Proto-Indo-European with respect to these consonants.

No scholar at the time thought to distinguish between sound correspondences that were without exception and those which appeared to be sporadic (i.e., which applied in some words but not in others). In fact, while the correspondences that Grimm noted were found to be true for very many words, there were at the same time many words in which the correspondences did not hold and other correspondences were apparent instead. For example, many voiceless stops in Sanskrit, Greek, and Latin corresponded to voiceless stops in Germanic instead of voiceless fricatives:

| Latin | Gothic | |
|--------|---------|---------|
| spuo | speiwan | ‘spit’ |
| est | ist | ‘is’ |
| noktis | naxts | ‘night’ |

The Gothic forms were not /sfeiwɑn/, /isθ/, and /naxθs/ as we might expect if the correspondences noted by Grimm were to be completely general. However, it was soon realized that the correspondence of Sanskrit, Greek, and Latin voiceless stops to Germanic voiceless stops, and Sanskrit, Greek, and Latin voiceless stops to Germanic voiceless fricatives were, in fact, in complementary distribution.

In chapter 5, you saw that when a conditioned sound change takes place in any of the daughter languages, the result is that the sound correspondence sets end up being in complementary distribution. So, once you have set out the full range of correspondence sets, you must check to see whether phonetically similar correspondence sets are in complementary or contrastive distribution. If it turns out that they are in complementary distribution, you need only reconstruct a single original phoneme that has undergone a conditioned sound change. The first of the two correspondences just mentioned was found only when Gothic had a preceding fricative, whereas the second correspondence was found when there was no preceding fricative. We can therefore reconstruct both correspondences as going back to a single voiceless stop series. This would make it necessary to reconstruct a conditioned sound change of the following form in the Germanic languages:

$$*\text{voiceless stop} > \begin{cases} \text{voiceless stop / fricative} \text{ —} \\ \text{voiceless fricative / elsewhere} \end{cases}$$

More and more sound correspondences came to be recognized as being due to the influence of phonetic factors of some kind, such as the nature of the preceding or following sounds, the position of stress, or the position of the sound in the word (i.e., whether it occurred word-initially, word-medially, or word-finally). By taking into account yet other phonetic factors, Herman Grassmann was able to account for a further set of consonant correspondences in these languages. Scholars had noted that some voiced stops in the Germanic languages corresponded to aspirated stops in Sanskrit and Greek (as covered by Grimm's statement, as you have just seen), but some voiced stops corresponded to unaspirated stops. Scholars were once again faced with a double set of correspondences.

Grassmann was able to show that these two sets of correspondences were also in complementary distribution and that both Sanskrit and Greek had undergone conditioned sound changes. Note the following forms in these two languages:

| Greek | | Sanskrit | |
|------------|---------------|-----------|---------------|
| do:so: | 'I will give' | a-da:t | 'he/she gave' |
| di-do:mi: | 'I give' | da-da:mi | 'I give' |
| the:so: | 'I will put' | a-dha:t | 'he/she put' |
| ti-the:mi: | 'I put' | da-dha:mi | 'I put' |

The first pairs of forms in these two languages indicate that there is a regular morphological process of partial reduplication involving the initial syllable of the verb. This process derives the present stem of the root of these verbs, which are seen more clearly in the Greek future and Sanskrit past tenses. When a syllable containing an initial aspirated stop is reduplicated, the reduplicated syllable contains an unaspirated stop. In chapter 2, this kind of change is described as dissimilation at a distance.

Grassmann related this kind of morphological alternation in these two languages to the unpredictable correspondence between Germanic voiced stops and Sanskrit and Greek unaspirated stops, as illustrated by the following example:

| Sanskrit | Greek | Gothic | |
|----------|--------|--------|-------|
| bo:dha | pewtho | bewda | 'bid' |

According to Grimm's earlier generalization about sound correspondences, where Germanic languages such as Gothic have /b/ we would have expected to find /bh/ in Sanskrit and /ph/ in Greek. Grassmann concluded that Sanskrit and Greek did in fact have these forms originally in words such as these but that the aspiration was subsequently lost under the influence of the aspiration of the stop in the following syllable. So, an earlier (and unrecorded) form of Sanskrit, for example, would have had /*bho:dha/, which would have corresponded regularly with Gothic /bewda/. However, with two adjacent syllables in Sanskrit containing aspirated stops, the first of these then lost its aspiration to become a plain stop. A parallel change was also suggested for Greek to explain the once apparently irregular correspondence for this language.

In 1875, Carl Verner was able to dispose of yet another set of apparently irregular forms according to Grimm's statement of sound correspondences in the Indo-European languages. If you compare Latin /pater/ with Gothic /fadar/, both meaning 'father', you will see that there is a correspondence here between Latin /t/ and Germanic /d/. However, you will remember from the statement of the correspondences that Grimm noted earlier that where Latin has /t/, we would normally have expected Germanic languages to have /θ/. Verner collected a full set of such irregular forms and showed that the correspondences of $t : d$ and $t : \theta$ were in complementary distribution, with one correspondence showing up when the following vowel was stressed in Proto-Indo-European, and the other correspondence showing up when the vowel was unstressed.

Grimm had stated earlier that "the sound shifts succeed in the main but work out completely only in individual words, while others remain unchanged." He stated this because of the large number of forms that did not fit his generalizations. However, with the discoveries of Grassmann, Verner, and others, most of these irregularities were eventually eliminated. Toward the end of the nineteenth century, scholars such as Brugmann and Leskien were stating that "sound laws operate without exception."

The sound correspondences that Grimm, Verner, and Grassmann had noted were restated as “laws” to emphasize the fact that they could not be “broken.” Newtonian physics gave Brugmann and Leskien a model of a closed system in which there could be no exceptions, just like the laws of gravity. Darwinian biology offered them a model of organisms developing according to unbendable laws of nature (i.e., the survival of the fittest). This was the birth of the Neogrammarian school, often also referred to as the *Junggrammatiker*.

The Neogrammarians argued that these phonetic laws operated without exception in a language, and they argued further that the only conditioning factors that could determine the course of a sound change were phonetic factors. They claimed that it was impossible for semantic or grammatical factors to be involved in the conditioning of sound changes. Thus, for example, it would be impossible for a particular change to affect all words referring to trees, but not words referring to birds as well, and it would be impossible for a change to operate in nouns without affecting verbs at the same time. The only factors that could condition a sound change were phonetic factors such as the nature of the preceding and following sounds, the position of the sound in the word, and so on.

This was a significant innovation in thinking for historical linguists. Once it was acknowledged that sound change was a regular process which operated without exceptions, it became possible for the study of etymology, or the study of the history of words (and therefore also of languages) to become scientific (i.e., rigorous and open to proof). Scholars now had a way of arguing scientifically against proposals such as those of Guichard, who tried to relate all languages to Hebrew, as you saw earlier in this chapter. A sound correspondence or a similarity between two languages is of no value for reconstruction or for determining linguistic relationships unless it is “systematic,” or regular.

In reconstructing the history of languages, you therefore need to make the important distinction between a systematic (or regular) sound correspondence and an isolated (or sporadic) correspondence. This is a distinction that I did not make in chapter 5 when I was talking about the comparative method, but it is very important. Between steps 2 and 3 of the comparative method as I summarized it at the end of chapter 5, therefore, we need to add a further step, which says the following:

Separate those correspondences that are systematic from those that are isolated (i.e., which occur in only one or two words) and set aside the isolated correspondences.

Let us look at an example of what I mean by this. In addition to the forms that I gave in chapter 5 for Tongan, Samoan, Rarotongan, and Hawaiian, let us also add the following cognate forms:

| | | | |
|--------|--------|------------|--------------|
| Tongan | Samoan | Rarotongan | Hawaiian |
| fonua | fanua | ?enua | honua 'land' |

If we were to set out the sound correspondences that are involved in that cognate set, we would have an initial correspondence of $f : f : ? : h$, followed by a correspondence of $o : a : e : o$, then $n : n : n : n$, then $u : u : u : u$, and finally $a : a : a : a$. There is nothing new in the correspondences involving the initial consonants, nor the final segments $/-nua/$, but correspondence involving the vowels of the first syllable is different from any other correspondence that you saw in chapter 5.

According to what I said in chapter 5, you should assume that each set of correspondences that is not in complementary distribution with any other correspondence should be reconstructed as going back to a separate original phoneme. If we were to reconstruct this new correspondence as going back to a separate protophoneme, however, you would end up reconstructing a new phoneme which occurs in just this single word. Rather than complicate the statement of the phonemes of the original language, what you do is simply ignore such isolated correspondences and reconstruct only on the basis of the evidence provided by systematic sound correspondences. You should therefore reconstruct the word for 'land' on the basis of regular correspondences only. There is not enough data in these four languages to allow you to decide whether the original vowel was $/*e/$, $/*o/$, or $/*a/$. The occurrence of reflexes of $*o$ in both Tongan and Hawaiian might suggest that $/*fonua/$ was the original form, with Samoan having undergone a sporadic shift of the vowel to $/a/$, and Rarotongan having unpredictably shifted the vowel to $/e/$. Comparing these languages with non-Polynesian languages which also have cognates of this word, such as Fijian $/vanua/$, we might be tempted to reconstruct Proto-Polynesian as having had $/*fanua/$ instead. But whatever the reconstruction, we are simply going to have to accept that there have been some completely unpredictable changes in the vowels of some of these languages. Sometimes we can resolve these problems by identifying the irregularities as loans, and sometimes we can isolate a previously identified environment for sound change, but at other times the cause of the irregularity cannot be identified.

Another example to illustrate the same kind of problem involves the additional following cognate set:

| | | | |
|--------|--------|------------|-------------|
| Tongan | Samoan | Rarotongan | Hawaiian |
| paʔi | paʔi | paki | paʔi 'slap' |

In this case, the medial correspondence of $aa : a : a : a$ is not attested outside this cognate set, and the same is true of the correspondence of $? : ? : k : ?$. The Samoan, Rarotongan, and Hawaiian data are perfectly consistent with what you saw in chapter 5, pointing to the original form having been $/*paki/$.

If the Tongan form were to behave as predicted, it should have been /paki/, but instead we find /paaʔi/. We must note that there has been an unpredictable change in Tongan of /*a/ to /aa/ and another unpredictable change of /*k/ to /ʔ/.

According to the NEOGRAMMARIAN HYPOTHESIS that sound change is without exception, there must be some kind of explanation for irregularities such as this. What Neogrammarians said was that, instead of being irregular, such correspondences must involve some other factors. It could simply be a matter of “undiscovered regularity”—in fact, there may be a regular phonetic conditioning factor that nobody has yet been clever enough to uncover. In this case, the explanation is perhaps that the Tongan form /paaʔi/ has been incorrectly identified as cognate with the forms in the other languages. Despite the similarity in the phonological shape and the meaning, it could be that this word is derived from the quite separate (and not cognate) root /paa/ and that the final syllable is a suffix /-ʔi/, which is added to many transitive verbs in Tongan.

The Neogrammarians did find some ways of accounting for some irregular sound correspondences as well, and it is to these that I turn my attention in the following sections.⁴

9.2 CONVERGENT LEXICAL DEVELOPMENT

When words undergo convergent development, you will also find that sounds do not have reflexes that you would have predicted from the earlier forms. What happens when two words converge in this way is that words which are largely similar in form (but not identical) and which have closely related meanings may end up combining their shapes and their meanings to produce a single word that incorporates features of the two original words. If somebody combines the words *dough* and *cash* into the previously nonexistent word *dosh*, you can say that in the speech of this person there has been convergent development of these two lexical items. Another example of this kind of change is in Bislama (in Vanuatu) where the English words *rough* and *rob* (*him*) end up as /ravem/, and not /rafem/ and /robem/ as we might have expected. The mixed word /ravem/ covers a wide range of meanings derived from the meanings of the two original words: that is, ‘rob, be rough to, do in a rough way, cheat, exploit’.

A similar development can be found when one language copies words from another language. What generally happens is that a language copies a single word from another language. However, there are cases when words in two different languages, which are partly similar in form and which are either the same or similar in meaning, are copied at the same time into a third language. When such words are copied, they may take on a form and a meaning that have elements from both of the source languages. For instance, in New Zealand the English word *kit* (which also occurs in the compound

kit-bag) seems to have taken on the meaning of the formally similar Māori word *kete* 'basket', and now Pākehā New Zealanders refer to traditional Māori baskets in English also as *kits*.

9.3 NONPHONETIC CONDITIONING

Another criticism that has been made of the Neogrammarian Hypothesis in more recent decades relates to the structuralist belief in strict separation of levels. Structuralist linguists in the 1930s to the 1950s held that, when we analyze the phonological system of a language, the only facts that we should concern ourselves with are purely phonological facts. Consideration from other levels of language such as morphology, syntax, and semantics should be carefully excluded when we come to working out the phonemes of a language. This view of phonology in which there is a strict separation of levels in linguistic analysis is often referred to as AUTONOMOUS PHONEMICS because phonemics is supposed to be completely autonomous, or independent of all kinds of facts except facts from the same "level" of analysis. In insisting on this rigid dichotomy between different levels of analysis, the structuralists were little different from the Neogrammarians, who also insisted that only phonetic conditioning factors could be involved in the statement of sound changes.

In more recent years, some linguists have questioned, and even denied, the need for the strict separation of levels that earlier linguists insisted on. If we allow reference to grammatical facts, for instance, we are able to state the distribution of the allophones of phonemes in a much more straightforward manner, as this allows us to use terms like "morpheme boundary" and "word boundary." As these are grammatical rather than phonetic concepts, structuralist phonemicists were of course unable to use terms such as these.

Although modern linguistics has now developed far beyond these methods and beliefs, it is still often argued that phonological changes over time should be stated only in terms of purely phonological conditioning factors and that sound changes are never conditioned by grammatical or semantic factors. It is indeed difficult to imagine a sound change that operates in a language only in words referring to the names of trees, or which only applies to verbs involving motion away from the speaker, so we probably can say that sound changes cannot be conditioned by semantic features. However, it seems that some languages do, in fact, provide evidence that at least some sound changes apply only in certain WORD CLASSES (or parts of speech) and not in others. Such a sound change clearly involves grammatical rather than purely phonological conditioning.

Paamese is an example of a language that has undergone a grammatically conditioned sound change. There is a correspondence of southern Paamese // to northern Paamese /i/, /l/, or zero. The southern varieties directly reflect the

original forms in Proto-Paamese with respect to this particular feature, with the northern varieties having undergone the following fairly complex set of conditioned changes:

$$*l > \begin{cases} \emptyset / \begin{cases} \# _ \text{non-high V} \\ \text{non-high V} _ e \\ e _ \text{non-high V} \end{cases} \\ l / \begin{cases} _ \text{high V} \\ \text{high V} _ \end{cases} \\ i / \text{elsewhere} \end{cases}$$

This rule states the following:

1. The lateral */*l/* is lost word-initially before the non-high vowels */*e/*, */*a/*, and */*o/*, and word-medially between */*e/* and any of these non-high vowels, for example:

| Northern Paamese | | |
|------------------|--------|------------------|
| <i>*leiai</i> | > eiai | 'bush' |
| <i>*alete</i> | > aet | 'flat area' |
| <i>*gela</i> | > kea | 'he/she crawled' |
| <i>*melau</i> | > meau | 'megapode' |

2. The lateral was retained unchanged when it was preceded or followed by a high vowel (i.e., */*i/* or */*u/*) in any position of the word, for example:

| Northern Paamese | | |
|------------------|----------|---------------|
| <i>*asilati</i> | > asilat | 'worm' |
| <i>*haulue</i> | > houlu | 'many' |
| <i>*gilela</i> | > kilea | 'he/she knew' |
| <i>*teilaji</i> | > teilaj | 'sky' |
| <i>*ahilu</i> | > ahil | 'hair' |
| <i>*tahule</i> | > tahul | 'wave' |

3. In all other situations, */*l/* changed to */i/*, for example:

| Northern Paamese | | |
|------------------|-----------|----------------|
| <i>*la:la</i> | > a:ia | 'kind of bird' |
| <i>*malou</i> | > maiou | 'kava' |
| <i>*meta:lo</i> | > meta:io | 'European' |

| | | | |
|---------|---|--------|------------------|
| *to:lau | > | to:iau | 'northeast wind' |
| *amalo | > | amai | 'reef' |
| *avolo | > | avoi | 'mushroom' |

The interesting point is that none of the examples of word-initial changes to /**l*/ that I have just given involves a verb. Verbs, it seems, are completely immune in Paamese to any changes involving initial /**l*/, though the same sound changes according to the regular rules in verbs in any other position in the word (as the preceding examples also show). Just so you can see that word-initial laterals in verbs are retained intact, examine the following changes:

Northern Paamese

| | | | |
|---------|---|-------|--------------------|
| *leheie | > | lehei | 'he/she pulled it' |
| *loho | > | loh | 'he/she ran' |
| *la:po | > | la:po | 'he/she fell' |

If these forms had obeyed the rule that I just presented, we would have predicted /*ehel*/, /*oh*/, and /*a:po*/, respectively. This is therefore a clear example of a sound change that does not involve purely phonological conditioning factors but also involves grammatical conditioning.⁵

9.4 THE WAVE MODEL AND LEXICAL DIFFUSION

The Neogrammarian Hypothesis on which the comparative method rests has never been free from attack. Even when it was being formulated in its most rigid form in the 1870s by Brugmann and Leskien, there were people who claimed that their position was overstated. One of the points on which the Neogrammarians were criticized related to their view of how languages diverge. In chapter 6, I discussed the notion of subgroups of languages within larger families of related languages. This model of language change suggests that languages undergo sudden splits into two (or more) quite different daughter languages and, once these splits have taken place, there is no longer any contact between the new languages. Each new language, it is assumed, then continues completely on its own, undergoing its own completely individual sets of changes.

However, many scholars have pointed out that this representation misses important facts about the nature of language change. Languages seldom split suddenly. Generally, what happens is that a language develops two (or more) closely related DIALECTS, which only gradually diverge into separate languages. While these languages are slowly becoming more and more different,

there is usually some degree of contact between the two speech communities, often with some kind of mutual influence between the two dialects. Even when the two dialects finally end up as distinct languages (i.e., when speakers have to learn the other speech variety as a separate system in order to be able to understand it), there can be mutual influence.

The Neogrammarian model also does not model diversity within languages well. As we all know, languages are heterogeneous, and there are often no distinct boundaries between languages or dialects at all. A detailed study of any language area (even very small ones) will generally reveal the existence of a number of dialects, or local varieties of the language. (Variation by place is not the only type of variation in language, either.) However, the dialect boundaries are also often indistinct, and it is often impossible to say where one dialect begins and the other ends.

I will now look at a particular example to show you what I mean. On the island of Paama in Vanuatu, the people speak a single language, the Paamese language, of which there are about 4,000 speakers. The island itself is quite small, being only about 10 kilometers from north to south and 4 kilometers from east to west. There are 20 villages on the island. Even within this speech community, which is tiny by world standards, there is dialect variation. Speakers of the language themselves recognize two dialects, a northern and a southern variety. These two dialects differ in the following respects:

1. Sequences of /ei/ and /ou/ in the north correspond to /ai/ and /au/, respectively, in the south; for example:

| Northern Paamese | Southern Paamese | |
|------------------|------------------|----------|
| eim | aim | 'house' |
| keil | kail | 'they' |
| oul | aul | 'maggot' |
| moul | maul | 'alive' |

2. The south often has /l/ where the north has /i/ or zero (as determined by the rule that I presented earlier); for example:

| Northern Paamese | Southern Paamese | |
|------------------|------------------|-----------------|
| amai | amal | 'reef' |
| aii | a:l | 'stinging tree' |
| tahe | tahel | 'wave' |
| mea | mela | 'get up' |

3. The south has initial /g/ and /d/ where the north has initial /k/ and /t/; for example:

| Northern Paamese | Southern Paamese | |
|------------------|------------------|---------------------|
| raho | daho | 'he/she is fat' |
| rei | dai | 'he/she chopped it' |
| kea | gela | 'he/she crawled' |
| keih | gaih | 'he/she is strong' |

4. The north often has /a/ when the following syllable contains an /a/, whereas the south has /e/ in the first syllable and /a/ in the second syllable; for example:

| Northern Paamese | Southern Paamese | |
|------------------|------------------|-----------|
| atau | letau | 'woman' |
| namatil | nematil | 'I slept' |

5. The south has /m/ and /v/ when the north has the labiovelars /m^w/ and /v^w/; for example:

| Northern Paamese | Southern Paamese | |
|----------------------|------------------|---------------------|
| m ^w ail | mail | 'left-hand side' |
| m ^w eatin | meatin | 'man' |
| v ^w e:k | ve:k | 'my sleeping place' |
| v ^w akora | vakora | 'coconut shell' |

In addition to these phonological differences between the two dialects, Paamese speakers are also able to point to numerous lexical and morphological differences between the northern and southern varieties of the language (though I do not give examples of these, as they are irrelevant to the point I want to talk about).

However, the picture is not nearly as simple as this. While the extreme north and the extreme south of this small island do differ in the ways that I have shown, it is in fact impossible to draw a single line that marks the boundary between the two dialects. To continue the discussion, I need to introduce the term "isogloss". An ISOGLOSS is a line that is drawn on a map that marks two areas that differ in one particular linguistic feature. On the map of Paama in figure 9.1, each dot represents a single village. It is possible to draw isoglosses for each of these linguistic features. You will find that, while the northern and southern ends of the island have the features that I have indicated, the villages in the center of the island share features from both the north and the south. So, for example, the isogloss dividing the features listed under 1, 2, and 5 above and the isogloss dividing the features listed under 2 and 4 are located as shown in the figure. There is therefore clearly no single boundary that can be drawn between the northern and southern dialects of Paamese, as the isoglosses do not run together. This has been a

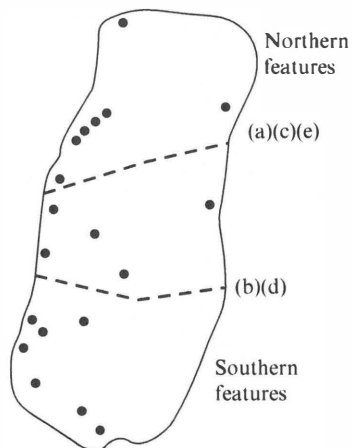


FIGURE 9.1 Isoglosses in Paama

simple example because the island is so small and the number of linguistic features that I have given to illustrate the two dialects is also fairly small.

In a larger language, the situation can become much more complicated. In a language such as German, for example, a huge number of isoglosses crisscross the German-speaking area. While many of these do bunch together (to form an ISOGLOSS BUNDLE), many other isoglosses cross the bundle, and some individual isoglosses move away from the bundle in a direction all of their own, perhaps to rejoin the bundle at a later point or perhaps to end up in a completely different part of the German-speaking area. The map in figure 9.2 shows the Rhenish fan of isoglosses in the Dutch-German speaking

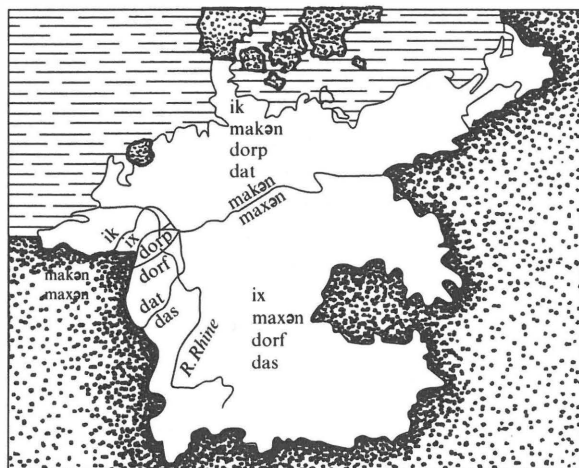


FIGURE 9.2 German dialect differences

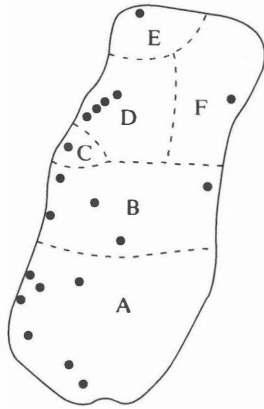


FIGURE 9.3 Further isoglosses in Paama

area, which divides areas with fricative and stop pronunciations in words like *machen* 'make', *ich* 'I', *Dorf* 'village', and *das* 'the'.

Returning to the relatively simply example of Paamese, it turns out that even this discussion has been oversimplified and the real situation is more complicated. Even though I have set out a number of phonological correspondences between northern and southern Paamese, some words behave individually, depending on whether they follow the stated correspondence. For instance, the correspondences between southern bilabial consonants and the northern labiovelar consonants (represented by /m^w/ and /v^w/) are grossly oversimplified. The reality of the situation is better shown by breaking these larger areas into much smaller areas, as set out in figure 9.3. These areas are characterized by the following facts:

Area A: There are no words containing labiovelar sounds, and all words contain plain labials.

Area B: There are some words containing /m^w/ but none with /v^w/.

Only a few words are consistently pronounced with the labiovelar nasal, including /m^weatin/ 'man' and /m^weahos/ 'male'.

Area C: There are some words containing /m^w/ and a few words with /v^w/. These words include those listed for Area B and also /am^we/ 'married man', /ti:mwe/ 'friend', and /v^we:k/ 'my sleeping place'.

Area D: There are some more words with /m^w/ and several more with /v^w/, including /m^weas/ 'dust', /rom^weite/ 'top', /um^we:n/ 'work', /v^weave/ 'cottonwood', and /v^waila/ 'footprints'.

Area E: More words contain each of these two sounds rather than plain labials: /m^wail/ 'left-hand side', /v^walia/ 'spider', /v^weihat/ 'coastal rocks', and /v^waiteh/ 'door'.

Area F: Yet more words contain labiovelars rather than plain labials: /m^wai/ 'he straightened it', /v^wakora/ 'coconut shell', and /av^we/ 'bell'.

The simple isoglosses that I drew earlier to separate the areas that have labiovelars from the areas that do not are a severe oversimplification.

You can see that the labiovelars are more prevalent in Area F and decreasingly prevalent until we get to Area A, where there are no labiovelars at all. Which words will have labiovelars in any particular area seems to be quite unpredictable. Each word, in fact, seems to have its own behavior. If the comparative method were strictly applied to these data, the facts that I have just described would need to be represented by recognizing six "dialects" in Paamese, with the following lexical correspondences between them:

| A | B | C | D | E | F |
|--------|----------------------|----------------------|----------------------|----------------------|----------------------|
| meatin | m ^w eatin | m ^w eatin | m ^w eatin | m ^w eatin | m ^w eatin |
| ame | ame | am ^w e | am ^w e | am ^w e | am ^w e |
| meas | meas | meas | m ^w eas | m ^w eas | m ^w eas |
| mail | mail | mail | mail | m ^w ail | m ^w ail |
| mai | mai | mai | mai | mai | m ^w ai |

On the basis of the earlier statement that there was a northern dialect with labiovelars corresponding to a southern dialect with plain labials contrasting with correspondences between both dialects involving plain labials, we would probably want to reconstruct for Proto-Paamese a contrast between labiovelars and plain labials. However, if we were strictly to apply the comparative method as I described it in chapter 5 to the data that I have just set out, we would be forced to reconstruct six separate nasal protophonemes, as there are six different sets of correspondences involving the nasals /m/ and /m^w/.

This brings us to the point where I should mention the French dialectologist Gilliéron. A DIALECTOLOGIST is a linguist whose speciality is the distribution of dialect features in a language. Gilliéron was a nineteenth-century scholar who opposed the view of the Neogrammarians, who were his contemporaries, when he made the famous statement that "every word has its own history." What he meant was that sound changes are not rigidly determined by purely phonetic factors, as the Neogrammarians had so forthrightly stated. Instead, he said that only some words undergo a particular change, while others do not. Which words undergo a particular change can, in fact, be quite arbitrary, as you have just seen with the Paamese example. Gilliéron's view is totally incompatible with a strict application of the comparative method.

Gilliéron's view of linguistic change is consistent with what is referred to today as the WAVE MODEL, and it contrasts sharply with the FAMILY-TREE MODEL of change on which the comparative method rests. The wave model implies that, instead of sharp linguistic splits, changes take place like waves spreading outward from the place where a stone is dropped into water, traveling different distances with different stones and crossing with waves caused by other stones (figure 9.4).

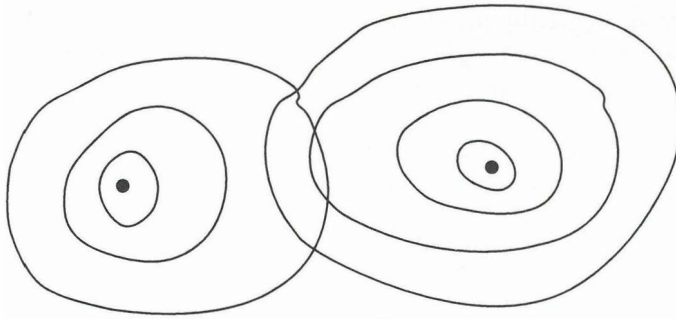


FIGURE 9.4 Wave model of change

Despite the success of the comparative method in reconstructing a large number of different protolanguages, the wave model of linguistic change has gained respectability in modern linguistics through recent work on lexical diffusion. This refers to the fact that sound changes do not operate simultaneously on every word in a language meeting the conditions for the application of a particular change. For example, if a language undergoes the devoicing of word-final voiced stops, what will often happen is that final voiced stops in just some words will lose their voicing first, and this change will then gradually spread throughout the lexicon to other words that are of basically the same phonological shape. That is exactly what seems to be happening in Paamese. The original distinction between /m^w/ and /m/ is being lost, with /m/ coming to replace the labiovelar in the south. However, the change is only gradually moving through the lexicon, having affected all words in the far south and just some words in villages further north. Over time, we can predict that increasing numbers of words in the central villages will undergo this change such that eventually the dialects of these villages will resemble those of the far south.

9.5 DIALECT CHAINS AND NONDISCRETE SUBGROUPS

In the preceding section I indicated that dialects cannot usually be separated by single lines of a map and that what you will find, instead, is that different linguistic features need to be mapped individually by means of isoglosses. While isoglosses do tend to bunch together in bundles, individual isoglosses frequently stray, making it impossible in many situations to draw a family-tree diagram showing dialect relationships.

In situations where isoglosses do not bundle together closely, a different kind of case can arise, which again demonstrates a fundamental weakness of the comparative method. With dialect differences such as these, it is possible

for there to be no clearly recognizable boundaries at all between one dialect and another, with dialects only gradually merging into each other.

You will note in the map of isoglosses (figure 9.2) that the entire German and Dutch language areas were included on a single map. The reason for this is that it is not possible to draw a single line on a map that separates the two languages. The Dutch-German political border represents a language boundary only in the sense that people on each side of this line have mutually unintelligible standard varieties. However, the local dialects of Dutch and German that are spoken on either side of the political border are little different, and people can readily understand each other.

What I am talking about in the case of Dutch and German is a dialect chain situation. Here, immediately neighboring dialects exhibit only slight differences from each other, but as geographical distance between dialects increases, so, too, does the extent of difference between dialects. Eventually, the point will be reached in a dialect chain where two different varieties will be mutually unintelligible, even though all of the neighboring dialects in between are mutually intelligible.

Even the languages spoken by relatively few people in Aboriginal Australia and in Melanesia commonly exhibit dialect-chain features. There is an area on the border between Queensland and New South Wales where cognate counts in the basic vocabulary of a number of neighboring speech communities are relatively high and where the two varieties are mutually intelligible. However, when we compare the basic vocabularies of the speech communities at the extreme ends of this chain, the cognate percentage drops to a level at which mutual intelligibility is not conceivable (figure 9.5). All of these speech communities are sharply differentiated from languages spoken outside the clearly definable area that is marked on the map, and cognate sharing between areas on either side of this boundary is low. Because there is mutual intelligibility between neighboring speech communities within this bloc, as well as a sharp contrast with speech communities that clearly do not belong to the bloc, some linguists have proposed the term *FAMILY-LIKE LANGUAGE* to refer to such situations.

The same principle that is involved in the phenomenon of dialect chains can extend to more distant levels of relationship as well. A comparison of the languages of central and northern Vanuatu has revealed that sometimes a particular language, or a number of languages, may satisfy the criteria for membership in more than one subgroup at a time. That is, we can have dialect chains and perhaps even language chains as well. The lines around the areas in the map of part of Vanuatu (figure 9.6) indicate which languages appear to belong together in lexicostatistically determined subgroups, and you will see that some of the areas overlap. This means that the languages in those areas appear to belong to two different subgroups at once.

Because the comparative method and the family tree model were so important in establishing historical linguistics in Europe, it has sometimes been claimed that those who wish to use the comparative method elsewhere

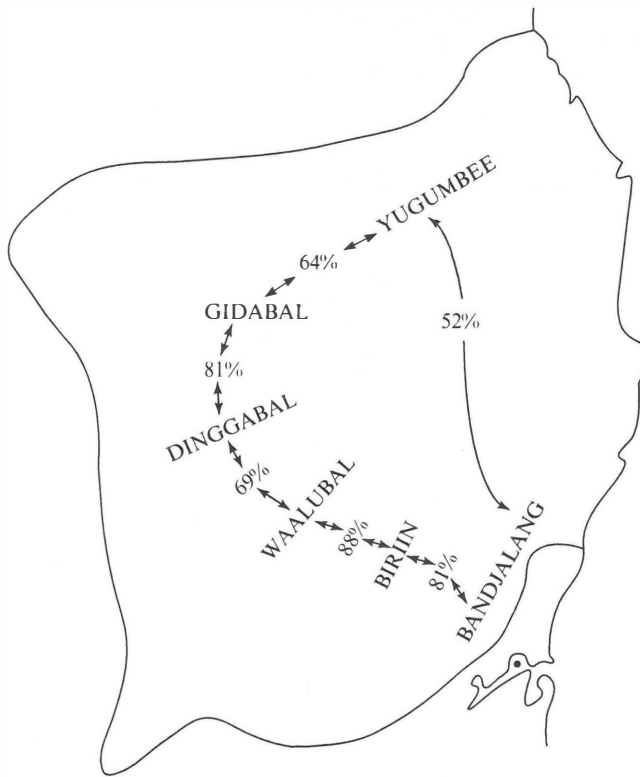


FIGURE 9.5 Bandjalang varieties in Australia showing lack of mutual intelligibility

are imposing a Eurocentric view on the rest of the world. This is simply not true (and it should be pointed out that the wave model also originated with European languages). To see this, we need to think about why the comparative method works. It works because regular systematic correspondences are created when a subset of a speech population converges on a particular change. That change might be a slight fronting of velars before high vowels, it might be the innovation of a particular morpheme, or it might be the generalization of the particular syntactic pattern to an environment it did not previously occur in. This process happens in all cultures, all over the world. Small hunter-gatherer societies, even ones with egalitarian social systems, still use language as a mark of social identity. This is not confined to large sedentary agricultural European societies. From the methodological point of view, the main difference between the major languages of Europe and the endangered languages of Oceania is that the European languages have had a few hundred years of intensive scholarship, whereas in many other parts of the world the languages have been described only relatively recently, the documentation has been done by nonnative speakers, and there have been fewer scholars working on the problems. There are so many interesting unsolved problems and unresearched

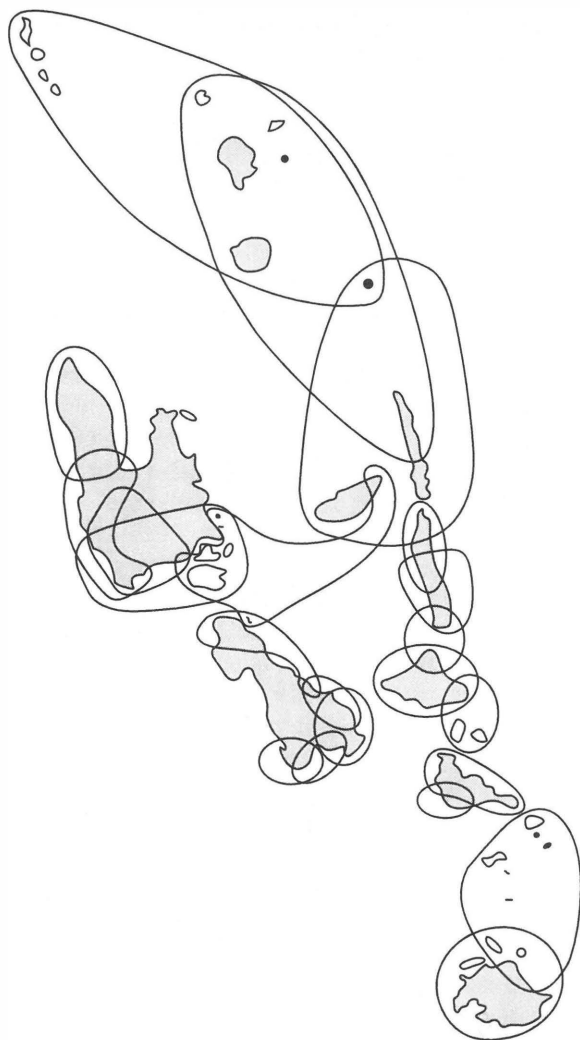


FIGURE 9.6 Language subgroups in part of Vanuatu

areas in historical linguistics that we cannot resort to simple statements like “our methods don’t work because they were developed in Europe.”

Reading Guide Questions

1. What is the basic difference between the study of etymology before the Neogrammarians and in the present day?
2. What was the importance of Sir William Jones’s statement in 1786 for the study of the history of languages?

3. What important contribution did Jakob Grimm make to the study of the history of languages?
4. What was the importance of Verner's and Grassmann's discoveries in the history of the Germanic languages?
5. What was the Neogrammarian Hypothesis? How did the neogrammarian view of language change differ from that proposed by Grimm?
6. How does the existence of sporadic sound correspondences affect the way that we apply the comparative method?
7. How does the wave model of linguistic change differ from the family-tree model?
8. What is lexical diffusion, and how does this affect the application of the comparative method?
9. What is an isogloss? What is significant about the fact that isoglosses do not always coincide (and sometimes cross over each other)?
10. What is autonomous phonemics, and what impact does the acceptance of this point of view have on the way that linguists view language change?
11. What is a dialect chain?
12. What is meant by "nondiscrete" subgroups, and why is this a problem for the application of the comparative method?

Exercises

1. Examine the data in Dataset 14. Identify the sound changes. Do you encounter any problems in working out what the relationship between Old Arabic and Cypriot Arabic is?

Further Reading

- Leonard Bloomfield, *Language*, chapter 23, "Analogic Change," pp. 404–24.
- Theodora Bynon, *Historical Linguistics*, chapter 4, "The Neogrammarian Postulates and Dialect Geography," pp. 173–97.
- Hans Henrich Hock, *Principles of Historical Linguistics*, chapter 15, "Linguistic Contact: Dialectology," pp. 426–71.
- Lyle Campbell and William Poser, *Language Classification*, chapter 11 "Why and how do languages diversify and spread?" pp. 330–64.
- William Labov, *Principles of Linguistic Change*, chapter 10, "Social Networks," pp. 325–365; chapter 13, "Transmission," pp. 415–445.
- William Labov, "Transmission and Diffusion".
- Betty S. Phillips, *Word Frequency and Lexical Diffusion*, chapter 1, "Word Frequency and the Neogrammarian Controversy," pp. 1–30.
- David Britain, "Space and Spatial Diffusion."

Morphological Change

So far in this book, I have been talking almost entirely about questions to do with sound change. There is much more to language than sounds, however. We also have to consider the grammar of a language—that is, the ways in which units of meaning are put together to make up larger units of meaning. Grammar is traditionally divided into MORPHOLOGY (the ways in which words are made up of smaller grammatical elements, or MORPHEMES) and SYNTAX (the way that words are combined with other words to form larger elements, or SENTENCES). The grammatical rules of a language are what link sounds to meanings. In talking about a language, we must also talk about the kinds of meanings that are expressed: that is, the SEMANTIC SYSTEM. Just as languages change in their sound systems, they can also change in their grammatical systems and in the meanings of their words. It is the purpose of the next few chapters to introduce the kinds of changes that take place in morphology, syntax, and semantics. This chapter is about morphological change. Chapter 11 covers change in semantics and the lexicon, while chapter 12 covers syntactic change.

I have concentrated so far on the study of sound change, with comparatively little emphasis on morphology, syntax, and semantics. This is no accident. The study of sound change has a long history, going back over 150 years. Scholars have therefore had lots of time to gather all kinds of information on sound change. Not only this, but it is probably inherently easier to study the changes to the sound system of a language than it is to study its grammatical and semantic systems. The number of individual phonemes of

a language ranges from around a dozen or so in some languages, to 140 or so at the very most in other languages. The range of possible variations and changes in phonology is therefore more restricted than in the grammatical system of a language, where there may be dozens (or even hundreds) of grammatical categories; not only that, we also have to consider the existence of thousands of particular grammatical constructions for any language. Also, when considering the semantic system of a language, the number of semantic relations that hold between different items in the lexicon would be so huge that they would be almost uncountable. So it is not really surprising that we know more about phonological change than we know about other types of changes.

10.1 CHANGES IN MORPHOLOGICAL STRUCTURE

As you know, morphemes are pieces of words.¹ While a great deal of change operates at the level of a word (or, in the case of syntactic change, at higher-order levels), we also find changes occurring within the word, at the level of morphemes. Just as we divided changes in some systems into a number of different types, so too we can identify different types of morphological changes. Some of the most common are given in this section.² Having a good idea of the range of morphological changes in the world's languages will help in reconstruction.

10.1.1 Allomorphic Change

One case when we get change in morphemes is when there is a sound change in the language, and that effects sounds in some environments but not others. That can lead to allomorphy.

In the history of the Djambarrpuyŋu language of northern Australia, there was a sound change that deleted the final vowels from three-syllable words. This affected two-syllable words that had a case marker on the end of them, as well as three-syllable words without a case marker. The sound change did not happen if it would have produced a word with a consonant cluster. There was also a change of lenition of voiced intervocalic stops. For example, the dative case marker was **-ku* in Proto-Yolŋu, but in Djambarrpuyŋu there are three allomorphs: *-w* after vowels (e.g., *yapaw* 'for sister'), *-gu* after nasals (*Wamuttjangu* 'for Wamuttjan'), and *-ku* after stops (*Wamutku* 'for Wamut').

Sound changes can also be morphologically conditioned. The Djambarrpuyŋu example I just talked about is part of a general sound change in the language. In other cases, we find the change only at morpheme boundaries, and not as a general process in the language. One example is in the history of the Arandic (Pama-Nyungan) language Kaytetye. In this language, there are

two allomorphs of the present imperfective suffix, *-rranytye* and *-ranytye*. The *rr* is a trill, while the *r* is a glide. In the history of the language, at some point there was an apical dissimilation change. That is, a trill changed to a glide whenever it followed an apical stop or trill. Thus the present imperfective of the root *ange-* 'scoop' is *ange-rranytye*, but the imperfective of *ate-* 'press with foot' is *ate-ranytye*.

Allomorphs can be lost through sound change, as well as created by it. English used to have lots of different ways of forming plurals. However, many of them were lost with the sound change that affected the final vowels of words in the late Old English and early Middle English period.

10.1.2 Changes in Conditioning

You have already seen an example of change in conditioning of allomorphs in the English plurals that we talked about in sec. 4.3. Remember that the stem changes in *foot* : *feet* and *mouse* : *mice* were originally allophonic vowel fronting caused by a final *i* vowel. The vowel was later lost, and further sound changes took place. This is a change in conditioning: the former conditioning environment has been lost.

10.1.3 Boundary Shifts

You've also seen examples of shifts in the history of morpheme boundaries, although we didn't talk about them as such. In sec. 7.1 we talked about the Samoan transitive marker and how the original final consonant of the stem has been reanalyzed as part of the suffix. This is an example of boundary shift.

A further example comes from the history of French (although this sort of change is fairly common in languages). In many languages, third-person verb forms are unmarked. One way that they become unmarked is when a former marker of third person is reanalyzed as being part of the stem or as marking something else. Then can lead to reshaping of the whole paradigm. Consider the following forms of the Provençal French verb 'sing':

| | Earlier | Later |
|-----------|-----------|------------|
| 1st sing. | canté-i | cantét-e |
| 2nd sing. | canté-st | cantét-es |
| 3rd sing. | canté-t | cantét-ø |
| 1st pl. | canté-m | cantét-em |
| 2nd pl. | canté-tz | cantét-etz |
| 3rd pl. | canté-ran | cantét-on |

Morpheme boundaries are also sometimes created, particularly when words are borrowed into another language. When the word *stool* was borrowed into Ndebele, it received a noun-class prefix. However, the initial *s* of the word was interpreted as being part of the prefix, not of the root. Thus the plural of the word *istulu* 'chair' is *iztulu*, and it conforms to one of the common noun classes. Another example of reanalysis and creation of morpheme boundaries comes in the morpheme *-burger* that is creeping into the English language in words such as *hamburger*, *cheeseburger*, *eggburger*, *fishburger*, and now even *Kiwiburger* (which refers to a hamburger marketed in New Zealand by McDonalds, and which contains not kiwi meat but a fried egg and pickled beetroot). The word *hamburger* was originally the only one of these four words to be used in English. Its derivation was from the name of the city Hamburg, with the suffix *-er* (on the same pattern as the noun *Berliner* derived from *Berlin*). However, speakers of English perceived an ambiguity between this explanation of the word's origin and the interpretation of *hamburger* as 'ham' (because of the meat filling in the bun) plus 'burger'. The second analysis has won out, and *hamburger* appears now to be a compound.

A special type of boundary reanalysis goes by the name of BACK FORMATION. An example of this process is involved in the development of the English word *cherry*. This word was originally borrowed from the French word *cerise*. In its pronunciation in French, the word is identical in both the singular and the plural: /səʁiz/. When *cerise* was copied into English, people analyzed the word as being plural (as cherries are small fruit that are generally seen in large numbers anyway). The final /-z/ of the word in French was thought to be the plural suffix, so when English speakers wanted to speak of a single *cerise*, they simply dropped off this /-z/ and came up with the previously nonexistent word *cherry*. If those earlier English speakers had not reanalyzed this root, we would today be speaking of "one cherries" and "two cherrieses." (Of course, English copied the French word *cerise* again at a later point in its history to refer to a deep purplish color, which is pronounced in English exactly as we would predict on the basis of its pronunciation in French: /səʁi:z/.)

Finally, morpheme boundaries may also be deleted. This also happens particularly when words are borrowed from one language to another, although it also happens with the languages even when no borrowing is involved. The word *caveat* is a borrowing from Latin. The Latin word is a verb that means 'let him/her beware'. It has a third-singular ending and a subjunctive ending on the stem. However, the English word is morphologically simple (and inflects like a noun, not like a verb).

10.1.4 Doubling and Reinforcement

Sometimes a word which is already inflected for a particular category will receive REINFORCEMENT by the addition of another instance of the morpheme.

This appears to happen particularly when a form is suppletive or otherwise irregularly marked. One example is the form of the word 'to be' in Latin and its subsequent development in the Romance languages. In Latin, the word is *esse*, unlike most other infinitives in the language which ended in *-Vre*; compare *amare* 'to love', *dormire* 'to sleep', and so on. Late in the Latin period, *esse* comes to be inflected like other infinitives. The infinitive becomes *essere*, and this is the form which descends as *être* 'to be' in French and so on.

There are also examples of this in English. The word *children* actually contains a double plural. In old English, the plural of *child* was *childer*, but at some point another plural marker *-en* was added. (These days the *-en* marker is very restricted, but it does occur in words like *brethren* and *oxen*.)

10.1.5 Change in Order of Morphemes

We find across the languages of the world that there are recurring similarities in the order of morphemes. For example, in just about all the languages with both tense marking and subject-person agreement, the tense marking occurs closer to the root than does the subject marking. Derivational morphology commonly occurs closer to the root than inflectional morphology does. Clitics almost never occur inside a word (that is, with very few exceptions, a clitic cannot appear closer to a root than an affix). However, there are exceptions to all of these statements, although they are rare.

We sometimes find that changes in the order of morphemes within a word have occurred, although this change is also somewhat rare. When this occurs, the change just about always results in a more common ordering. For example, we might find that a word has grammaticalized as a derivational morpheme. This would potentially create a derivational morpheme outside inflection. We do find cases where the order of the derivational marker and inflectional marker have changed so that the derivational marker is closer to the root. We do not find any cases where the reverse has happened (although the reverse can be created through other processes).

Here is an example from the Chukchi language. In this language, both the future and the desiderative ("wanting") forms of words are made with the prefix *re-* and a suffix *-j*. However, they differ in the order they occur with other morphemes in the word. In the desiderative (which is derivational), the suffix occurs closer to the root than the aspect marker, whereas when the form is marking future tense, the suffix occurs on the other side of the aspect marker. In this case, we know that the desiderative form reflects the earlier order and the creation of the future tense from the desiderative also involved the swapping of places between the continuous aspect and the suffix portion of the tense.³

a. Desiderative

re- viri -ŋə -rk -ət
 DES- descend -DES -CONT -3pl
 'They want to descend.'

b. Future

re- viri -rkæani -ŋ -ət
 FUT- descend -CONT -FUT -3pl
 'They will descend.'

10.2 ANALOGY

A major type of change which affects morphology is known as ANALOGY; although analogy does not affect morphology only, it is one of the major types of morphological change, and so it will be discussed here in some detail. We will also discuss analogical changes in chapter 12.

The term ANALOGY is used in a nontechnical sense to mean that we find similarities between things that are not ordinarily regarded as similar. In presenting an argument, we often "draw an analogy" by taking a concept that we know our audience is familiar with and showing how it is similar to the new concept that we are talking about. For example, if you were trying to explain the unfamiliar concept of complementary distribution of the allophones of a phoneme to a beginning student of linguistics, you could use an analogy to help get your point across. You might say that complementary distribution can be compared with the relationship between formal and nonformal education. Formal education is carried out only in certain contexts and by certain people (i.e., by qualified teachers in approved schools). Nonformal education also takes place in particular sets of contexts, but different ones, and is generally carried out by different people as well (i.e., out of school; by our parents, community leaders, agricultural extension officers, village leaders in Pacific villages, and so on). Similarly, you could say that certain allophones of phonemes may occur only in certain phonetic contexts, and other allophones in other contexts. Although there is nothing else in common between phonemes and education, we can use the similarity that does exist to illustrate this concept.

Analogies can be represented by using a formula of the following type:

A : B :: C : D

This formula is read as follows:

A is to B as C is to D.

Alternatively, it can be read as follows:

The relationship between A and B is the same as the relationship between C and D.

Using this formula, we can represent the analogy that I just drew between phonemes and education as follows:

Formal education : nonformal education :: one allophone : another allophone

This can be read as follows:

The relationship between formal and nonformal education is the same as the relationship between two allophones of the same phoneme.

10.2.1 Analogical Change by Meaning

Analogy is a very powerful force in language change, and this fact was recognized by the Neogrammarians. Speakers of a language often perceive a partial similarity between two forms on the basis of their meaning alone, even when there is no similarity in their actual forms. Speakers of languages sometimes even change the shape of a word to become more like that of another word to which it is related only by meaning. To do this is to change the phonetic shape of a word by analogy, and we can express this using the following formula:

$$\text{meaning}_a : \text{meaning}_b :: \text{form}_a : \text{form}_b$$

Given that the relationship between form and meaning in language is by and large arbitrary (as Saussure noted toward the beginning of this century), we would not ordinarily expect that two related meanings would be expressed by related forms. However, similarities in meaning sometimes do cause words to change their shape so that they end up being phonologically closer to each other than they would have been if they had been subject to all of the regular sound changes. Consider the following months of the year:

September
October
November
December

Three of the four are similar; the word *October* is a little different. However, in some languages, the word for the tenth month of the year has changed to be something like **Octember*, by analogy with the other months of the year. The Russian word for 'October', *okt'abr'*, goes back to **oktember* and not to **oktober*. However, there's no particular reason that *October* should have changed to **Octember*; *November* could have changed to **Novober*, with the analogy going in the opposite direction.

As a further illustration of the point that analogy operates unpredictably, let us turn our attention to the words *deux* 'two', *trois* 'three', and *quatre* 'four' in some nonstandard varieties of modern French. When the word *quatre* appears before a noun that is pronounced with an initial vowel, some speakers of French now add a final /-z/ to the word *quatre*, making it *quatres*, on the analogy of the /-z/ at the end of the words *deux* and *trois*. So, compare the following examples:

| | Standard French | Nonstandard French | |
|-----------------|-----------------|--------------------|------------------|
| deux articles | døz aʁtikl | døz aʁtikl | 'two articles' |
| trois articles | tʁwaz aʁtikl | tʁwaz aʁtikl | 'three articles' |
| quatre articles | katz aʁtikl | katz aʁtikl | 'four articles' |

10.2.2 Analogical Change by Form

Analogy need not take just meaning as the basis for comparing two forms, as in the examples that we have just looked at. Analogical change can also operate when there is a perception of partial similarities between two forms without any consideration of meaning. For instance, earlier in the history of English there was a word *ewt*, which referred to a creature that looks like a small lizard. In modern English, this word has become *newt*, having unpredictably added an initial /n-/. It was not a regular change in English for /n-/ to be added to words that have initial vowels, so we need to find an explanation for this particular irregularity.

Once again, we can invoke analogy as the explanation. In English, we also have words like *name*, which have always had an initial /n-/, and words like *apple*, which have always had an initial vowel. The indefinite article in English varies in shape between *a* and *an*, with *a* occurring when the following noun begins with a consonant, and *an* occurring when there is a vowel at the beginning of the noun. So, compare the following:

- a name
- an apple

The old word *ewt* began with a vowel, so, according to this rule, the indefinite article should have taken the form *an* rather than *a*: that is, *an ewt*. However, in saying *an ewt*, earlier speakers of English evidently stopped breaking up the words between *an* and *ewt* as they started to associate this phrase with phrases like *a name* rather than with other phrases such as *an apple*. So, by analogy of one form with another, *an ewt* became *a newt*.

10.2.3 Analogical Extension and Leveling

In the preceding two sections we talked about analogy in terms of what was being analogized—that is, whether the analogy was on the basis of meaning. However, we can also describe analogy in terms of how it affects particular paradigms. Linguists have often drawn a distinction between extension and analogical leveling. In the first case, a form is extended from one paradigm into another. An example of this is the history of the Greek noun paradigms. Originally, the nominative plural of *o*-stems (a common noun class) ended in *-o:s*. At some point, however, the plural form from the pronominal paradigm—*oi*—was extended to this noun class. Eventually, the ending was analogically extended to the *a*-stem nouns as well, where **-a:s* was replaced by *-ai*.

An example of leveling comes from the history of German. Some German verbs have a stem-vowel alternation. There are several different classes of verbs, but one of them had *eu* in the singular and *ie* in the plural. For example, to say ‘I fly’ in early New High German, the form was *fleuge*, but to say ‘we fly’, the form is *fliegen*. Sometime in the early New High German period, this alternation was replaced, and the verbs that had *eu* in the singular came to be inflected with *ie* throughout the paradigm.

Analogical extension and leveling creates problems in morphological reconstruction. You might not be able to discover a clear conditioning environment, or you might be able to reconstruct a set of affixes but not necessarily recover all their meanings. A case like this is the distribution of core case marking in the Karnic languages of Central Australia. We can reconstruct an ergative case marker **-lu* and an accusative marker **-nha* (and the nominative/absolutive, which was zero-marked). But we can’t tell exactly what distribution these affixes had. The form **-nha* occurs on pronouns in all the languages, but in some languages it also occurs on nouns:

| | Pitta-Pitta | Arabana | Diyari | Yandruwandha | Wangkumara |
|-----------|-------------|---------|-----------|--------------|------------|
| nouns? | Yes | No | Sometimes | No | Yes |
| pronouns? | Yes | Yes | Yes | Yes | Yes |

We could say that Proto-Karnic had **-nha* on pronouns alone, and some languages started to mark nouns for accusative case by analogy with the pronouns. Alternatively, we could argue that Proto-Karnic marked accusative case on both nouns and pronouns and the accusative marking was later lost in some languages (giving those languages as split-ergative system of a different type). We cannot choose between those scenarios with the data available.

10.3 DOING MORPHOLOGICAL RECONSTRUCTION

In general terms, doing reconstruction in morphology (as opposed to reconstruction at the word level) is little different from the procedures you have learned so far. We can use both the comparative method and internal reconstruction in discovering morphological change, and the same caveats for both apply. If a change has happened in all languages and is not dependent on another change, we are unlikely to be able to reconstruct it. However, changes often leave records in the history of the languages.

Doing comparative morphological reconstruction requires you to assemble cognate sets, just as you would for comparative reconstruction for words. However, establishing cognacy for morphemes is more difficult than doing so for lexical items, because the strings tend to be shorter and because of the changes that we talked about in sec. 10.1. You need to consider not only the form of the morpheme but also its etymological history. By way of an example, consider the verb forms in the related languages Bardi and Nyikina (table 10.1). These inflected words are all cognate, in that they go back to the same Proto-Nyulnyulan verb forms. However, when we come to compare the

TABLE 10.1 Verb Forms in Related Languages Bardi and Nyikina

| Nyikina | | |
|------------------|-------------------------------|-----------------------------|
| Person | Nonfuture Intransitive Realis | Nonfuture Transitive Realis |
| First | ʔa-ʔ-kama | ʔa-n-kama |
| First and second | ya-ʔ-kama | ya-n-kama |
| Second | nyi-ʔ-kama | mi-n-kama |
| Third | yi-ʔ-kama | yi-n-kama |
| Bardi | | |
| Person | (Intransitive) Past Realis | Transitive Present Realis |
| First | ʔa-ʔ-kama | ʔa-n-kama |
| First and Second | a-ʔ-kama | a-n-kama |
| Second | mi-ʔ-kama | mi-n-kama |
| Third | i-ʔ-kama | i-n-kama |

individual morphemes, things become a great deal more complicated. This is because there was a change in the history of the Nyikina language which collapsed the present and past paradigms into a single nonfuture paradigm. The transitive forms go back to the earlier present paradigm, and the intransitive forms go back to the earlier past paradigm. This has led to a reanalysis of the content of the prefixes. The Nyikina marker that we would say denotes “intransitive” is actually cognate with the Bardi “past tense” morpheme. Therefore, it is important that morphemes are considered in the context of whole words, not only as isolated pieces. This is especially true if the languages are morphologically complicated.

Another thing to remember when the morphological reconstruction is that the changes you posit should be plausible. Reconstruction in morphology, like reconstruction in other areas of linguistics, requires assembling a case for content of a protolanguage and the changes that are hypothesized to derive the modern languages. To do this, you need to use all the available evidence, combined with what you know about language change in general and about synchronic systems.

Reading Guide Questions

1. Why is it more difficult to do reconstruction in morphology than in phonology? (There are reasons other than the ones we have discussed here, too. See if you can think of some.)
2. What is rule inversion?
3. Name some of the changes that may occur in the formal realization of a morpheme.
4. What are some of the ways that morphemes are lost?
5. Give an example where a morpheme boundary has been created.
6. How common is it for morphemes to be reordered within a word?
7. What is analogy?

Exercises

1. In Udi there are clitics, called PMs, or person markers, which attach to verbs. The following table has an example (*tayɣ-* means ‘go’, *-a* is the subjunctive marker):

| | Singular | | Plural | |
|------------|------------------|-------------------|--------------------|----------------------|
| 1st person | <i>tayɣ-a-zu</i> | ‘I would go’ | <i>tayɣ-a-yan</i> | ‘we would go’ |
| 2nd person | <i>tayɣ-a-nu</i> | ‘you would go’ | <i>tayɣ-a-nan</i> | ‘you (pl.) would go’ |
| 3rd person | <i>tayɣ-a-ne</i> | ‘he/she would go’ | <i>tayɣ-a-q’un</i> | ‘they would go’ |

Udi is a member of the Lezgian family. The Proto-Lezgian pronouns are given below, along with the Udi free pronouns and person markers.

| | Proto-Lezgian Independent Pronouns | Udi Independent Pronouns | Udi Person Markers |
|--------------------------|--|-----------------------------|-----------------------|
| 1st-person sing. | *zwə | zu | -zu |
| 2nd-person sing. | *G ^w ən | un | -nu |
| 1st-person pl. inclusive | *x:ə | | |
| 1st-person pl. exclusive | *ʒlən | yan | -yan |
| 2nd-person pl. | *z ^w ən | vən | -nan |

Consider these data and answer the following questions:

- Explain generally how person markers of the first and second person originated. Do not worry about details at this point.
- As shown in the second chart, Proto-Lezgian had both inclusive and exclusive pronouns. Examine the data in the first chart closely; why was the exclusive form used for the first-person plural in Udi, when either the exclusive or the inclusive could have been used?
- Notice the difference between the second-person singular independent pronoun in Udi and the person marker of the same person and number. Metathesis was not a regular process (did not apply in similar words). Can you suggest a reason it may have applied here?
- Notice the difference between the second-person plural independent pronoun and the person marker of the same person and number. Can you suggest a reason the person marker would have become *-nan* instead of *-van* or *-vən*?

In the third person, Udi uses proximate ('this one'), medial ('that one'), and distal ('yon one') deictic pronouns. It is believed that the third-person singular person marker, *-ne*, developed out of *-no*, the shared portion of these three pronouns in the absolutive singular. It is believed that the third-person plural person marker, *-q=un*, on the other hand, developed out of *-t=oyon*, the shared portion of these three pronouns in the ergative plural. (The first vowel was syncopated, producing *-t=yon*, then the two newly juxtaposed consonants fused into *q=*, producing *-q=on*. The vowel is a problem in both the singular and the plural.)

| | Proximate Pronoun | Medial Pronoun | Distal Pronoun | Person Marker |
|---------------------|----------------------|-------------------|-------------------|------------------|
| Singular absolutive | me-no | ka-no | še-no | -ne |
| Ergative plural | me-t'in | ka-t'in | še-t'in | |
| Absolutive | me-nor | ka-nor | še-nor | |
| Ergative | me-t'oyon | ka-t'oyon | še-t'oyon | -q'un |

- e. Can you explain why the singular would be based on the absolutive and the plural on the ergative? Hint: Look at the whole paradigm in the first chart and at the two choices for the third-person singular person marker and the two choices for the third-person plural person marker.
2. Bardi and Nyikina are both Nyulnyulan languages. Consider the following phrases in the two languages:

| | | |
|------------|------------|--------------|
| Bardi | Nyikina | |
| ɲamaɭa | nimaɭajanu | 'my hand' |
| ɲimaɭa | nimaɭajija | 'your hand' |
| nimaɭa | nimaɭajina | 'his hand' |
| ɲalma | nalmaɲanu | 'my head' |
| ɲalma | nalmaɲija | 'your head' |
| nalma | nalmaɲina | 'his head' |
| ɲana ba:wa | baba ɲanu | 'my child' |
| ɲija ba:wa | baba ɲija | 'your child' |
| ɲina ba:wa | baba ɲina | 'his child' |

- a. Describe the Bardi system, then do the same for the Nyikina system.
- b. What language is likely to reflect the order system? Why?
- c. Which language has changed? What was the change?
- d. What is the name for this change?
3. Consider the following inflected forms in Turkic languages. (This problem is based on R. L. Trask (1994:243), but revised, retranscribed, and expanded.)

| | Kazakh | Uzbek | Uyghur | Turkish | Tatar | Yakut | Turkmen | |
|----|---------|---------|-----------|----------|----------|-----------|----------|--------------|
| 1 | ɯolım | yolim | yolum | yolum | yulum | suolum | yolum | 'my way' |
| 2 | kölüm | kolim | kölüm | gölüm | külüm | küölüm | kölüm | 'my lake' |
| 3 | tuzdi | tuzli | tuzluq | tuzlu | tozlo | tu:sta:ɣ | duðlu | 'salty' |
| 4 | sütü | sutli | sütlük | silü | söttö | ü:tta:ɣ | θüytli | 'dairy' |
| 5 | ɯolımız | yolimiz | yolumiz | yolumuz | yulbuz | suolbut | yolumuð | 'our way' |
| 6 | kölümüz | kolimiz | kölümüz | gölümüz | külbüz | küölbüt | kölümüð | 'our lake' |
| 7 | onınfı | onintfi | ʔonuntfi | onundɯu | ununtfu | (onus) | onuntfu | 'tenth' |
| 8 | üfinfi | utfinçi | ʔütfüntfi | ütfündɯü | ötföntfö | (ühüs) | ütfündzi | 'third' |
| 9 | ɯoldı | yolni | yolni | yolun | yulunıɣ | suolun | yoluɣ | 'the way's' |
| 10 | köldi | kolni | kölni | gölü | külünüɣ | küölün | köluɣ | 'the lake's' |
| 11 | tuzsı | tuzsı | tussız | tuzsuz | tozsoz | tu:s suoɣ | duðið | 'salt-free' |

| | | | | | | | | |
|----|--------|--------|--------|--------|--------|---------|---------|----------------------------|
| 12 | sütsiz | sutsiz | sütsiz | sütsüz | sötsöz | üt suoχ | θüýtθið | 'milk-free' |
| 13 | zoli | yoli | yoli | yolu | yulu | suolu | yolu | 'its way' |
| 14 | köli | koli | köli | gölü | külü | küölü | kölü | 'its lake' |
| 15 | tuzdı | tuzni | tuzni | tuzu | tozno | tu:hu | duðu | 'the salt (<i>acc.</i>)' |
| 16 | sütti | sutni | sütni | sütü | sötnö | ü:tü | θüýtü | 'the milk (<i>acc.</i>)' |

- a. First look at the consonants. What would you reconstruct? What morphophonemic consonant processes are there in the daughter languages?
- b. Now consider the Kazakh data as a whole, concentrating on the vowels. On the basis of internal reconstruction, what do you think the prior forms looked like?
- c. Now look at the Uzbek data. Does Uzbek provide you with more information for reconstructing Turkic morphology? Why or why not?
- d. Do the same thing for the Turkish data.
- e. Consider the Uyghur data. What additional complications are there? (Internally reconstruct the Uyghur system, and compare it to your interim Turkic reconstruction.)
- f. Now, consider the whole dataset, using the Tatar, Yakut, and Turkmen data, too. What additional difficulties (and enlightenments) do these data provide?

(Hint: this problem is not about looking at the correspondences, although all the words are cognate. The trick to solving this problem is to think about the internal systems in the individual languages.

Finally, consider the following words for 'father' and 'horse':

| | Kazakh | Uzbek | Uyghur | Turkish | Tatar | Yakut | Türkmen | |
|----|--------|--------|--------|---------|--------|---------|---------|------------------------------|
| 17 | atım | otim | atam | atm | atım | atım | atım | 'my horse' |
| 18 | atımız | otimiz | atimiz | atımız | atıbız | atpıt | atımıθ | 'our horse' |
| 19 | atı | oti | atisi | atı | atı | atın | atı | 'its horse' |
| 20 | atsız | otsiz | assiz | atsız | atısız | at suoχ | atθið | 'without a horse' |
| 21 | attar | otlar | atlar | atlar | attar | attar | atlar | 'horses' |
| 22 | æke | ota | ata | ata | eti | aça | ata | 'father' |
| 23 | ækeni | otani | atini | atanı | etine | açanı | atanı | 'the father (<i>acc.</i>)' |
| 24 | ækeniŋ | otani | atiniŋ | atası | etineŋ | açatın | atanıŋ | 'father's' |
| 25 | ækeler | otalar | atilar | atalar | etiler | açalar | atalar | 'fathers' |

- g. These words provide a further clue for your problems in reconstructing Turkic morphology. What is it?

4. Consider the following data from Georgian:

| Old Georgian | | | | Modern Georgian | |
|--------------|-----------------|---------------|------------------|-----------------|---------------|
| Present | | Aorist | | Present | Aorist |
| v-t'ir | 'I cry' | v-i-t'ir-e | 'I cried' | v-t'ir-i | v-i-t'ir-e |
| t'ir | 'you cry' | i-t'ir-e | 'you cried' | t'ir-i | i-t'ir-e |
| t'ir-s | 'he/she cries' | i-t'ir-a | 'he/she cried' | t'ir-i-s | i-t'ir-a |
| v-i-cin-i | 'I smile' | gan-v-i-cin-e | 'I smiled' | v-i-cin-i | gan-v-i-cin-e |
| i-cin-i | 'you smile' | gan- i-cin-e | 'you smiled' | i-cin-i | gan- i-cin-e |
| i-cin-i-s | 'he/she smiles' | gan- i-cin-a | 'he/she smiled' | i-cin-i-s | gan- i-cin-a |
| v-i-marx-av | 'I fast' | v-i-marx-e | 'I fasted' | v-marx-ulob | v-i-marx-e |
| i-marx-av | 'you fast' | i-marx-e | 'you fasted' | marx-ulob | i-marx-e |
| i-marx-av-s | 'he/she fasts' | i-marx-a | 'he/she fasted' | marx-ulob-s | i-marx-a |
| v-mep-ob | 'I reign' | v-mep-e | 'I reigned' | v-mepob | v-i-mep-e |
| mep-ob | 'you reign' | mep-e | 'you reigned' | mepob | i-mep-e |
| mep-ob-s | 'he/she reigns' | mep-a | 'he/she reigned' | mepob-s | i-mep-a |

- In Old Georgian the prefix *i-* occurs in some paradigms and not in others; its function is not clear. Identify the function that this prefix has developed or is developing in the modern language.
- Explain the morphological changes involving the prefix *i-*. You do not need to account for the changes in suffixes.

Further Reading

Eugene Nida, "Analogical Change."

Harold Koch, "Reconstruction in Morphology."

Aditi Lahiri, Introduction to *Analogy, Levelling, Markedness*.

Henning Anderson, "Morphological Change: Towards a typology."

Jeffrey Heath, "Hermit Crabs: Formal Renewal of Morphology by Phonologically Mediated Affix Substitution."

Stephen Anderson, "Morphological Change."

Semantic and Lexical Change

I mentioned at the beginning of chapter 10 that phonological change has been fairly intensively studied in the world's languages. Grammatical change is less well studied, but it is an area that is receiving a lot of attention from linguists at the present. Semantic change, however, seems to be the area of diachronic linguistics that is least well understood. However, there are some observations that we can make as to the kinds of semantic changes that occur in languages and the forces that are involved in bringing these changes about.

If we read Shakespeare, Chaucer, or even Jane Austen, it's easy to see that some words are used in different ways from the way that we would use them. In other cases, there are phrases in English that are still current but that don't really make sense if we think about them. For example, in many translations of the Bible there is a phrase "the quick and the dead." This makes no sense on the face of it: Why would people who can move fast be grouped with the dead? It makes more sense if you know that the word *quick* used to mean 'living', and the phrase used to mean "the living and the dead" (and is translated as such in modern editions).

Changes in meaning can be divided into four basic types: broadening, narrowing, bifurcation (or split), and shift. In the following sections, I will discuss each of these in turn. We will then go on to talk about some of the other changes that happen to words over time.

11.1 BASIC MEANING CHANGES

11.1.1 Amelioration and Pejoration

Words can change their connotation over time. Some words acquire positive connotations, while others acquire negative ones. The word *silly*, for example, originally meant 'blessed', but today it is not a positive word. AMELIORATION is the technical term for words whose meaning changes to be more positive over time, while PEJORATION is the opposite.

11.1.2 Broadening

The term BROADENING is used to refer to a change in meaning that results in a word acquiring additional meanings to those that it originally had, while still retaining those original meanings as part of the new meaning. Quite a number of words have undergone semantic broadening in the history of English. The modern English word *dog*, for example, derives from the earlier form *dogge*, which was originally a particularly powerful breed of dog that originated in England. The word *bird* derives from the earlier word *bridde*, which originally referred only to young birds while still in the nest, but it has now been semantically broadened to refer to any birds at all.

11.1.3 Narrowing

Semantic NARROWING is the exact opposite of broadening. We say that narrowing takes place when a word comes to refer to only part of the original meaning. The history of the word *hound* in English neatly illustrates this process. This word was originally pronounced *hund* in English, and it was the generic word for any kind of dog at all. This original meaning is retained, for example, in German, where the word *Hund* simply means 'dog'. Over the centuries, however, the meaning of *hund* in English has become restricted to just those dogs used to chase game in the hunt, such as beagles. The word *meat* in English has also been semantically narrowed. It originally referred to any kind of food at all (and this original meaning is still reflected in the word *sweetmeats*), though now it only refers to food that derives from the flesh of slaughtered animals.

Words may also come to be associated with particular contexts, which is another type of narrowing. One example of this is the word *indigenous*, which when applied to people means especially the original inhabitants of a country which has been colonized, not "original inhabitants" more generally, (for example, Spanish people are never referred to as "indigenous" to Spain).

11.1.4 Bifurcation

A third type of semantic change can be called SEMANTIC SPLIT, OR BIFURCATION. These terms describe the change by which a word acquires another meaning that relates in some way to the original meaning. For instance, if you take the phrase *pitch black* in English, you will find that some people do not realize that the word *pitch* comes from the name of a very black substance like tar (or bitumen). These speakers of English might simply regard *pitch* in this example as meaning 'very' or 'completely'. If you were ever to hear anybody saying *pitch blue* or *pitch yellow*, then you would know that, for these people, the original meaning of *pitch* has split into two quite different meanings.

11.1.5 Shift

The final kind of semantic change that I will talk about is SEMANTIC SHIFT, where a word completely loses its original meaning and acquires a new meaning. In all of the examples of semantic shift that you have just learned about, at least something of the original meaning is retained, but this is not the case with semantic shift. The history of the word *silly* in English illustrates this process. This word is cognate with the German word *selig* 'blessed', and it is derived from *Seele* 'soul'. The meaning of the German word represents the original meaning of the word, so there has clearly been a major semantic shift to get from the meaning 'blessed' to the meaning in modern English of 'stupid' or 'reckless'.

Words obviously do not jump randomly from one meaning to another when they undergo semantic shift of this kind. They may shift in smaller steps that go under some of the headings that I have already presented, but as some original meanings are lost, the points of connection between intermediate semantic stages may also be lost. The German word *selig* has also acquired the meaning 'blissful' from its original meaning of 'blessed'. This represents an understandable semantic broadening, as somebody who is blessed is likely to feel blissful at the prospect of getting into heaven. From 'blissful', the more general meaning of 'happy' was acquired in German. Perhaps somebody who is happy ends up skipping around and being silly, giving us the modern English meaning of the word.

11.2 INFLUENCES IN DIRECTION OF CHANGE

When talking about semantic change, we can recognize a number of different forces that operate to influence the directions these changes take, including

metaphor, euphemism, hyperbole, and interference. I discuss each of these in turn.

11.2.1 Metaphor

A METAPHOR is an expression in which something is referred to by some other term because of a partial similarity between the two things. For example, if you say *Kali is a pig*, you do not mean literally that he is a pig but that there are certain things about his appearance or his behavior that remind you of a pig. Perhaps he eats a lot, or he eats sloppily, or he is an extremely dirty or untidy person. Sometimes the metaphoric use of a word can cause the original meaning to change in some way. The word *insult* in English originally meant 'to jump on'. Presumably, if you insulted someone, it was as though you had metaphorically jumped on them. However the metaphoric use of the word then completely took over the original word, and a semantic shift took place.

11.2.2 Euphemism

A EUPHEMISM is a term that we use to avoid some other term which has some kind of unpleasant associations about it or that is completely taboo in some contexts. For instance, in colonial Papua New Guinea, Europeans often referred to Melanesian people as "natives." As Papua New Guineans became more aware of the connotations of the word *native* (as it implies a certain backwardness), people had to find a new word to talk about Papua New Guineans that was not offensive. This is how the expression "a national" became the accepted expression to replace *native*. The term *national* has therefore undergone a semantic broadening in Papua New Guinea English under the pressure of euphemism. In Vanuatu, the word *native* was also felt to have offensive connotations, and a new term was also created there, but in this case out of local lexical resources, and the word *ni-Vanuatu* (literally: of + Vanuatu) was created. This word has become accepted, but those Europeans who still insist on putting Melanesian people down (but who dare not use the word *native*) have re-created their own insulting word from this new word and refer to *ni-Vanuatu* as *ni-Vans*.

11.2.3 Hyperbole

Some words in languages are felt to express meanings in a much stronger way than other words referring to the same thing. For instance, the two words *good* and *fantastic* can be used to refer to more or less the same things, but it is the second word that has the greater impact. Stronger words can often change to become more neutral if used often enough. This force in

semantic change is referred to as *HYPERBOLE*; this means that an originally strong connotation of a word is lost because of constant use. An example of this kind of development involves the change of earlier French *extonäre*, which originally meant 'strike with thunder'. This form has developed into modern French *étonner*, which simply means to 'surprise'.

11.2.4 Interference

A final force that operates in semantic change is *INTERFERENCE*. Sometimes one of a pair of similar words, or a pair of homonyms (i.e., words with the same form but totally different meanings) can undergo semantic change of one kind or another to avoid the possibility of confusion between the two meanings. The word *gay* in English is undergoing semantic shift at the moment as a result of interference. Until 30 years ago, in mainstream society, this simply meant 'happy' or 'cheerful'. Then the word *gay* underwent a semantic split and acquired the second meaning of 'defiant and proud homosexual'. When the heterosexual majority of the English-speaking population became aware of this new meaning of the word *gay*, they tended to avoid the word altogether when they wanted to express the fact that they were happy. People are now unlikely to say "I am gay" unless they want to declare that they are homosexual.

Another example of semantic interference involves the Bislama word *melek*. When the English word *milk* was originally copied into Bislama, this was the form that it took. The word *melek* then acquired a second meaning, that of 'semen'. The association of the word *melek* with the taboo connotations of the meaning 'semen' has recently become so strong that younger speakers of Bislama tend to avoid using the word *melek* to refer to plain milk and have reborrowed the English word 'milk' in the shape *milk*.

11.2.5 Folk Etymology

Another kind of analogy that we often find is referred to as *FOLK ETYMOLOGY* or *POPULAR ETYMOLOGY*. Etymology, as you have already seen, is the study of the history of words. When we speak of folk or popular etymology, we mean that people who speak a language often make their own guesses about what the history of a word is on the basis of partial similarities to some other words (and in doing this, they obviously have no interest in what the professional etymologist might have to say about the history of the word). Speakers of the language may then actually change the word so that its pronunciation comes more into line with what they think is the origin of the word.

Folk etymology tends to take place in words that are relatively long and in some sense felt to be "unusual" by speakers of the language. Speakers may then take part of this word, or all of it, and change it so that it looks more

like a word that they already know. For instance, the word *crayfish* in English was originally copied from an older French word *crévisse* (and it had nothing to do with fish at all). Ordinarily, such a word would probably have been copied into English as something like *creviss*. Although this word was a single morpheme in French, English speakers apparently felt that it was long or unusual enough in its sound that it must “really” be two morphemes. They noted a partial similarity in meaning between French *crévisse* and English *fish*, as both are edible creatures that live in water, and they also noticed the partial similarity in shape between French-*visse* and English *fish*. So, these earlier speakers of English changed the word to become *crayfish* because they felt that was what the word should have been according to their own view of where it came from. Professional linguists, of course, would say that the word *fish* originally had nothing to do with this word.

Folk etymology can be seen to be taking place when speakers make certain mistakes in pronunciation. A person who says *ashfelt* instead of *asphalt* is operating under this influence. Presumably they see the greyish-black color of the *asphalt* (which is referred to as ‘bitumen’, ‘tar’, ‘tar-seal’, or ‘tar macadam’ in other varieties of English) and equate it with the greyish-black ash from a fire, as well as the black color of felt cloth, and rename it accordingly. A person who refers to *watercress* as *water grass* is doing the same thing, and so is somebody who says *sparrow grass* instead of *asparagus*.

11.2.6 Hypercorrection

In chapter 13, you will see how variability is involved as a factor in causing the spread of language change, and one of the concepts that you will come across there is hypercorrection. HYPERCORRECTION refers to the situation when a word may have two possible pronunciations, one of which is regarded as prestigious (i.e., looked up to, or having positive social value), while the other is stigmatized (i.e., looked down on, or having negative social value). In many varieties of English, for example, there are two different ways of pronouncing the word *dance*: /dæns/ and /da:ns/. Of these, the second generally has higher social value than the first, and if you want to indicate that you are not from the working class, you are, or you want to indicate that you are not from the working class, you might use the more “posh” /da:ns/ pronunciation. However, if somebody substitutes a variable sound in a word or in an environment where it is not appropriate, then that person is engaging in hypercorrection, or “overcorrecting.” For instance, if someone were to accidentally say /ʌndəstænd/ instead of /ʌndəstænd/, this could be the reason. Another example from historical linguistics is the student who insists that the plurals of *suffix* and *prefix* are “prefices” and “suffices”, as though the words were Latin.

Another example comes from Bahasa Malaysia. In the standard variety of this language, there are words containing the phoneme /r/, and there are also words borrowed from Arabic that contain the voiced velar fricative /ɣ/.

In the area of Malaysia known as Perak, there is a variety of the language that is known locally as Celaka Perak, which translates as 'the Perak misfortune'. You will no doubt guess from its name that people think that this dialect sounds "funny" and that it is a stigmatized dialect. One of the features of Celaka Perak is that it merges the distinction between /r/ and /ɣ/, and all words containing these sounds are pronounced in Celaka Perak with the velar fricative. The result is that we find the following regular correspondences between standard Bahasa Malaysia and Celaka Perak:

| Standard Bahasa Malaysia | Celaka Perak | |
|--------------------------|--------------|------------|
| ratus | yatuih | 'hundred' |
| ribu | yibu | 'thousand' |
| buruk | buyuk | 'rotten' |
| loyat | loyat | 'accent' |

When somebody from Perak is trying to speak the standard language, one thing that they have to remember to do is to substitute /r/ for /ɣ/ in order to avoid sounding like Perak bumpkins. Mostly people can do this without making mistakes, but as there are only very few words containing /ɣ/ in the standard dialect, it is not too difficult to find people hypercorrecting in those few cases where there is supposed to be a velar fricative. So, if somebody from Perak pronounces /lorat/ 'accent' instead of /loyat/, they are producing an irregular sound correspondence (at least in their own speech) as a result of hypercorrection.

11.3 LEXICAL CHANGE

If you study the history of particular words in themselves rather than the changes in their actual pronunciations, you are engaging in a study of LEXICAL CHANGE (which is sometimes known as etymology). While some lexical items can be traced back all the way to a reconstructible protolanguage, there are almost certainly going to be some words in the lexicon of any given language that represent innovations since the breakup of the protolanguage.

11.3.1 Borrowing

Innovations in the lexicon can come from a number of different sources. One of the most common sources of new words in a language is words from a different language. Traditionally, linguists refer to this process as BORROWING. While using this term, many linguists express their unease about it, as a language that "borrows" a word from another language does not give it back, nor is the first language denied the use of a word that it has "lent" to

another language. It is more accurate to speak of one language COPYING words from another language, because this is precisely what happens. In this book, therefore, I have generally used the term “copying” rather than “borrowing” to refer to this process, though it should be kept in mind that both terms can be used to refer to the same process.

When a language copies a lexical item, it takes the form of a word in one language and generally reshapes that word to fit its own phonological structure. This means that non occurring phonemes may be replaced with phonemes that are present in the system of the language that is taking in the new word, and words may be made to fit the phonological pattern of a language by eliminating sounds that occur in unfamiliar positions or inserting sounds to make words fit its patterns. For instance, Tongan does not allow consonant clusters at all, nor does it allow word-final consonants. Tongan has no distinction between [l] and [r], either, so when Tongan speakers want to talk about an ice cream, they use a word that has been copied from English into Tongan, with the shape /aisikilimi/.

Languages are more likely to copy words from other languages in the area of cultural vocabulary than in core vocabulary. Core vocabulary is basically vocabulary that we can expect to find in all human languages. It is difficult to imagine any language that does not have some convenient way of expressing meanings like the following: *cry*, *walk*, *sleep*, *eat*, *water*, *stone*, *sky*, *wind*, *father*, and *die*. Cultural vocabulary, on the other hand, refers to meanings that are culture-specific, or which people learn through the experience of their own culture. Culture-specific meanings are obviously not core vocabulary, as only some languages have words to express these meanings: *tepee*, *potlatch*, and *peacepipe* (in North America), *frost* and *snow* (in nontropical climates), *kava* and *tapa* cloth (in the South Pacific), *dreamtime* and *rainbow serpent* (in Aboriginal Australia), *earthquake* and *lahar* (in geologically unstable areas), *television* and *internet* (in Western technological societies), *muezzin* and *haji* (in Muslim societies), and *trinity* and *resurrection* (in Christian societies).

There is some other terminology which is culture-specific, but this fact may not be obvious at first glance. “Thank you” is one good example of such an expression. Western children are constantly reminded to say thank you at every appropriate opportunity, but the verbal expression of thanks is a very Western habit. Many languages in the South Pacific, for example, do not have words to express this meaning, and it is not considered necessary in these cultures to express thanks in words (though thanks can still be expressed in other ways, of course). Even such apparently basic words as the numbers *one* to *ten* are not found in all languages. Very few Australian Aboriginal languages, for example, have separate words for numbers above three. Anything more than three is simply expressed by the word for ‘many’, or an awkward compound of the existing numbers could be used. In the Bandjalang language of northern New South Wales in Australia, for example, there are the numbers *jabur* ‘one’ and *bula:bu* ‘two’, and if you needed to express ‘seven’, you would say *bula:bu-bula:bu-bula:bu-jabur*. Given that this is awkward once the numbers get

any larger, it is clear that counting is something that was not done very often. The obvious explanation for this is that counting was not a major part of the nonacquisitive cultures of the Australian Aborigines.

No culture is constant, and often cultural changes are brought about as the result of contact with culturally or technologically different people. As European technology and beliefs have spread into the Pacific, many words of English origin have been copied into the languages of this region. Speakers of Motu in Papua New Guinea use the word *botolo* for 'bottle', the Māori use the word *hikareti* for 'cigarette', the Tongans refer to a 'car' as *motuka*, and the Paamese in Vanuatu refer to a 'letter' as a *ve:va* (from the English word *paper*). The expression *thank you* has now also been copied into Paamese, where it has been reshaped into the single word /tagio/. (In Paamese, sequences of [ju] are not possible, so the final vowel has been changed.)¹ It is not just English words that have been copied into Pacific languages; colonial powers have been introducing cultural changes to this part of the world since the mid 1800s. The French, for example, have contributed the word /lalene/ 'queen' into the languages of Wallis and Futuna (from *la reine*), and the Germans have contributed words like /beten/ 'pray' into some of the languages of New Guinea. There are loans from Arabic and Sanskrit into northern Australian Aboriginal languages: they entered the languages via Macassan traders over the last few hundred years. The widespread Yolŋu word *djorra*' (IPA [ju:ra]) 'paper' is ultimately from Arabic *surat* 'Koranic verse', for example.

While the non-core component of the lexicon is highly susceptible to change in a language because of the need to express technological and cultural change, lexical copying is not restricted just to the expression of new meanings. Younger generations of Paamese speakers frequently use the English-derived words /bu:s/ 'bush' and /ka:ren/ 'garden' instead of the indigenous words /leiai/ and /a:h/ (respectively) that their parents and grandparents use. There is no need for this, as the Paamese language already had perfectly good words to express these meanings. These are not the only "unnecessary" words that Paamese has copied. For instance, we also find words like /sta:t/ 'start', /ma:s/ 'must', and /ale/ 'OK then' (from French *allez*). Although there are perfectly adequate ways of expressing these meanings using indigenous Paamese words, few people use these words (and younger people would even have trouble saying what the Paamese word for 'start' actually is). Paamese has an efficient counting system, yet few younger speakers of the language can count in their language beyond five, preferring instead to use the English-derived terms /wan/, /tu/, /tiri/, /vo:/, /vaiv/, and so on. The same thing has happened in the Ndebele language of Zimbabwe, where English numbers above five tend to be used by younger people rather than the Ndebele numbers (probably because of the influence of English in schooling and mathematics education).

Why do people do this? It is difficult to find a good explanation. However, if a speaker of English uses the French-derived expression *coup de grâce* instead of 'final blow', many people would suspect that the speaker

is trying to demonstrate his or her level of education. In the same way, when speakers of Pacific languages use words that are copied from English, they may simply be trying to say that they consider themselves to be much more of the modern world than the old-fashioned world of their grandparents.

Although lexical copying is frequently associated with dominant economic and political powers, any kind of cultural contact can bring about lexical copying between languages. There had been long-term contact between Tongans and Fijians from well before the first European arrived in the Pacific, and there has been much copying of vocabulary between these two languages. Similarly, there are many words of Kiribati origin in the lexicon of the Tuvaluan language.

The Rotuman language of Fiji shows evidence of having copied words from Polynesian languages at different periods in history. Sometimes we find the same original form being regularly inherited with one meaning and later copies with a slightly different meaning. For instance, the form *toka 'come ashore' has been directly inherited as *foʔa* with the same meaning. However, the word was later copied from another language, where it had not changed its shape, so we now find the word /toka/ meaning 'settle down' in Rotuman. Cases such as this are referred to as DOUBLET—historically related pairs of words in which one is directly inherited while the other is a later copy from a related language.

Obviously, if a Pacific language has copied a word from a language with which it is not related genetically, it is going to be fairly easy to identify the word as being a relatively new addition to the lexicon. When a language copies words from a language with which it is fairly closely related, it might be more difficult to recognize it as a later lexical innovation, especially if the borrowing is extensive.

There are other reasons why languages undergo lexical change. In many cultures in the Pacific and Australia, for instance, there is a strong tendency to name people after some particularly noticeable occurrence in the environment at the time of the child's birth. For instance, a child born during a violent thunderstorm might be called Lightning. One child born out of wedlock in Vanuatu in the 1980s was called Disco because it was after a night of dancing that he was conceived. In some societies, there are powerful social restrictions against mentioning people's names in certain situations. In many Australian Aboriginal societies, for example, it is forbidden to mention somebody's name for a period of time after they have died.² In modern times, this restriction carries over to a prohibition against hearing their voice on tape or seeing their face in a photograph or on video. If somebody is named after some common thing and that person dies, then speakers of that language cannot use the name of that thing, either. In situations like this, the easiest way of avoiding the problem is to copy a word meaning the same thing from a nearby language. Australian Aborigines

traditionally spoke more than one language anyway, so this was often easy to do.

In the Kabana language (spoken in the West New Britain province of Papua New Guinea), people typically have personal names that also refer to everyday objects. In this society, as in many other Melanesian societies, there is a strong restriction against saying the names of one's in-laws. This is true even if you want to refer to the actual thing that your in-law is named after, and you are not using the word as a personal name at all. In cases such as these, the language has a set of special words that are held "in reserve." These special, reserved items are either words in the Kabana language itself (but with a different meaning) or words copied from neighboring languages with the same meaning. For example, the word in Kabana for a particular kind of fish is *urae*. If your in-law is called Urae, this fish must be referred to instead as *moi*, which is usually the word for 'taro'. The word for 'crocodile' in Kabana is *puaea*, but this word cannot be used if your in-law is called Puaea, and the crocodile must be referred to instead as *bagele*. This form *bagele* is apparently copied from a nearby language, where the word for 'crocodile' is actually *vagele*.

A similar kind of cultural practice is found in Polynesia, though here the restriction against the use of words is associated with chiefly status. There is a custom in Tahiti, for example, that is known as *piti*, and this custom states that the name of a chief (or even a part of the name of a chief) cannot be used by ordinary people. So, for instance, during the time that the very powerful chief called Pomare was in power, the very common words *poo* 'night' and *mare* 'cough' became taboo simply because they sounded like parts of the chief's name. The word *poo* was replaced by the word *ru?i* and the word *mare* was replaced by the word *hota*.

Another kind of restriction among the Wampar speakers of Morobe province in Papua New Guinea involves place-name taboo. Certain places are regarded as sacred, perhaps because the people's ancestors' blood had been spilled there or because their ancestors are buried there. If Wampar people today use the names of these places, it is believed that the ancestral spirits will punish the people by causing disasters, sickness, or the failure of the crops on which they depend for food. The people of this area also have a similar kind of restriction to the Kabana practice of not saying the names of in-laws. People have a range of options available that allow them to talk about things and at the same time avoid breaking these taboos. Some languages have two or three synonymous terms to refer to the same thing, especially for very common words. Another possibility is for people to substitute a word that is semantically related to the taboo word in some way. For example, in the Mari language of this area, if the word /zah/ 'fire' is restricted, the word *pakap* 'ashes' can be used to talk about fire instead.

Words can be lost in a language, and new words can be created for reasons that are not at all obvious. Sometimes when a new word appears in a language,

we have no idea where it came from. The English word *man*, for example, has a very long history. It has cognates in other Germanic languages such as the German *Mann*, and it can be traced all the way back to Proto-Indo-European (compare the Sanskrit word *manu*). The English word *boy*, however, is something of a mystery, as it appears in the historical record only after English became a separate language, and it has no known cognates in any other Indo-European languages. There are several possible explanations for this. One is that the word from which *boy* was derived was in fact present earlier, but it was lost at the same time in all other languages related to English. Another possibility is that *boy* was borrowed from some other language. However, we have no idea what language that might have been.

A final possibility is that *boy* represents a genuine lexical innovation in English. It is hardly ever the case that words genuinely spring out of nowhere. Occasionally a word like *googol* is invented (in this case by a mathematician's child, to refer to the numeral 1 followed by 100 zeroes), but generally words have some basis in preexisting forms. Presumably what happened in the case of *boy* is that some other existing word took on this new meaning and the old meaning was lost altogether. However, we have no evidence that this is what actually happened, so what we are left with is a word that looks as though it suddenly sprang into the lexicon out of nowhere.

11.3.2 Internal Lexical Innovation

Lexical copying is not the only source of lexical changes as a way of expressing cultural changes. Speakers of languages also make use of their own linguistic resources in creating new words. If they take an existing word and extend its area of reference to express a new meaning, this becomes an example of semantic change which has been used to fill a lexical gap in the language. For instance, when the Paamese people in Vanuatu saw their first airplane, it must have looked to them like a large bird. The word for 'bird' in Paamese is /aman/, and this word is now also used as the Paamese word for 'airplane'. People also fill lexical gaps by generating new words and joining existing words together in new compounds, according to the existing rules of the language, in order to express new meanings. When the Fijians first saw planes, they called them instead /waga-vuka/, which is derived from the words /waga/ 'canoe' and /vuka/ 'fly'. An airport in Paamese is an /out ten aman/, which literally means 'place of birds (i.e., airplanes)'.

11.3.3 Shortening Words

There are a number of ways in which words end up being shortened. One special category of lexical innovations involves compression, clipping, or

shortening. This typically applies only to a few words in a language, although the productivity of these processes varies greatly from language to language.

COMPRESSION is the process of dropping off one or more syllables from the end or middle of a word, for example:

| | | |
|----------------|---|--------------|
| administration | → | admin |
| university | → | uni, varsity |
| David | → | Dave |
| Thomas | → | Tom |

In fact, in Australian and New Zealand English there is often an additional syllable added to the compressed forms to express a kind of diminutive meaning:³

| | | |
|------------------------|---|---------------|
| football | → | footie |
| biscuit | → | bikkie |
| Christmas | → | Chrissie |
| present | → | prezzie |
| hotwater bottle | → | hottie |
| truck driver | → | truckie |
| wharf laborer | → | wharfie |
| Salvation Army | → | Salvo, Sallie |
| journalist | → | journie |
| politician | → | pollie |
| conscientious objector | → | conshie |
| Brisbane | → | Brizzie |
| documentary | → | doco |

Another particular kind of compression involves the use of initials. Examples of this kind of lexical change using only initials include the following:

| | | |
|---------------------------------------|---|------|
| Canadian Broadcasting Corporation | → | CBC |
| television | → | TV |
| World Health Organization | → | WHO |
| Ministry of Foreign Affairs and Trade | → | MFAT |

It is sometimes possible for initials completely to lose their association with the forms from which they are derived and to be reanalyzed as a new lexical item. For instance, in Bislama (in Vanuatu), there is a word /kao/ meaning

'flat out, fast asleep, completely used up'. This derives from the French pronunciation of the first letters of the English abbreviation K.O., which stands for 'knockout' (in boxing). Few speakers of Bislama would be aware of the source of this item as an abbreviation for K.O., however, and a genuinely new word has entered the lexicon in this way.

Another possible source for new lexical items is word mixes or blends. By this, I mean new words that are created by taking parts of two different words and adding them together to make up a completely new word. For instance, the following word mixes are frequently used in Papua New Guinea:

| | | |
|--------------------------|---|---------|
| Administrative College | → | Adcol |
| Electricity Commission | → | Elcom |
| University of Technology | → | Unitech |

This kind of change seems to be particularly common in government departments and in relation to administration generally. In fact, in Indonesia, there has developed a special register of Bahasa Indonesia that is commonly used in the newspapers where there are many word mixes of this kind (as well as many abbreviations). People in Indonesia sometimes find it difficult to read some parts of the newspaper because so many word mixes and abbreviations are used as totally new lexical items. New lexical items of this type also seem to be entering the English vocabulary in advertisements. For instance, forgettable kettles that switch themselves off when the water has boiled are called *forgettles*, and folding, environmentally friendly bottles are referred to as *fottles*. In Namibia, many government organizations begin with Nam-; for example, the electricity company is called *Nampower*. This is also a type of blending.

11.4 CONSEQUENCES OF BORROWING AND IRREGULAR LEXICAL CHANGE

11.4.1 Semantic Change

All the types of change we've discussed in this chapter cause potential problems for reconstruction. In the first case, if a word has shifted, it might be difficult to show that it belongs with other cognates that look similar. Such arguments often come down to one linguist's intuition about what is or isn't a likely semantic change. Is it likely that a word for 'time' could shift into a word meaning 'water'? Maybe not in a single step, but words that mean both 'time' and 'tide' are attested around the world (note that English *tide* is cognate with the German word *Zeit*, and in fact exhibits the change of 'time' > 'tide'); and 'tide' may be an alternative expression for 'water', as in *The tide was lapping against the top of the seawall*. How about 'money', 'hill',

and 'pumpkin-fish' (a type of tropical fish)? Proto-Nyulnyulan **wanaŋarti* has cognates in daughter languages in all these meanings. In such cases it might be possible to reconstruct a word, but not its meaning.

11.4.2 Borrowing and Copying

Lexical copying is another factor that can cause sound correspondences between two languages to show up as irregular or unpredictable. As we saw earlier in this chapter, it is possible for a language to copy a cognate form from another language that has undergone different sound changes to its own words. If a sufficiently large number of words have been copied into a language, it sometimes becomes difficult to establish what the correct sound correspondences should be. Another result of lexical copying is that sometimes a single word in a protolanguage may appear to have two reflexes, both of which clearly derive from the same original form.

In English, for example, the regular reflex of /**sk*/ is /*f*/, but alongside words such as *ship* and *shirt* (which correctly reflect the original pronunciation) we also find words such as *skiff* and *skirt*, which are derived from the same sources. It might be tempting to say that /**sk*/ sporadically became /*sk*/ in English, while generally being reflected as /*f*/. However, /**sk*/ did in fact regularly become /*f*/, and the /*sk*/ forms were reintroduced at a later date in words from Danish (which had not undergone the same change as English had by that stage). If you were trying to reconstruct the history of English phonology by applying the comparative method, you would therefore need to exclude *skirt* and *skiff* when you drew up your list of sound correspondences. You should not let the fact that there is a *sk* : *sk* correspondence between English and Danish force you to reconstruct an additional contrast in the protolanguage, as it is only the *sk* : *f* correspondence that goes directly back to a phoneme in the protolanguage.

Sometimes when there are several different sets of sound correspondences in a number of related languages, some of these correspondences may be the result of lexical copying rather than being directly inherited forms. While repeated (rather than sporadic) correspondences are normally taken to point to separate original forms, as you saw in chapter 5 (as long as they cannot be shown to be in complementary distribution with other correspondences), it is possible for large-scale lexical copying at different points in history to show up as separate sound correspondences. One famous case involves the Rotuman language of Fiji. Rotuman is spoken on the island of Rotuma in what is politically part of Fiji, yet it is closely related to the Polynesian languages. In addition to words that are clearly derived directly from Proto-Polynesian, there are separate sets of sound correspondences between Rotuman and other Polynesian languages which suggest that there have been two waves of other Polynesian words that have been copied on a large scale into the vocabulary of Rotuman since it diverged from its sister languages.

When words are copied from languages that are unrelated, or only distantly related, this causes few problems in recognition, as there will normally be sufficient difference in shape between the kinds of words found in both languages to make their source obvious. However, it can become difficult to distinguish copied forms from directly inherited forms when words from one dialect are copied into another closely related dialect (as often happens in some of the smaller languages of Melanesia, for example), as these are generally similar to each other. Look at the following examples from the Sinaugoro and Motu languages of Central Province in Papua New Guinea:

| Sinaugoro | Motu | |
|-----------|-------|--------------|
| yita | ita | 'see' |
| yutu | utu | 'lice' |
| yate | ase | 'liver' |
| yulita | urita | 'octopus' |
| tuliya | turia | 'bone' |
| yatoi | yatoi | 'egg' |
| leyi | rei | 'long grass' |

From this set of cognates, there are two sound correspondences involving the velar fricative in Sinaugoro. First, there is a correspondence of Sinaugoro /ɣ/ to Motu /ø/; second, there is a correspondence of Sinaugoro /ɣ/ to Motu /ɣ/. Clearly, however, you should be suspicious of the $\gamma : \gamma$ correspondence, as there is only one example in the data. If you had more data, you would be in a better position to judge whether there is a single example of this correspondence or whether there are more words in these two languages that correspond in the same way. If it turns out that, in fact, this is a sporadic correspondence in these two languages, its irregularity could easily be explained by saying that Motu copied the Sinaugoro word /yatoi/ for 'egg' instead of keeping its own original word /atoi/, which no longer exists in the language. However, there is no way of deciding just by looking at the Motu word /yatoi/, as it looks like a perfectly ordinary Motu word.

When dealing with copied vocabulary, things can get very complicated indeed when you come to carry out the reconstruction of linguistic history. Some languages have relatively little vocabulary that is of foreign origin, while other languages have incorporated huge numbers of words from other languages. Sometimes there has been so much vocabulary entering a language from outside sources that linguists are genuinely confused about what family the language belongs to. For instance, the Maisin language of Oro Province in Papua New Guinea has been variously described by linguists as being Austronesian with considerable non-Austronesian influence, non-Austronesian with considerable Austronesian influence, and finally as a truly

mixed language. The confusion has arisen because whatever conclusion we come to, we must recognize that there has been massive copying of vocabulary from some outside source.

Reading Guide Questions

1. What is folk etymology?
2. What is meant by lexical copying? How can this cause sound correspondences between languages to become unpredictable?
3. What is semantic broadening?
4. What is semantic narrowing?
5. What does the term "bifurcation" mean with respect to semantic change?
6. What is semantic shift, and how does this kind of change differ from the other kinds of semantic change mentioned in this chapter?
7. How can metaphor influence the direction of a semantic change?
8. What is euphemism? How can it influence semantic change?
9. What is meant by hyperbole, and how is this involved in semantic change?
10. What is meant by interference when speaking of change of meaning?
11. What is lexical borrowing, or copying?
12. What is the difference between cultural and core vocabulary?
13. What possible ways are there for a language to fill lexical gaps?
14. What problems can lexical copying cause in reconstructing the phonological history of a language?
15. What is the possible effect of lexical taboo in vocabulary change?
16. What do we mean by lexical innovation?
17. What is lexical compression?
18. What are word mixes?
19. What is analogical sound change? How can it affect the way we apply the comparative method?
20. In what way can semantic or grammatical factors influence the direction of a sound change?

Exercises

1. A thesaurus is a book that lists words by meaning, and which makes it possible to find out the synonyms of a word. Look up some synonyms for the following words in a thesaurus: *popular*, *fantastic*, *native*, *juvenile*. Then find a dictionary that goes back a couple of hundred years, if possible (for instance Samuel Johnson's), and see how these words have changed semantically.
2. Compare the meanings of the following forms in English and Tok Pisin (with the meanings in Tok Pisin given on the right). How would you describe the nature of the changes that have taken place?

| English | Tok Pisin | |
|-----------|-----------|--|
| arse | as | 'buttocks, basis, foundation, tree trunk, stem of plant' |
| bed | bet | 'bed, shelf' |
| box | bokis | 'box, crate, cardboard carton, vagina' |
| garden | garen | 'plot of ground planted out to food crops for a single season' |
| grass | gras | 'grass, hair, whiskers' |
| hand | han | 'hand, arm, wrist, branch of tree' |
| cargo | kago | 'material possessions' |
| copper | kapa | 'roofing iron' |
| cry | krai | 'cry, weep, wail, moan' |
| straight | stret | 'straight, correct' |
| take away | tekewe | 'peel (of skin)' |

3. What is the plural of Walkman? If you use more than one mouse with your computer, what do you say? If you say *Walkmans* and *mouses* rather than *Walkmen* and *mice*, why might this be?

Further Reading

- Anthony Arlotto, *Introduction to Historical Linguistics*, chapter 10, "Semantic Change," pp. 165–83.
 Leonard Bloomfield, *Language*, chapter 24, "Semantic Change," pp. 425–43.
 Elizabeth C. Traugott and Richard B. Dasher, *Regularity in Semantic Change*, chapter 1, "The Framework," pp. 1–50.
 David P. Wilkins, "Natural Tendencies of Semantic Change and the Search for Cognates."

Syntactic Change

12.1 STUDYING SYNTACTIC CHANGE

Studying syntactic change has proceeded rather differently in linguistics from the way some change and morphological change have been studied. Some have argued that it is not possible to study syntactic change in the same way. These arguments are primarily due to David Lightfoot. When we study sound change and reconstruct using the comparative method, we compare forms and meanings in different languages with one another. It is the form-meaning pairs together which allow us to make these hypotheses. We cannot study the sounds alone unless they are arranged in words. This is because it's only when we compare full words that we get environments for sound change. When we come to syntax, however, it becomes a little more difficult to define our correspondence sets. It is because while phonemes have a set place in a word, and can only be substituted for one another in limited circumstances, words and sentences can very freely be substituted for one another. When we study syntax, we are studying the rules that we infer from sentence data. If we treat these as correspondence sets, they are sets of very abstract items.

Others have taken a less pessimistic view. They have pointed out that there are some cases where we can study words in a syntactic context. For example, question words like 'what' and 'who' are lexical items that have grammar associated with them. For example, English question words must come at the start of a sentence. Complementizers are another example of a word class with both syntax and comparable words. Harris and

Campbell (1995: chapter 1) have quite a lot of discussion about the different arguments.

Another difference between historical syntax and the study of sound change is that there has been quite a bit of focus on internal reconstruction in syntax and the changes that are attested in the history of individual languages. There has also recently been much work on the historical relationship between syntax and morphology. This is part of grammaticalization theory, and we discuss some of this next.

12.2 TYPOLOGY AND GRAMMATICAL CHANGE

Languages of the world can be classified according to their grammatical typology. A typological classification of languages is one that looks for certain features of a language, and groups that language with another language that shares the same features. A TYPOLOGICAL CLASSIFICATION differs fundamentally from a GENETIC CLASSIFICATION of languages. While two languages may be grouped together typologically, this does not mean that they are genetically related, though, of course, it may turn out that this is the case. Similarly, it is possible for two languages that are genetically related to be typologically quite different. English and the Tolai language of Papua New Guinea, for example, belong to the same typological grouping if we consider the fact that they both share the same basic word order: subject + verb + object. Tolai and Motu (also of Papua New Guinea) are both genetically related in the Austronesian language family, yet they belong to different typological groups if we consider their basic word orders. The basic word order in Motu is subject + object + verb.

While it is possible for a language to belong to only one genetic classification, we can group languages into as many typological groups as we want, depending on which particular linguistic feature we want to classify them by. If we were to classify languages according to the way in which they express inalienable possession in noun phrases, we would find that Tolai and Motu both belong to the same typological group, while English behaves quite differently. In both Tolai and Motu, pronominal suffixes are added to nouns, whereas in English, a separate possessive pronoun precedes the noun to express the same meaning. Examine the following examples:

| | |
|-----------|-----------|
| Tolai | Motu |
| bilau-gu | idu-gu |
| nose-my | nose-my |
| 'my nose' | 'my nose' |

(In this particular case, Tolai and Motu are typologically similar because both have inherited a feature that was present in the protolanguage through which they are genetically related.)

Typological classifications of languages can be based on whatever features we might find it useful to base them on. Some shared features are of little general interest, while other features are of much greater interest. In the study of grammatical change, linguists are interested in looking at how languages evolve from one grammatical type to another. I now describe some of the major grammatical typologies, and you will see how languages that belong in each of these typological groups may have come to be like that, or how they might change typologically in the future.

It can be observed that diverse languages tend to change independently in similar sorts of ways. For instance, certain types of lexical items—especially verbs or locational items—often change to become prepositions or postpositions (which can be collectively referred to as ADPOSITIONS). Adpositions can then become attracted to nouns to become AFFIXES. Affixes can then be lost, which means that other grammatical strategies must be developed to express the functions originally expressed by the now-lost forms.

Typological changes such as I have just described are not always unidirectional, however. By this I mean that it is possible for a variety of different sorts of changes to follow from a single starting point, as it is also possible for some of these changes to operate in the reverse direction. If language change were unidirectional, then human language—in all the typological diversity that we find today—would be inexorably moving toward a single type of language. What we find, in fact, is that the typological mix of the world's languages has been constantly changing in a variety of directions at once, resulting in today's typological mix.

12.2.1 Morphological Type

Languages can be grouped according to their MORPHOLOGICAL TYPE—that is, the way in which the main features of the grammar are expressed morphologically.

The first type of language that I will talk about is the ISOLATING TYPE of language. Such a language is one in which there tends to be only one morpheme per word; that is, there are many free morphemes with very few bound morphemes. A language of this type would be the Hiri Motu language of Papua New Guinea. If you examine the following sentence, you will see that each word expresses only a single meaning:

| | | | | | | | |
|------------------------------------|--------|---------|-----|-----|--------|------|----------|
| Lauegu | sinana | gwarume | ta | ia | hoia | Koki | dekenai. |
| My | mother | fish | one | she | bought | Koki | at |
| 'My mother bought a fish at Koki.' | | | | | | | |

A second type of language is what we call the **AGGLUTINATING TYPE**. An agglutinating language is one in which a word may contain many separate morphemes—both free morphemes and bound morphemes. However, the boundaries between morphemes in an agglutinating language are clear and easy to recognize, and it is as if the bits of the language were simply “glued” together to make up larger words. In such a language, each morpheme typically expresses a single meaning, while words typically consist of several—perhaps even many—morphemes combined together. A language such as *Sye* (spoken on the island of Erromango in Vanuatu) has agglutinating constructions in sentences of the following type:

ov-nevyarep yu-tw-ampy-oyh-or u-ntoy
 plural-boy they-will-not-want-to-see-them in-sea
 ‘The boys will not want to see them in the sea.’

The single word /yu-tw-ampy-oyh-or/ ‘they will not want to see them’, for example, expresses several meanings, some expressed by the prefixes (*yu-* ‘they’, *tw-* ‘will not’, *ampy-* ‘want to’), one by the suffix (*-or* ‘them’), and one by the root (*oyh* ‘see’).

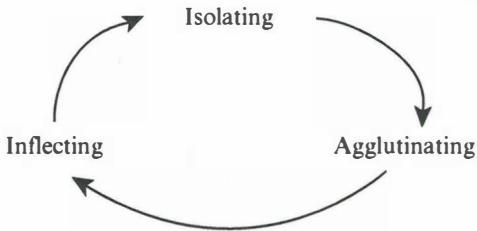
A third type of language that we can consider is the **INFLECTIONAL TYPE**. Inflectional languages are those in which there are many morphemes included within a single word, but the boundaries between one morpheme and another are not clear. So, in inflectional languages, there are many meanings per word, but there is not a clear “gluing” together of the morphemes, as is the case with agglutinating languages. An example of an inflecting language is Latin. Examine the following sentence:

Marcellus amat Sophiam.
 M-subject loves S-object
 ‘Marcus loves Sophie.’

Each of these words contains a number of different meanings. In the first word, we can recognize the root *Marcell-*, but the single suffix *-us* expresses a number of different meanings. For one thing, it indicates that *Marcell-* is the subject of the verb (rather than the object), and it also indicates that *Marcell-* is both masculine in gender and singular in number. In the case of *Sophiam*, the root is *Sophi(a)-*, and the suffix *-am* indicates that she is the object (rather than the subject), that she is feminine, and that she also is singular. Finally, the word *amat* includes the meaning of ‘love’, as well as indicating that this particular activity takes place in the present tense, and that the one performing the activity is in the third person, as well as being singular.

If any one of these items of meaning in any of these words were to be changed, then a different form of the word would have to be used. As Latin is an inflectional language, you should also note that although we can recognize a suffix of the form *-us* on the root *Marcell-*, and a suffix *-m* on the noun *Sophia-*, we cannot further subdivide either of these suffixes corresponding to the various meanings that these both express. That is, there is no single morpheme that expresses the meaning of 'singular', for example, or 'feminine', or 'subject'. The fact that a singular masculine subject is indicated by means of the single suffix *-us* is a typical characteristic of an inflectional language.

There is a tendency for languages to change typologically according to a kind of cycle. Isolating languages tend to move toward agglutinating structures. Agglutinating languages tend to move toward the inflectional type, and, finally, inflecting languages tend to become less inflectional over time and more isolating. This cycle can be represented by the following diagram:



Isolating languages become agglutinating in structure by a process of phonological reduction. By this I mean that free-form grammatical markers may become phonologically reduced to unstressed bound form markers (i.e., suffixes and prefixes). If we look at modern Melanesian Pidgin, for example (at least as it is spoken, rather than written), we can see that a number of grammatical changes appear to be taking place. First, the prepositions that are written as if they are pronounced /loŋ/ 'on, at, in' and /bloŋ/ 'of, for' tend to be pronounced nowadays as prefixes to the following noun phrases. The forms of these evolving prefixes are:

lo-/blo- before consonants

l-/bl- before vowels

So we find that changes such as the following seem to be taking place:

aus bloŋ mi > aus blo-mi
 house of me house of-me
 'my house'

loj aus > l-aus
 at home at-home
 'at home'

Not only are these two prepositions being phonologically reduced in this way, but so too are some of the preverbal tense and mood markers. For instance, the future marker /bai/ is now sometimes reduced to the prefix /b-/ when the following word begins with a vowel rather than a consonant. Compare the following:

bai yu go
 future you go
 'you will go'

b-em i go
 future-(s)he predicate go
 'he/she will go'

As I have said, languages of the agglutinating type tend to change toward the inflectional type. By the process of morphological fusion, two originally clearly divisible morphemes in a word may change in such a way that the boundary is no longer clearly recognizable. We could exemplify this process of morphological fusion by looking at the following example from Paamese (spoken in Vanuatu). The marker of the first-person singular subject on verbs can be reconstructed at an earlier stage as /*na-/ , and the second-person singular subject marker can be reconstructed as /*ko-/ , and these are the forms that are still retained in modern Paamese, for example:

na-lesi-ø
 I-see-it
 'I see it'

ko-lesi-nau
 you-see-me
 'you see me'

Other tenses, as well as the negative, are expressed by adding other prefixes and suffixes in sequence, for example:

ko-va-ro-lesi-nau-tei
 you-immediate future-not-see-me-not
 'you are not going to see me'

The distant future tense was also originally marked in the same way, by a prefix of the form **i-* which appeared after the subject marker, in the same position as is occupied in the example that I just gave you by the prefix *va-*. However, the future tense marker **i-* fused morphologically with the preceding subject prefix. So, what was originally **na-* followed by **i-* became *ni-*, and what was originally **ko-* followed by **i-* became *ki-*:

| | | |
|---------------------|---|------------------|
| <i>*na-i-lesi-ø</i> | > | <i>ni-lesi-ø</i> |
| I-future-see-it | | I+future-see-it |
| 'I will see it' | | 'I will see it' |

| | | |
|-----------------------|---|--------------------|
| <i>*ko-i-lesi-nau</i> | > | <i>ki-lesi-nau</i> |
| you-future-see-me | | you+future-see-me |
| 'you will see me' | | 'you will see me' |

In modern Paamese, we can no longer divide the *ni-* and *ki-* prefixes into a subject marker and a future tense marker, as *n-* and *k-* do not occur anywhere else in the language as recognizable morphemes, and there is no longer any clearly recognizable *i-* morpheme as a future marker. We must therefore regard these two prefixes in modern Paamese as expressing two meanings at once. Such morphemes are called "portmanteau morphemes". This situation has arisen as a result of the fusion of two originally separate morphemes into one form. When this kind of fusion affects the grammar of a language in a major way, then the language can be said to have changed from an agglutinating type to an inflectional type.

Finally, languages of the inflectional type tend to change to the isolating type; this process is called MORPHOLOGICAL REDUCTION. It is common for inflectional morphemes to become more and more reduced, until sometimes they disappear altogether. The forms that are left, after the complete disappearance of inflectional morphemes, consist of single morphemes. The functions that were originally expressed by the inflectional suffixes then come to be expressed by word order or by free-form morphemes. As I indicated earlier, Latin was an inflectional language. So many ideas were expressed in a single word that there was no need in Latin for word order to be rigidly fixed. Words could occur in any order because the one who was performing an action and the one who was on the receiving end of an action were always marked in the suffixes that were attached to the noun phrases themselves. So, the meaning of the sentence that you saw earlier could be equally well expressed in Latin in any of the following ways:

Marcellus amat Sophiam.

Sophiam amat Marcellus.

Sophiam Marcellus amat.

Amat Sophiam Marcellus.

'Marcus loves Sophie.'

To indicate that the roles are reversed in this situation (i.e., that it is Sophie who is keen on Marcus), we would need to change the marking on the nouns, but the word order could be just as variable. We could indicate that it is Sophie who loves Marcus by the following sentence:

Sophia-ø amat Marcell-um.

Sophie-subject loves Marcus-object

'Sophie loves Marcus.'

However, any of the following would do just as well to express the same meaning in this inflectional language:

Marcellum amat Sophia.

Sophia Marcellum amat.

Amat Sophia Marcellum.

Latin evolved into modern Italian, and in the process lost a lot of its original inflections, thereby moving toward the isolating type. Nouns in Italian are no longer marked by suffixes to indicate whether they are the subject or the object, and they do not change in form as they did in Latin. In modern Italian, the only way to express the fact that Marcus loves Sophie is the following:

Marcello ama Sophia.

Marcus loves Sophie

'Marcus loves Sophie.'

Whereas, in Latin we would be free to change the order of these words without changing the meaning, this is no longer possible in Italian, as the nouns have lost their suffixes that indicate subject and object. If we were to change the Italian sentence that I just gave you into the following sentence, we would change the meaning as well:

Sophia ama Marcello.

Sophie loves Marcus

'Sophie loves Marcus.'

In modern Italian, it is now word order alone that marks the difference between the subject and the object of a verb, whereas before it was the presence or absence of an inflectional suffix on the noun.

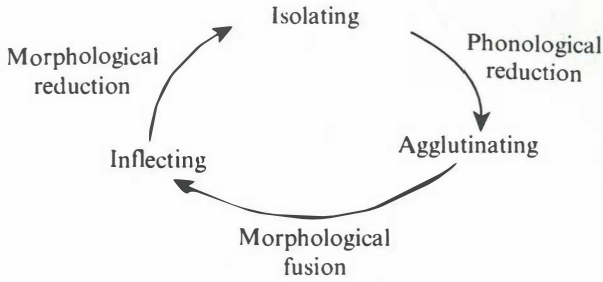


FIGURE 12.1 Typological cycle of language transformation

This typological cycle, and the processes involved in the transformation from one type to another, can be summarized in figure 12.1.

There is, in fact, a fourth type of language: those having **POLYSYNTHETIC** morphology. Such languages represent extreme forms of agglutinating languages in which single words correspond to what in other kinds of languages are expressed as whole clauses. Thus, a single word may include nominal subjects and objects, and possibly also adverbial information, and even noncore nominal arguments in the clause, such as direct objects and spatial noun phrases. The following example from the Yimas language of Papua New Guinea illustrates a polysynthetic structure:

na- ŋa -mpa -na -ŋkan -mpan -ra amtra
 plural give now imperative few them -class marker food
 'You few give them food now!'

Polysynthetic languages can develop out of more analytic (i.e., non-polysynthetic) languages by a process of argument incorporation. In English, we find some evidence of this kind of construction in the form of incorporated objects, such as the following:

Professor Hawne took up pipe smoking to make himself look pompous.

In the example, a generic object such as pipe can be preposed to a transitive verb such as *smoke*, instead of its usual position after the verb. In fact, we can even incorporate spatial noun phrases in the same sort of way, as in the following:

He just sat there star gazing.

Since *gaze* is an intransitive verb, this sentence can only be derived from the following, in which the incorporated noun stars appears in a prepositional phrase:

He just sat there and gazed at the stars.

It is possible for such patterns to become established as the normal pattern in a language and for these to completely replace earlier patterns in which there are free form nominal arguments and other kind of arguments in a clause. This is called UNIVERBATION. It is currently believed that the only way polysynthetic languages arise is through progressive univerbation and analogical extension of the patterns created by univerbation. (We see some more examples of this in sec. 12.3.)

12.2.2 Accusative and Ergative Languages

Languages of the world can also be grouped typologically according to the way in which they mark the subject and object noun phrases in a sentence. In a language like English, we speak of the subject of a verb and its object. The subject is the noun that comes before the verb and which causes the verb to choose the suffix *-s* if it is singular and *-ø* if it is plural, when the verb is in the present tense. The object is the noun phrase that comes after the verb in English. So we have sentences like the following in English:

The vice-chancellor is praising the students.

Subject (singular) Verb (singular) Object

The vice-chancellors are praising the students.

Subject (plural) Verb (plural) Object

Other languages may differ from English in the way that the subject and the object noun phrases are marked. Look at the following sentences in the Bandjalang language of northern New South Wales (in Australia):

Mali-ju bajgal-u mala ja:jam buma-ni.

the man the child hit-past

'The man hit the child.'

Mala bajgal gaware-:la.

the man run-present

'The man is running.'

Mali-ju ja:jam-bu mala bajgal ja:-ni.

the child the man see-past

'The child saw the man.'

Here the noun *bajgal* 'man' appears in two separate forms, either *bajgalu* (with the suffix *-u*) or just *bajgal* (with no suffix). The word that precedes

it also varies in its shape. When the word for 'man' appears with the suffix *-u*, this word has the form *maliju*, but when the word for 'man' appears without any suffix, the preceding word has the shape *mala*. If you examine the sentences carefully, you find that the noun phrase appears as *maliju bajgalu* when it is the subject of the transitive verb *buma-* 'hit', but when it is the subject of the intransitive verb *gaware-* 'run', it appears without any suffixes, as *mala bajgal*. You also see that when the same noun phrase appears as the object of the transitive verb *ja:-* 'see', it also has the unsuffixed form *mala bajgal*. The noun phrase referring to 'the child' behaves in exactly the same way. When the child is the object of the verb *buma-* 'hit', the object appears without any suffix as *mala ja:jam* 'the child', but when the child functions as the subject of the transitive verb *ja:-* 'see', it appears with suffixes: *maliju ja:jambu*, for example. (The forms of the suffix on the word *bajgal* 'man' and *ja:jam* 'child' are different, but these are phonologically determined allomorphs of the same morpheme.)

If you compare the structure of English and Bandjalang sentences, you see that there are three basic grammatical functions that are being expressed in the two languages, but in different ways in both cases. In English, we have intransitive subject and transitive subject being marked in the same way, and being distinguished from transitive object. In Bandjalang, however, we have intransitive subject and transitive object being marked in the same way, while these two functions are distinguished from transitive subject. In a language like English, the transitive and intransitive subject functions are referred to collectively as the **NOMINATIVE NOUN PHRASES**, while the transitive object is said to be the **ACCUSATIVE NOUN PHRASE**. In a language like Bandjalang, the transitive subject is referred to as the **ERGATIVE NOUN PHRASE**, while the intransitive subject and the transitive object noun phrases are referred to collectively as the **ABSOLUTIVE NOUN PHRASES**.

Languages in the world fall into one of these two basic typological groupings, though the type represented by English is about twice as common as the type represented by Bandjalang. (It is also possible for languages to be structurally intermediate between the two patterns.) With such different types of languages, we cannot really use the term "subject" for all languages of the world because it will have to mean different things, depending on which of these two types of languages we are looking at. To make it clear which type of system we are talking about, we need to distinguish between two basic types of languages: **nominative-accusative languages** (such as English) and **ergative-absolutive languages** (such as Bandjalang). Sometimes these labels can be shortened, so English can also be called an **accusative language**, and Bandjalang can be called an **ergative language**.

Just as it is possible for a language to change its basic morphological type over time, it is also possible for an accusative language to evolve into an ergative language and for an ergative language to become an accusative language. Most Australian languages behave like Bandjalang—that is, they are ergative rather than accusative—and we would reconstruct ergativity at

least to Proto-Pama-Nyungan (the ancestor language of Bandjalang and about 150 of the 250 languages traditionally spoken in Australia). However, there are other cases where ergativity has a known source. Hittite (one of the languages of ancient Anatolia [modern Turkey]) had an ergative case marker *-ants* (transliterated *-anza*) which was originally a form of the ablative case. Hittite ergativity is assumed to have originated in sentences where there was an instrument but no overt subject. (The data and analysis are from Garrett 1990.) Consider the following made-up Hittite sentence:

- | | | |
|-------------------------------------|---------------------------|------------|
| n=at | witenanza | parkunuzi |
| particle=he/she | water-ablative (singular) | makes pure |
| a. 'He/she purifies it with water.' | | |
| b. 'Water purifies it.' | | |

In the first interpretation, there is an overt subject and the 'water' is marked in the ablative case. The second sentence shows the presumed reanalysis, where the presumed subject was lost and the instrument was reinterpreted as the subject of the clause.

Another origin of ergative marking is in passives. To see how this might have happened, consider the common properties of ergatives and passives. In an ergative sentence, the object is in an unmarked case (usually the absolutive) and the subject has the ergative case (that is, it is overtly marked). Furthermore, the ergative-marked noun phrase is usually the agent of the verb and the object is usually the patient. In passive sentences, the patient is also in an unmarked case, although it is the subject of the clause. The agent is in an oblique case (or has a preposition, as in English passives). However, over time, the construction lost the passive meaning, and the agent came to be interpreted as the subject of the sentence. Agents tend to be subjects, so the reanalysis brings the construction into line with a common pattern. The old case marking pattern of the (former) passive remained. Here is a summary of the stages:

- | | | |
|----------|------------------------|--------------------------|
| Stage I | Patient is subject. | Agent is oblique. |
| Stage II | Patient is object. | Agent is subject. |
| Stage I | Patient is nominative. | Agent is oblique-marked. |
| Stage II | Patient is absolutive. | Agent is ergative. |

This is not the only way that ergative marking can arise, although it is a common way. The other main way that ergativity arises is also through reanalysis, but of a different construction.

One of the ways that languages express perfective aspect is through possession. There are many examples in the world's language of possessive constructions being reanalyzed. The English construction with *have* plus

another verb is an example of this. Consider the following sentence: *I have written a letter*. The construction originally meant the equivalent of 'I have a written letter': that is, I possess a letter which was written (by someone, not necessarily by me). Over time, the possessor came to be equated with the writer of the letter, and the implication of the clause changed. The phrase *I have a written letter* has an emphasis on the result (that is, on the letter and the fact that it is written). The implication shifted to an implication that the possessor of the letter did the writing and has finished writing (and has letter to show for it). This is essentially what the perfective means.

This is one type of possessive, but not all possession is marked with the verb like *have*. Some languages use the verb *be* and put the possessor in a possessive or oblique case. Think about what would happen if the same reanalysis that we just talked about happened with this other type of possession. In the first case, the possessor ended up as the subject of the sentence and the writer of the letter. The participle *written* became reanalyzed as dependent on the verb *have* rather than an adjective going with the object. The possessive marker also became a marker of aspect. If we assume the same type of reanalysis, we would expect the possessor to end up as the subject, except this time the possessor is not in the nominative case. If the possessor is reanalyzed as the subject but retains its earlier case, we get the case-marking patterns of ergativity. This seems to have happened in some Central American languages and perhaps also in the history of Hindi.

Of course, ergative languages can also change to become accusative languages. Just as accusative languages often have passive constructions, ergative languages often have what are referred to as ANTIPASSIVE constructions. In an antipassive sentence, a transitive verb with an ergative subject is structurally marked and detransitized, with the original subject receiving absolutive marking. The original absolutive object is then marked in some other way. If the original antipassive function of the marker on the verb were to have this function obscured over time—perhaps by phonological reduction or loss, or the acquisition of new functions—then we would be left with a system of accusative marking. This actually happened in some Australian Aboriginal languages spoken in western Australia.

12.2.3 Basic Constituent Order

When I talk about basic constituent order, I am referring to the relative order in the sentence of the three major components: the verb and the noun phrases that are centrally associated with it, these being the subject and object noun phrases. Languages of the world can be grouped typologically according to the way that these three major constituents in the sentence are ordered. Most languages have the order subject + verb + object (SVO)—English is a language of this type. The next most frequently

found order is subject + object + verb (SOV). The only other commonly found order is verb + subject + object (VSO). (There are three other logical possibilities for the order of constituents in a sentence: OVS, OSV, and VOS. However, these orders are much rarer among languages of the world.)

Many of the Austronesian languages of the Pacific, along with English as I have already said, are SVO languages. The Tolai language of New Britain in Papua New Guinea is a language of this type, as shown by the following example:

A pap i gire tikana tutana.
 the dog it see one man
 SUBJECT VERB OBJECT
 'The dog saw a man.'

The Austronesian languages of Central and Milne Bay Provinces of Papua New Guinea, however, are generally of the SOV type. For example, the same sentence in Motu would be expressed as:

Sisia ese tau ta e-ita-ia.
 dog subject man one it-see-him
 SUBJECT OBJECT VERB
 'The dog saw a man.'

The Austronesian languages of Central and Milne Bay Provinces appear to have changed their word order from the earlier order of SVO to the SOV order that they now have. Some scholars have argued that this change took place when the ancestor language from which Motu and its closer relatives are descended came into contact with the non-Austronesian languages of the area, as all of these non-Austronesian languages are SOV languages. For instance, in the non-Austronesian Koita language, which is spoken by the neighboring group to the Motu, the sentence that I have just given for Tolai and Motu would be expressed as follows:

Tora ata be eraya-nu.
 dog man one saw-him
 SUBJECT OBJECT VERB
 'The dog saw a man.'

Language contact is not the only possible explanation for a change in basic word order, as languages clearly do undergo these sorts of changes without any evidence that language contact is involved. Many languages that have one particular basic constituent order often allow competing patterns in

certain structural contexts. German, for example is an SVO language in main clauses, as shown by the following:

Der Mann sah den Hund.
 the man saw the dog
 'The man saw the dog.'

In subordinate clauses, however, German has SOV order, as shown by the following:

Ich glaube, dass der Mann den Hund sah.
 I believe that the man the dog saw
 'I believe that the man saw the dog.'

When there are competing structures of this type, it is possible for one of the two patterns to be generalized to other contexts and for the typology of the language to change. Note, however, that I am not trying to say here that German is moving from SVO to SOV constituent order. In this case, there is good reason to believe that the earliest Germanic languages were SOV and that this older order is preserved in subordinate clauses in German. Here are two examples from early texts. The first is a runic inscription from roughly 1,500 years ago; the second is a ninth-century Old English inscription.

ek Hlewagastiz Holtijaz horna **tawido**
 I H. H. horn did
 'I, H. H., made this horn.'

Æþred mec **ah**. Eanred mec **agrof**.
 Æ me owns E. me carved
 'Æthred owns me; Eanred carved me.'

One often finds that main clauses have innovated a new pattern, while the old one is preserved in less-frequent clause types such as subordinate structures.

Other languages allow alternative word orders as a way of expressing purely stylistic contrasts in particular contextual environments. For instance, in an SVO language, it may be possible to focus attention on the object by moving that noun phrase to the beginning of the sentence, or by moving the subject to the end of the sentence. Even though English is an SVO language, we sometimes find OSV orders in sentences such as the following:

I quite like Harry, but John I can't stand.

Similarly, although French is an SVO language, we also find constructions such as the following in the colloquial language which appear to have a VOS order (the pronoun *il* has many properties of being a clitic to the verb):

Il aime bien sa petite fille le vieux mec.
 he love much his little daughter the old guy
 'The old guy really loves his little daughter.'

Again, if constructions such as these originally purely stylistic variants were to take over from the dominant patterns, then a change of constituent order typology would have taken place.

12.2.4 Verb Chains and Serialization

While there are many grammatical facts that we could consider when setting up language typologies, the final example of typological change that I look at in this chapter is the development of what is called in some languages VERB CHAINS OR SERIAL VERBS. In some languages, we find that whole series of verbs can be strung together, sometimes in a single phonological word, with just a single subject and a single object. For instance, in the non-Austronesian Alambak language of the East Sepik in Papua New Guinea, we find sentences such as these:

Wifërt fir gëngimë-t-a.
 wind blow cold-past-it-me
 'The wind blew me and I got cold (i.e., 'the wind blew me cold').'

Another example comes this time from the Paamese language of Vanuatu (which is an Austronesian language):

Keik ko-ro: vul a:i.
 you you-sat break plank
 'You sat on the plank, breaking it.'

Verb-serializing languages sometimes even allow three (or more) verbs to be chained together in single constructions of this type. For instance, in the Yimas language, which is a close neighbor of the Alambak language, we find complex examples of clause chaining such as the following:

Na-bu-wul-cay-pra-kiak.
 him-they-afraid-try-come-past
 'They tried to frighten him as he came.'

Such constructions are not possible at all in English. Thus, we do not use equivalent constructions such as the following:

- *The wind blew-colded me.
- *You sat-broke the plank.
- *They tried-frighten-he-cam him.

Serial verb constructions of this type are common in the languages of eastern and southeastern Asia and in western Africa, as well as in the non-Austronesian languages of Melanesia. There is also evidence of serial verb constructions in some of the Oceanic languages, as well as Australian languages.

In languages that have these kinds of constructions, it is often possible to show that these chains of verbs originate from much simpler constructions in which each verb had its own set of subject and object noun phrases. For instance, the complex Alambgak structure that you have just seen could be derived from the Alambgak equivalents of the following:

- 'The wind blew me.'
- 'I got cold.'

Languages which develop serial verbs of this type are generally (but not always) SOV languages. This is not surprising, as this order allows speakers simply to state the subject and the object once at the beginning and then string the verbs together one after the other following these two noun phrases. It is then a relatively small step for these chained verbs to be "collapsed" into a single grammatical unit or even a single word.

12.3 GRAMMATICALIZATION

Words in languages can be grouped into two basic categories: lexical words and grammatical words. LEXICAL WORDS are those which have definable meanings of their own when they appear, independently of any linguistic context: *elephant, trumpet, large*. GRAMMATICAL WORDS, in contrast, only have meanings when they occur in the company of other words, and they relate those other words together to form a grammatical sentence. Such words in English include *the, these, on, and my*. Grammatical words constitute the mortar in a wall, while lexical words are more like the bricks.

If a particular meaning is expressed by a grammatical rather than a lexical word, the form is obligatorily present. For instance, in the sentence *I will come later*, the meaning of 'future tense' is expressed twice — first in the auxiliary

will and second in the adverb *later*. Of these, *will* is a grammatical word and *later* is not because we cannot omit the future marker *will*, whereas we can omit the future marker *later*:

*I come later.

I will come.

Words in languages can often change from being lexical words to grammatical words. This process is referred to as GRAMMATICALIZATION. We can see evidence of grammaticalization in progress in English with the following sentences:

I'm going to cut a piece of chocolate cake.

I'm going to the supermarket.

Although these two sentences both contain the sequence *going to*, these two words do not have the same status in both cases; it is only in the first sentence that we can contract *going to* to give *gonna*. Thus:

I'm gonna cut a piece of chocolate cake.

*I'm gonna the supermarket.

In the first example, it is clear that the meaning of *going to/gonna* is different from the meaning of *going to* in the second example. Rather than expressing the purely lexical meaning of the intransitive verb *go*, this sequence in the first sentence expresses a kind of intentional future tense. In this case, then, we say that *going to* has been grammaticalized and English has acquired a new kind of auxiliary, along with other auxiliaries such as *can*, *will*, and *might*, and other more recently grammaticalized auxiliary-like constituents such as *oughta*, *wanna*, and *hafta*.

Grammaticalization can affect lexical words in a variety of ways, though there is a tendency for forms to become increasingly closely linked to some lexical form in a sentence as the process continues. The change from lexical word to grammatical word is only the first step in the process of grammaticalization, with the next step being MORPHOLOGIZATION—the development of a bound form out of what was originally a free form. In fact, morphologization can also involve degrees of bonding between bound forms and other forms as it is possible to distinguish between clitics and affixes. A clitic is a bound form which is analyzed as being attached to a whole phrase rather than to just a single word. An affix, however, is attached as either a prefix or a suffix directly to a word.

In the Syle language of Erromango in Vanuatu, the free form /im/ 'and' is currently developing into a clitic with the shape /m-/ and this attaches to the beginning of whatever happens to be the second element of two coordinated

noun phrases. It is possible to say either of the following in this language, in which /im/ appears as a free form:

| | | | | |
|---------------------|-----|----------|---------------------|------------|
| netor | im | nevyarep | netor | m-nevyarep |
| Netor | and | boy | Netor | and-boy |
| 'Netor and the boy' | | | 'Netor and the boy' | |

However, when some other constituent intervenes between the coordinator and the second noun, the coordinator can be attached to whatever happens to be the first constituent of the second noun phrase:

| | | | | | | |
|----------------------|-----|--------|----------|----------------------|------------|----------|
| netor | im | ovon | nevyarep | netor | m-ovon | nevyarep |
| Netor | and | plural | boy | Netor | and-plural | boy |
| 'Netor and the boys' | | | | 'Netor and the boys' | | |

Morphologization can proceed one step further, with lexical forms (or clitics) becoming genuine word-level affixes. There are many languages in which locative affixes on nouns began as free postpositions or prepositions, while before this they were ordinary lexical items with some kind of locational meaning. In this discussion of morphologization, it is impossible not to refer back to the earlier discussion of morphological change in languages, where I demonstrated that isolating languages tend to move toward agglutinating structures, while agglutinating structures tend to move toward inflecting structures (Sec. 12.2 ff). These kinds of changes clearly involve increasingly grammaticalized (and correspondingly delexicalized) patterns.

Lexical items can obviously grammaticalize to varying extents and in differing ways in languages. Despite the varying possible end results, the process is a strongly unidirectional one in that lexical items generally become grammaticalized, while grammatical items generally do not become lexical items. As an example of how grammaticalization can develop along a continuum from a fully lexical item to a fully morphologized affix, let us consider some developments affecting some verbs in Oceanic languages.

In the Paamese language of Vanuatu, there are two verbs of the shape /kur/ 'take' and /vul/ 'break':

| | | | | | |
|---------------------|--------|-------|----------------------|---------|-------|
| inau | na-kur | a:i | inau | na-vul | a:i |
| I | I-took | stick | I | I-broke | stick |
| 'I took the stick.' | | | 'I broke the stick.' | | |

In Paamese, the verb /vul/ 'break' can also enter into a serial verb construction in which both verbs retain their lexical status, as follows:

inau na-kur vul a:i
 I I-took broke stick
 'I took the stick, thereby breaking it.'

However, in languages to which Paamese is related, the form that originally occupied the second slot in this kind of serial verb construction no longer occurs as an independent verb. There is typically a restricted set of forms in such languages that can behave in this way, so what was originally a lexical verb has been grammaticalized to become a kind of postverbal modifier of some kind. Examine the following example from the Numbami language of Papua New Guinea:

i-tala ai tomu
 he-chopped tree broke
 'He chopped the tree, thereby breaking it.'

In this case, the form /tomu/ 'break' cannot be used as a verb in its own right. Thus, it is not possible to say:

*i-tomu ai
 he-broke tree
 'He broke the tree.'

Other languages may then undergo further grammaticalization in which forms behaving like *tomu* in Numbami end up as verbal affixes that express meanings that are still clearly related to the meanings of the verbs from which they were originally derived. In some cases, a preverbal grammaticalized item may become a kind of classificatory verbal prefix that is attached to a general semantic category of verbs. For instance, all verbs that involve some kind of finger action, such as pinching, picking, plucking, flicking, and so on, might be marked by a prefix that derives from a verb that perhaps originally meant something like 'pinch'. In the Manam language of Papua New Guinea such a development has taken place, so we find the verb *sere?* 'break', along with the prefixed form /ʔin-sere?/ 'break with the fingers'. The verb *sere?* 'break' is then free to appear with other classificatory prefixes, such as *tara-* 'do by chopping', which therefore gives *tara-sere?* 'break by chopping'.

Given that grammaticalization is a diachronic process, it is possible for synchronic descriptions of languages to represent situations that are still only partly grammaticalized. In such cases, the distinction that I made at the beginning of this section between lexical and grammatical items will seem somewhat arbitrary. Instead of a clearcut distinction between these two categories of words, there will appear to be a continuum between two extremes.

For instance, the Paamese serial verb construction that I described earlier (sec. 12.2) is already moving along the way toward grammaticalization with some verbs. For one thing, the great majority of verbs in Paamese cannot appear in the second structural slot in such constructions. While some verbs can appear in either the first or second slot, other forms can never appear as independent verbs. Such forms have therefore already undergone functional restriction to postverbal modifiers. Thus, the form /vini:/ 'kill'—which derives from an earlier genuine verb with the same meaning—can now only ever occur as a serialized verb, never as an independent verb. Thus:

| | | | | | | |
|-------------------------------|-----------|--------|------|---------------------|----------|------|
| inau | na-sal | vini: | vuas | *inau | na-vini: | vuas |
| I | I-speared | killed | pig | I | I-killed | pig |
| 'I speared the pig to death.' | | | | 'I killed the pig.' | | |

Occasionally, partially grammaticalized forms may have very unusual features, which makes it difficult to assign them to one word class or another. For example, in some Admiralty Islands languages (spoken just north of mainland Papua New Guinea), serial verb constructions have partially grammaticalized into prepositional phrases. They have the distribution and functions of prepositions, and they do not behave like verbs. For example, regular verbs in the language take subject agreement forms, whereas these verbs do not agree with anything. However, they are not regular prepositions, either, because they take tense marking.¹

12.3.1 Direction of Grammaticalization

Grammaticalization tends to be a unidirectional process, with forms moving along a continuum of increasingly grammaticalized status:

lexical word > grammatical word > clitic > agglutinated affix > portmanteau affix

So far, we have looked in some detail at the creation of grammatical structure from lexical items. Occasionally, however, the change goes on the opposite direction, and lexical structure is created from items that were previously only grammatical.

While grammaticalization is quite a common process, the reverse—DEGRAMMATICALIZATION (or lexicalization)—is attested, though it is much rarer. Some examples can be given of this kind of change, however. For instance, a grammatical item such as the suffix *-burger* in words such as *hamburger*, *cheeseburger*, and *fishburger*, has become a genuine noun in English, and it is possible nowadays to ask for just a *burger*. The forms *pro-* and *anti-* were originally just prefixes in English in words such as *pro-democratic*

and *anti-Castro*. However, these days, they can also be used as lexical adjectives:

Are you pro or anti?

She is more anti than I am.

The affix *-ism* was borrowed into English on words of Greek origin, but these days, it is a separate word in its own right:

All these isms are getting really irritating.

12.3.2 Grammaticalization and Reconstruction

Grammaticalization covers many aspects of change. It has a syntactic component (the syntax of utterances changes), a morphological component, and a semantic component (and sound change, too).

Reconstruction using grammaticalization theory is somewhat similar to using the principles of internal reconstruction, except a greater emphasis is placed on typological parallels and on general pathways of meaning and grammar change. For example, if we were to consider an inflectional language like Latin, the principles of grammaticalization theory suggest that we should look for the origins of Latin inflection in structures that are more agglutinating. Furthermore, we should look for the origins of agglutinating languages in isolating structures that have been phonologically reduced. Of course, we know that isolating languages tend to develop morphology through phonological reduction, so it is logical to look for agglutinating structures in prior isolating structures. However, that is not the only way that agglutinating structures arise. Languages also develop new morphology through other processes (some of which we have seen in chapter 10). Such processes hold as general principles, but they are not deterministic enough at the level of individual morphemes to allow them to be used alone as solid evidence for reconstruction.

Another problem is that construction types may be generalized from the initial locus. Consider the following example:²

The Kyng *had Werre*, with hem of Sithie.

The King *had war* with the Scythians.

This is a Middle English example from approximately the year 1400 (in the writings of Sir John Mandeville). It is an example of a very common construction, where the verb *have* is used without its full lexical possessive meaning. This is called a LIGHT VERB. Now, from the standpoint of grammaticalization,

we would argue that light verbs originally had their full meaning in such a construction, and over time the verb was bleached and the construction acquired the idiomaticity that it has today. In the case of *have war*, that particular noun and verb no longer combine for most English speakers. However, this is not actually what happened. What seems to have happened is that the verb *have* acquired an idiomatic meaning in some contexts (such as *have a brother* or *have a cold*), and it was this already partially bleached meaning that was extended into phrases such as *have war*. We can tell this because of the long documentary history of these constructions in English. Relying on universal pathways would give us the wrong answer here.

In conclusion, reconstruction using grammaticalization heavily privileges general known pathways as the type of evidence used (just as the comparative method heavily privileges regularity of correspondences).

12.4 MECHANISMS OF GRAMMATICAL CHANGE

Three general factors seem to be involved in one way or another whenever grammatical change occurs. These factors are reanalysis, analogy, and diffusion. We see some of these mechanisms in chapter 10 where we talk about morphology, but the same processes also apply in syntax and morphosyntax. I discuss each of these mechanisms in this section.³

12.4.1 Reanalysis

REANALYSIS in grammatical change refers to the process by which a form comes to be treated in a different way grammatically from the way in which it was treated by speakers in previous stages of the language. This happens when a string of words is ambiguous in some way. The ambiguity can lie in one of a number of different areas, including constituency, the grammatical categories of the items, and the grammatical relations.

We have already seen a number of examples of reanalysis in syntax. Our analysis of ergativity in sec. 12.2.2 and the discussion of serial verb constructions in sec. 12.2.4 both relied crucially on reanalysis as the mechanism for change. When we discussed the rise of ergativity in Hittite, for example, we saw that the crucial construction was one where the singular instrument noun in the ablative case could be analyzed as the subject of the sentence. This would be reanalysis of constituency and of grammatical relations.

The Hittite example brings up another point: the analysis in syntax can result in the same surface structure of sentences, even though they have a different underlying structure. In such cases, we may not see that a reanalysis

has taken place initially. In other cases, a reanalysis brings with it a change in surface form. My example here comes from the Yandruwandha language of central Australia (Breen 2004). Yandruwandha, along with a number of other languages in the area, has verb suffixes that provide further information about the manner that the action of the verb was done in. One of these is *-thalka*. As a free adverb it means 'up', but as a verb suffix it specifically means that the action of the verb is directed upward:

Mayatha-li nhulu yadamani pangki-parndri-thalka-na
 boss-ergative he.ergative horse rib-hit-up-imperfect
 mardrawita-ngadi.
 stone.hill-dative
 'The boss galloped his horse up the hill.'

Here an independent word has been reanalyzed as belonging with the verb stem, and we can see this because now tense morphology such as the imperfective marking goes after *thalka*. (*Thalka* is one of quite a few adverbs and former verb stems which have been reanalyzed as manner suffixes.)

12.4.2 Analogy and Extension

We saw some examples of analogy already, in sec. 10.2. Those examples were in morphology, but analogical change also applies in syntax. Remember that analogical extension is the extension of a pattern from one part of the grammar to an area of the grammar where it previously did not apply. We saw the example of plural marking in some Greek nouns being taken from pronoun paradigms. An equivalent example in syntax comes from Laz, a Kartvelian language spoken in Turkey and the Caucasus.⁴ In the ancestor language of Laz and its close relatives, there were two sets of case marking rules for the arguments of the verb, depending on what class the verb belonged to. In the first series of rules, the subject was marked by the nominative case, and direct and indirect objects had dative cases. The second series of rules were more complex and depended on the conjugation class of the verb. In some classes, the subject received the so-called narrative case, the object got the nominative case, and the indirect object is marked with the dative. In other classes, subject is marked in the nominative and the indirect object appears in the dative. The following chart summarizes the situation:

Series I

- a. Subject is nominative.
- b. Object is dative.
- c. Indirect object is dative.

Series II

- a. Class 1
 - i. Subject is narrative.
 - ii. Direct object is nominative.
 - iii. Indirect object is dative.
- b. Class 2
 - i. Subject is nominative.
 - ii. Indirect object is dative.

The Kartvelian languages Georgian and Mingrelian preserve the set of rules intact. Laz generalized the second series to verbs that originally took the first series. This has resulted in different case marking patterns with words that otherwise are cognate. Have a look at the following parallel sentences; the first is Mingrelian and shows the old series I pattern, while the second is from Laz and shows the series II pattern which has been generalized:

| | | | |
|------------|------------------------|----------|----------------|
| Mingrelian | k'otʃi | ʔviluns | ye-s |
| | man-nominative | kill | pig-dative |
| Laz | k'otʃi-k | q'vilups | yedʒi |
| | man-narrative | kill | pig.nominative |
| | 'The man kills a pig.' | | |

12.4.3 Diffusion or Borrowing

A third factor that can influence the direction of grammatical change is *DIFFUSION*. You have already seen that languages can influence each other in their vocabulary, as words are frequently copied from one language to another. Languages do not copy just words, as they can also copy grammatical constructions and sometimes even the morphemes that are used to construct sentences in a language. This happens when there are enough people who speak two languages, and they start speaking one language using constructions that derive from the other language. In sec. 12.1, I discuss the suggestion that an original SVO word order in Austronesian languages switched in the languages of the Central and Milne Bay Provinces to SOV under the influence of the neighboring non-Austronesian languages. This means that the SOV word order in this case has diffused to the Austronesian languages. In chapter 14, I look in more detail at how languages can change grammatically as a result of diffusion.

Reading Guide Questions

1. What is the difference between a genetic grouping and a typological grouping of languages?
2. What is an isolating language?
3. What is an agglutinating language?
4. What is an inflectional language?
5. How can phonological reduction cause a language to change its grammatical typology?
6. What is morphological fusion? What sort of typological change can result from this kind of change?
7. What is morphological reduction? What kind of grammatical type results from this kind of change?
8. What is meant by the terms “ergativity” and “accusativity” with respect to language typology? How can a language change its type from one to the other?
9. How can languages change their basic word order?
10. What are verb chains? How can these develop in languages?
11. What is meant by the term “grammatical reanalysis”?
12. What is back formation?
13. How can analogy cause grammatical change?
14. What is grammaticalization?

Exercises

1. In Bislama (Vanuatu), it is possible to express contrast by shifting a noun phrase to the front of a sentence, for example:

Mi no stap slip long haos ya.
 I negative habitual live at house that
 ‘I do not live in that house.’

Haos ya mi no stap slip long hem.
 house that I negative habitual live at it
 ‘It is not that house that I live in.’

The basic word order of Bislama is SVO. How might the existence of the following sorts of variations affect the basic word order of the language in the future?

Saki i bonem haos ya.
 Saki predicate burndown house that
 ‘Saki burned down that house.’

Haos ya Saki i bonem.
 house that Saki predicate burndown
 'It was that house that Saki burned down.'

2. Many speakers of Tok Pisin (Papua New Guinea) express a relative clause by simply putting the relative clause inside the main clause without any special marking at all, except that a repeated noun phrase is expressed by means of a pronominal copy; for example:

Dispela man ol paitim em asde i dai pinis.
 that man they beatup him yesterday predicate die completeive
 'That man who they beat up yesterday has died.'

Mi no stap long ples ol paitim em long-en.
 I negative be at place they beatup him at it
 'I wasn't there where they beat him up.'

Some speakers of Tok Pisin (especially, but not exclusively people from the Highlands area) are coming to mark relative clauses by adding *longen* at the end of the relative clause; for example:

Em i bin draiv long bris i bruk longen.
 he predicate past drive over bridge predicate broken relative clause
 'He drove over the bridge that was broken.'

Mi paitem em long diwai mi holim longen.
 I beat him with stick I hold relative clause
 'I beat him with the stick which I was holding.'

How has this new function of *longen* evolved in Tok Pisin?

3. Tok Pisin has an interrogative *husat* 'who', which occurs in sentences such as the following:

Husat i kukim dispela haus?
 who predicate burndown this house
 'Who burned down this house?'

Some speakers of Tok Pisin are coming to mark relative clauses by placing *husat* in front of the relative clause (at least in written forms of the language). Here is an example of such a construction that was taken from a student's

essay in Tok Pisin for a course in linguistics at the University of Papua New Guinea:

Bai mi toktok long ol asua husat bai i kamap sapos Tok Pisin i kamap
nambawan tokples bilong Papua Niugini.

'I will discuss the problems that would arise if Tok Pisin were to become the
national language of Papua New Guinea.'

How has this construction arisen?

4. Transitive verbs in Tok Pisin carry an obligatory suffix of the form *-im* (which is illustrated in the forms *paitim* 'beat up', *kukim* 'burn down', and *holim* 'hold' in the preceding exercises). A small number of transitive verbs in Tok Pisin are exceptions in that they do not take any transitive suffix, including *save* 'know', *kaikai* 'eat', and *dring* 'drink'. Most speakers of Tok Pisin would say the following:

Yu laik dring sámpela bia?
you want drink some beer
'Would you like to drink some beer?'

But there are others who prefer to say the following to express the same
meaning:

Yu laik dringim sampela bia?

What factor would you say is responsible for bringing about the change from
dring to *dringim* in this example?

Further Reading

- Alice Harris and Lyle Campbell, *Historical Syntax in Cross-Linguistic Perspective*.
Robert J. Jeffers and Ilse Lehiste, *Principles and Methods for Historical Linguistics*,
chapter 4, "Morphological Systems and Linguistic Change," pp. 55–73; chapter
7, "Syntactic Change," pp. 107–25; chapter 8, "Lexical Change," pp. 126–37.
Mary R. Haas, *The Prehistory of Languages*, "Problems of Morphological Recon-
struction," pp. 51–58.
Theodora Bynon, *Historical Linguistics*, chapter I.8, "Morphological and Syntactic
Reconstruction," pp. 57–61; chapter I.9, "Lexical Reconstruction," pp. 61–63.
Hans Henrich Hock, *Principles of Historical Linguistics*, chapter 9, "Analogy: General
Discussion and Typology," pp. 167–209; chapter 10, "Analogy: Tendencies of
Analogical Change," pp. 210–37; chapter 11, "Analogy and Generative Gram-
mar," pp. 238–37; chapter 12, "Semantic Change," pp. 280–308; chapter 13, "Syn-
tactic Change," pp. 309–79; chapter 14, "Linguistic Contact: Lexical Borrowing,"
pp. 380–425.

- R. M. W. Dixon, *Ergativity*, chapter 7, "Language Change," pp. 182–206.
- Frantisek Lichtenberk, "Semantic Change and Heterosemy in Grammaticalisation."
- Anthony Kroch, "Syntactic Change."
- Mark Hale, "Syntactic Change."
- Ans van Kemenade and Nigel Vincent (eds.), *Parameters of Morphosyntactic Change*.
- Andrew Garrett, "The origin of NP split ergativity."
- Claire Bowern, "The diachrony of complex predication."
- David Lightfoot, *How New Languages Emerge*.

Observing Language Change

13.1 EARLY VIEWS

If you ask a speaker of any language the question, Can you think of any changes that you can see taking place in your language now?, you will likely get a positive answer. It seems that people are usually aware of some kinds of changes that are occurring in their language at any particular time. For instance, if you were to ask somebody what sorts of changes are taking place in English, you might get answers like this: The word *whom* in sentences like “This is the person whom I saw yesterday” is being replaced with *who*. If you were to ask speakers of Tok Pisin in Papua New Guinea what sorts of changes they were able to observe taking place in the language, they might comment on the fact that people are starting to say *moskito* instead of *natnat* for ‘mosquito’ and that some words are being shortened, with *mipela* ‘us’ becoming *mipla* and *bilong* ‘of’ becoming *blo*. You saw in chapter 1 that speakers of languages are often quite aware of changes that are taking place in their language and that these changes tend to be regarded as “corruptions” of the “correct” or unchanged form of the language.

Even though speakers of languages are often quite aware of changes that are taking place in their languages at a given time, it is surprising to find that, for a long time, linguists claimed that language change was something that could never be observed. Linguists claimed that all we could do was study how a language behaved before a change and after a change and then compare the two different stages of the language, but to study the change actually taking

place was impossible. They argued that language change was so slow and so gradual that the differences between the different stages of language would end up being so far apart in time that we could not hope to be alive to see a change through from beginning to end, or even to see it having any significant effect on the language. One of the most important linguists of this century after Saussure was the American linguist Leonard Bloomfield, and he stated quite unambiguously in 1933 (p. 347) that “the process of linguistic change has never been directly observed; . . . such an observation, with our present facilities, is inconceivable.” Why did linguists say this, ignoring the obvious facts around them which ordinary speakers of any language can clearly see?

As I mention in chapter 1, Saussure is regarded as the originator of modern linguistics, and one of his major achievements was to divert attention away from the purely diachronic study of language to the synchronic study of language. In chapter 14, you will learn about the Neogrammarians, who were nineteenth-century scholars who claimed that they had established linguistics as a genuine empirical science (see also sec. 9.1). By this, they claimed to have developed linguistics as a field of study that was based on the observation of physical data, with generalizations that could be tested by referring back to a different (but comparable) set of data. The Neogrammarians were able to point to earlier etymological studies and claim that there were never any scientific “checks” on the conclusions that people made. This was because there was no distinction between systematic sound correspondences and sporadic sound similarities. Only in the case of systematic sound correspondences can we claim to have made any scientifically valid generalizations.

However, Saussure reacted against the Neogrammarians by claiming that, in fact, their position was basically unscientific. He said this because it was impossible to describe the changes in a language over a period of time without first describing the language at a particular point in time. To study scientific diachronic linguistics, we must first have two synchronic descriptions of the language taken at two different times: before and after the changes that we are studying. Saussure’s *Course in General Linguistics* set out to describe the basic concepts that he felt were needed before it was possible to sit down and write scientific synchronic descriptions of languages.

Saussure proposed a rigid distinction between diachronic and synchronic methodology, and he expressed the point of view that historical information was totally irrelevant in a synchronic analysis of a language. By implication, we can assume also that Saussure would regard any guesses about the future changes a language might undergo as being irrelevant, and so they should not be included in a synchronic description of a language. In a synchronic description, all we should be interested in is describing the relations between the units in a system at a particular point in time.

This distinction between diachrony and synchrony can perhaps be compared to a movie film. A movie film is a sequence of still photographs, or frames. A description of an individual frame would be like a synchronic description of a language. But these individual frames, when they are viewed

quickly one after the other, indicate movements like real life. A study of these movements would therefore be like a diachronic study of language; to carry out such a study, you would need to compare one of the frames with another further up or down the filmstrip.

So Saussure, and the linguists who followed him in the same tradition (including Bloomfield, as you have just seen), were in a sense blinded by their own theoretical approach. They failed to see language change in progress, even though everybody else could see it. Because they did not believe that language change could be seen, they did not even look for it. There are two different headings under which we can see language change in operation: indeterminacy in language and variability in language. In this chapter, I cover both of these areas.

13.2 INDETERMINACY

To understand the concept of INDETERMINACY (OR FUZZINESS, as it is sometimes called), take a look at the following English sentences. Would you judge them to be grammatical or ungrammatical?

1. James is chopping the firewood.
2. Daffodil must sells something at the market before she goings home.
3. The dogs don't try to keep off the grass.
4. Remy isn't wanting any money from me.
5. Who isn't that?
6. I saw a man coming from the bank get robbed.
7. Who did you come to the pictures without?
8. Jennifer said she will come yesterday.
9. I doesn't goes to church at Christmas.

Some of these sentences are clearly grammatical. I am sure you could imagine an English speaker actually saying sentences such as 1, 3, and 6. Some are also clearly ungrammatical, and people who speak English know they could not say things like 2, 8, and 9. But what about sentences 4, 5, and 7? Are they grammatical or ungrammatical? They are clearly not as grammatical as 1, 3, and 6, but at the same time they are not completely ungrammatical like 2, 8, and 9. In fact, they seem to be neither one nor the other, or perhaps they are both at the same time. Thus, these sentences are indeterminate: neither clearly one nor the other.

There are many similar examples of indeterminacy in language. For instance, it is possible to derive many nouns from verbs in English by adding the suffix that we write as either *-tion* or *-ion*. So:

| Verb | Noun |
|------------|--------------|
| emancipate | emancipation |
| isolate | isolation |
| speculate | speculation |
| subject | subjection |
| connect | connection |
| delegate | delegation |

We also have the noun *aggression* in English. If somebody were to say something like the following, we would feel that, while the sentence is not really all that good, we could still imagine that somebody might actually say it:

The toddlers aggressed en masse against their teddy bears.

This sentence is also indeterminate, lying somewhere between grammatical and ungrammatical. The verb *to aggress* does not appear in the dictionary, but it is clearly not as bad as totally nonexistent verbs such as *to teapot* and *to underneath*:

*This saucepan can teapot if we're desperate.

*The dog is underneathing the house.

(Note that the asterisk * marks the sentences as being ungrammatical in these cases.)

Because there are many examples like this in languages, linguists in the past have tended to deal only with those constructions that are clearly grammatical, distinguishing them from those that are clearly ungrammatical. Linguists who view judgments on grammaticality as either yes or no situations feel that the categories of language must be viewed as absolutely "watertight." In fact, however, categories in language are often "fuzzy." Grammars are not watertight—they leak all over the place. Categories and rules are often fuzzy or indeterminate in their application. By insisting that languages consist of a number of strict and rigid either/or types of rules, such linguists have ignored a lot of what is actually going on when people use their languages.

Indeterminate or semigrammatical sentences are often evidence that grammatical change is in progress. Some people, for example, still object to the use of *access* as a verb, as in Remy accessed the internet, arguing that this can only be used as a noun. While for some people it is no doubt true that *access* is only a noun, for other people it is also possible now to use it as a verb. My suspicion is that the people who use it as both a verb and a noun are going to win, and the variability that we see between English speakers is evidence that change is now in progress.

The concept of linguistic indeterminacy also relates to the idea of the linguistic system as used by Saussure. He argued that in describing a language synchronically, we are listing the various units in the system of a language (i.e., phonemes, morphemes, words, phrases, clauses, and so on), and describing the ways in which these units interrelate (i.e., the grammatical rules for putting them together to make up larger units). In talking about describing the system of a particular language, Saussure is implying that for every language, there is one—and only one—linguistic system.

But here, too, the theoretical assumption does not always fit neatly with what happens in individual languages. There is sometimes a need to recognize that within a single language, there might be more than one system in operation, even if these systems are partially interrelated in some way. Let us look at the phonology of the Motu language of Papua New Guinea as an example. If we look at the basic vocabulary of Motu, we will find that the language has five vowel phonemes: /i/, /e/, /a/, /o/, and /u/, as well as the following set of consonant phonemes:

p t k k^w
 b d g g^w
 m n
 v ɣ h
 l
 r

Of these consonant phonemes, only /t/ has any major allophonic variation. You saw in chapter 4 of this book that we can state the distribution of the allophones of this phoneme as follows:

$$/*t/ > /t/: \begin{cases} [s] \text{ before front vowels} \\ [t] \text{ elsewhere} \end{cases}$$

The labiovelar phonemes /k^w/ and /g^w/ are treated as unit phonemes in Motu for reasons of simplicity. If we treated them as sequences of sequences of two phonemes—that is, as velar stops followed by the phoneme /w/—then we complicate our description of the phonology in two ways:

1. We have to introduce a separate phoneme /w/, which occurs only in this environment and in no other environment. This phoneme would be unlike all other phonemes in the language, which are restricted in their distribution.
2. We would have to revise our statement of the phonotactics of the language to allow consonant clusters of just this type and no other type. Otherwise, syllables in Motu would be entirely of the type CV (i.e., a single consonant followed by a vowel).

But if we include other words in the language which have been more recently introduced, we find forms such as the following:

| | |
|-----------|------------|
| tini | 'tin' |
| maketi | 'market' |
| su | 'shoe' |
| traka | 'truck' |
| hospitala | 'hospital' |

Those words violate some of the rules of Motu phonology that I have just described. The original neat complementary distribution between [t] and [s] has been destroyed, for one thing. And for another thing, the language now allows words containing consonant clusters in initial position (e.g., /tr-/) and in medial position (e.g., /-sp-/).

Linguists of Bloomfield's generation would probably have ignored these introduced words by saying that they were not really part of the language. They would probably include those introduced words that had been modified in some way in order to fit completely into the original sound pattern of Motu, so they would be happy to include words such as *makedi* 'market', which some older people actually do use instead of *maketi*. This word avoids the disruption in the complementary distribution between [t] and [s] by substituting another sound that does not undergo the same kind of allophonic variation. Linguists of Bloomfield's generation would probably also be happy to include words like *gavamani* in a dictionary of Motu as well, because the original consonant clusters of English have been totally eliminated. But, to ignore words such as *traka* and *hospitala* (or to describe only those words that fit "the system") is to ignore the way that people actually use the language. Such a description, which recognizes only a single phonemic system for the language, is clearly inadequate.¹

To describe Motu adequately, we need to recognize that there are two phonemic systems, one for original Motu words, where [t] and [s] are not phonemically distinct and there are no consonant clusters, and another for introduced words, where [t] and [s] are phonemically distinct and some consonant clusters do occur. Speakers of the language are subconsciously aware of the existence of these two different systems and could probably tell you which system a word belongs to if you asked them. There is variation between some forms (such as the variation that I have already mentioned between *maketi* and *makedi*) which is evidence of competition between the two systems. Change is clearly under way in Motu, with the original single system being supplemented by a second partial system. Some introduced words are completely adapted to the original single system, while other words belong to the more recent subsystem. Finally, some words for some speakers belong in the original system, while for other speakers they belong in the introduced system.

More and more introduced words are coming to be assigned to the introduced system by all speakers, so we can assume that when the number of such words comes to be large enough, there is a possibility that the two systems will eventually be reanalyzed as a single new system, and the original system will cease to operate. So, eventually, we could expect [t] and [s] to come to be treated as completely separate phonemes in Motu (as you saw earlier in sec. 4.3). Motu [t] and [s] are actually indeterminate in their status at the moment. In some ways we can say that they are allophones of one phoneme, while in other ways they seem to be phonemically distinct. This is therefore another example of how linguists in the past have ignored the fact that linguistic systems tend to be fuzzy.

13.3 VARIABILITY

The other important concept is the concept of VARIABILITY. Linguists traditionally believed that language was basically a yes-or-no kind of thing. While geographical varieties of languages (i.e., dialects) and social class varieties of language (i.e., sociolects) can to some extent be described as single systems of their own, independent of other systems, it is more difficult to deal with differences of style within the speech of a single individual in the same way. Probably all speakers of all languages alter their speech so that it matches the nature of the social situation they are in, even though they may not consciously be aware that they are doing it.

Linguists in the past found it difficult to describe these different styles of speech within a single fixed set of rules. For them, a rule either applied or it did not. The other possibility was that a rule could be completely optional, in which case it was entirely up to the speaker whether she or he would apply it in a given context. This led to the use of the phrase FREE VARIATION in linguistics. Two variants that were said to be in free variation were supposed to be completely equivalent in all respects, and the choice of one over the other was supposed to be completely up to the individual.

Let us examine an example of a supposedly optional rule in the grammar of English. As you know, there is a rule in English which changes sentences such as the following:

Pipira chased the boys along the beach.

into sentences like this:

The boys were chased along the beach by Pipira.

Those two sentences express the same event: that is, the same participants are involved in the same action, which takes place in the same location.

Because of this equivalence between the two sentences, linguists in the past have argued that the choice of one form over the other to refer to this event is random. In grammatical terms, therefore, the passive rule is a completely optional rule in the grammar of English.

But the choice of a passive sentence over an active sentence is not completely random, if you look at the way that people actually use the two types of sentences. Compare the following two paragraphs, which differ only in that one contains active sentences, while the other contains the corresponding passive sentences:

1. We expect children to learn to behave like adults by the time they are teenagers. We give them models of behavior to follow, and we punish them when they do not follow them. When they do things the way we like them to, we reward them.
2. Children are expected to learn to behave like adults by the time they are teenagers. They are given models of behavior to be followed, and they are punished when these are not followed. When things are done the way we like them to be done, they are rewarded.

I think that you will probably agree that paragraph 1 sounds more "conversational," while paragraph 2 sounds more "literary," even though both are saying basically the same thing.

If you examine the overall use of active and passive sentences in English, you will probably find that the active form is predominant in situations such as the following: in letters to friends, in private conversations with close relatives, in messages scribbled on the dust of somebody's car window, at home (if you speak English at home), or in a note pinned to a lecturer's door saying why you couldn't hand in your assignment on time. In contrast, the passive is probably likely to be more frequently used in situations such as these: in letters applying for jobs, in formal speeches, in public notices, by a lecturer in front of a class, and in a student's essay for a course. The difference between these two sets of situations is that the first set is considered to be more casual or informal, while the second set is considered to be more formal. When speakers feel that the situation is casual, they tend to use more active sentences, but when the situation is more formal, they tend to use a greater number of passive constructions.

It is difficult to write these kinds of facts into a grammatical rule, so linguists in the past just tended to ignore these social considerations and simply described the passive rule as "optional." One of the most influential linguists of the past few decades, Noam Chomsky, expresses this view when he said that a grammar should describe an "ideal speaker-hearer relationship," (Chomsky 1965: Ch 1) and it should ignore factors from outside language itself (such as the formality of a social situation). But language is not an ideal

system at all; in fact, it is highly variable, and much of the variability is closely tied in with social considerations.

13.3.1 Class-Based Variation

The importance of the concept of variability in language as an indication that language change is in progress was first described in detail by the American linguist William Labov (1972). He studied the way people speak in the city of New York. He clearly documented the fact that there was no such thing as a single “New York dialect,” as different people there speak differently according to their own social class background and also depending on the social context that they find themselves in while they are speaking. For instance, he found that some New Yorkers would say [k^ha:] for *car*, while others would say [k^haɪ]. He did an extensive survey to find out which people used which form, and in what kinds of situations. He came up with the results in table 13.1, according to the social background of the speaker.

These figures indicate that working-class New Yorkers have no [ɪ] in such words in 82 percent of cases, whereas upper-middle-class New Yorkers have no [ɪ] in only 30 percent of cases. The lower-middle-class speakers lie somewhere in the middle linguistically, as they “drop” their [ɪ] in 41 percent of cases. Clearly, as we go higher up the socioeconomic scale in New York City, we find that people use [ɪ] more and more.

But the story does not end there. Labov also found that the same person might use [k^ha:] sometimes and [k^haɪ] at other times. The choice between these two variants was not completely free, just as I showed you earlier that the choice between the active and the passive form of a sentence was not completely free. What Labov found was that all speakers, regardless of their social background, were likely to increase their use of [ɪ] in situations that they felt to be more formal. However, when the situation was more casual, they preferred instead to say [k^ha:]. Look at the graph in figure 13.1, which shows how people of the working class and the upper middle class increase their use of [ɪ] when the social situation increases in formality.² Something very interesting happens if we add in the figures for the lower middle class to the same graph (Figure 13.2). From this graph, you can see that when

TABLE 13.1 [ɪ] Variation by Social Class in New York City (%)

| | [ɪ] Always present | [ɪ] Sometimes present | [ɪ] Never present |
|--------------------|--------------------|-----------------------|-------------------|
| Working class | 6 | 12 | 82 |
| Lower middle class | 22 | 37 | 41 |
| Upper middle class | 24 | 46 | 30 |

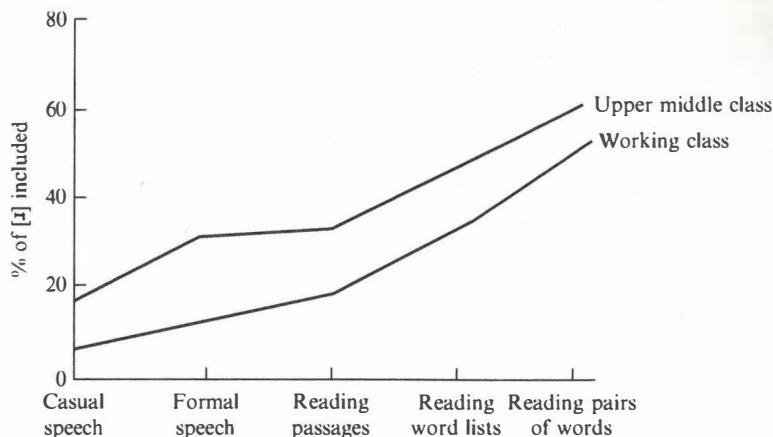


FIGURE 13.1 Percentage of [ɹ] tokens in two classes in New York City (Labov 1972)

people from the lower middle class are using their most careful form of speech, they actually put in more [ɹ] sounds than the people who are above them socially. This might seem to represent some kind of contradiction to the earlier generalization that I gave that the higher one's social class, the more likely it is that an individual will say [k^haɹ] rather than [k^ha:ɹ].

What this apparent contradiction shows is that all speakers are aware that it is more socially acceptable to use the forms with [ɹ] than the forms without it. The more careful New Yorkers are being when they are speaking, the more likely it is that they will use the [ɹ] pronunciation. It is also clear that the higher classes in society are perceived as speaking "better" English than the lower

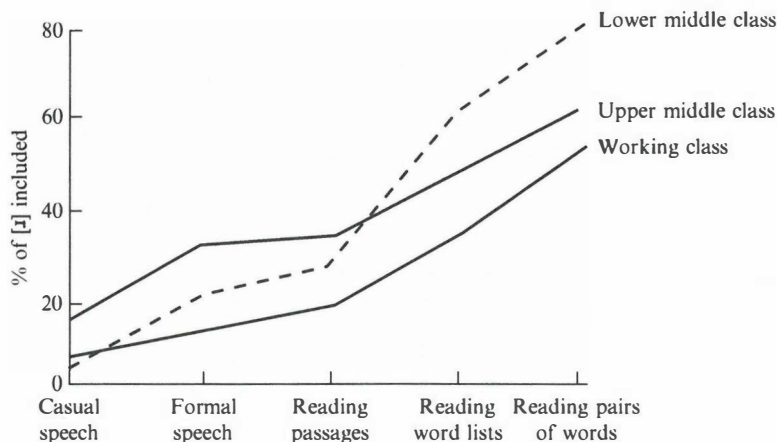


FIGURE 13.2 Percentage of [ɹ] tokens in three classes in New York City (Labov 1972)

social classes. Even though the upper middle classes do not always pronounce their [ɪ]s, the lower classes feel that their speech sounds better than their own. People from the lower social classes in stratified societies commonly try to adopt the behavior of their social superiors and try to speak like them as well. There is social prestige not only in the clothes that we wear, the cars we drive, and who we mix with but also in the way we speak.

The usual explanation for the middle-class crossover effect (as it is called) goes something like this, although the linguistic marking of social categories is complex. Some class categories are better defined than others. The category working class (or lower class), for example, has been defined by particular employment categories, education levels, income levels, and the suburb or town where someone lives. It also carries with it a certain 'anti-prestige' (that is, there are contexts where it is a positive attribute to belong to this class; the technical term in sociolinguistics is COVERT PRESTIGE). The upper middle and upper classes have overt prestige, both linguistically and materially; the category is marked by flashy cars, designer clothes, pasta machines, and trendy technological gizmos. They are often touted as modern urban society's success stories. But the lower middle class lies somewhere in between. They are not working class, but they do not have the social trappings of the upper middle class either. However, one way that social status can be marked for free is by language use. Linguistic class features (such as using or not using coda [ɪ]) become a proxy for the other absent class markers. However, in imitating upper middle class dialect features, the lower middle class end up producing more tokens than the upper middle class. This is known as the "middle class crossover effect," as shown in figure 13.2. This kind of crossover is common in studies of linguistic variation, and it shows that there is a prestigious form that speakers are consciously trying to adopt. You can probably think of examples of politicians or famous actors using "folksy" or "working class" dialect features to appear more popular; the middle class crossover effect is a similar phenomenon, except it is more systematic. It can also be an indication that linguistic change is in progress. Of course, this explanation for this type of language change only works in a stratified class-based society; not all societies have class distinctions, and explanations of language shift based on class alone ignore the fact that there are many other relevant cross-cutting social categories, including gender and race/ethnicity.

This kind of situation can lead to HYPERCORRECTION. That is, people sometimes actually use a particular linguistic variable in a place where the higher classes would never use it and where we would predict that it would not occur on purely historical grounds. For instance, in words like *father*, *Dakota*, and *data*, the lower middle class might pronounce these as if they were spelled *farther*, *Dakotar*, and *datar*, respectively, when they are trying to accommodate to higher-class members of society.

From all of these observations, it is clear that the English of New York City is in a state of change, and we are in a position to watch that change taking place. Back in the 1930s, the normal pronunciation of words like *car*

in New York was without the [ɹ]. In old movies set in New York City, the characters hardly ever pronounced their [ɹ] sounds in such words. In many words, there is instead a [j] glide, such as in *thirty*, which was pronounced as [θəɹtɪ]. (This is where the stereotype arose that New Yorkers say “Toidy-Toid Street” instead of “Thirty-Third Street.”) In the 1950s and the 1960s, there was an increase in the number of [ɹ] sounds that people used, and this led to the situation in the 1990s. Presumably, after some time, the whole city will be pronouncing the [ɹ] all the time, and change in language will then have been completed. However, while we can see that the change is taking place, and we can see the direction in which the language is headed, there is no way that we can predict how long it will take for the change to work its way right through the language.

This discussion of the role of linguistic variability and social prestige in language very neatly explains how a language like English changes. English is the language of a large-scale society that is socially stratified. Moreover, this society is one in which upward mobility is possible, so people can aspire to reach greater social heights than the level in society that they were born into. However, not all societies are like this. Some societies may be stratified, but once an individual is born into a particular place in the society, there may be no hope of moving either up or down. The caste systems of many Indian societies are examples of such rigidly stratified societies, as is the division of Tongan society into commoners, nobility, and royalty. Linguists need to do more research to find out what motivates the spread of linguistic changes in societies such as these.

13.3.2 Variation in Small Communities

Another kind of society we should consider is the small-scale society of areas like Melanesia and Aboriginal Australia. In these societies, people either know or are related to almost everybody else in the society. It is pointless to speak of upper middle class and working class in such societies, and high social status is something that is achieved by individuals who “buy” their way up the scale by killing large numbers of pigs or by demonstrating great generosity to other people in the form of presentations of goods and food. There are no privileged or underprivileged classes in these societies. Yet Melanesian languages change, just like any other language. Just as we can find synchronic evidence that change is taking place in English in New York City, we can also find synchronic evidence in the form of variability that change is taking place in the Lenakel language of the island of Tanna in Vanuatu. In this language, there is variability between the sounds [s] and [h] word-finally. While there are minimal pairs word-initially and word-medially to show that these two sounds are phonemically distinct, this distinction is being lost word-finally, and we find variation between forms such as the following:

| | | | |
|-------|---|-------|--------|
| mis | ~ | mih | 'die' |
| os | ~ | oh | 'take' |
| pugas | ~ | pugah | 'pig' |

So, what factors are involved in the spread of changes in such societies? Linguists do not yet know the answer. Perhaps what is needed is a well-trained Melanesian linguist who really understands the dynamics of a Melanesian society as an insider. Until such a person carries out some detailed studies on the dynamics of language change in Melanesian societies, our understanding of language change will be incomplete. We would expect that the mechanisms of change would work in the same ways as in other societies. After all, the same triggers of change (social differentiation, accommodation to or differentiation from other speakers and groups, misparsing, reanalysis, and so on) are a fact of language and interaction and exist whatever society we are examining. The details of the spread of change could be quite different, however. Do changes spread faster or slower in small societies? Both have been claimed. On the one hand, it has been claimed that norms are easier to reinforce in small communities, so the possibilities for drift are more constrained. On the other, small societies tend to have dense and multiplex social networks, and that facilitates the spread of changes.

13.4 THE SPREAD OF CHANGE AND LEXICAL DIFFUSION

In the preceding section, you saw how a linguistic change can spread from one small group of society so that it eventually affects the whole of society. You saw that in a socially stratified society in which there is social mobility, one force behind the spread of a change is that of social prestige.

When the Neogrammarians (about whom you learned something in chapter 9) were speaking of sound change, they claimed that these changes were conditioned by purely phonetic factors. They said that if a sound change applied in a particular phonetic environment in one word, then the same change also took place in all other phonetically comparable words at the same time. So, for instance, when final voiced stops were devoiced in German (a change that I have referred to a number of times already), the Neogrammarians would argue that all final stops underwent this change simultaneously. Now that we are in a position to observe language change taking place, we can check on the accuracy of this assumption. In fact, we can show that this view of language change is misleading in some cases. Not all sound changes work like mechanical processes, in which every word submits to an overriding rule at the same time as all other words.

For instance, the variation in Lenakel between the sounds [h] and [s] (which I described at the end of sec. 13.3) is not totally free. For one thing,

some speakers are more likely than others to use the [h] pronunciations rather than the [s] pronunciations. For another thing, the variation is more likely to occur in some words than in others, and in yet other words there is no variability at all. The less common words in the language tend not to exhibit the alternation between [s] and [h] at all, and we find only the [s] pronunciation. In words that are in more frequent daily use, the [h] pronunciation is more likely to be found. Although a change of [s] to [h] is clearly taking place in Lenakel, it is a change that is slowly creeping through the lexicon rather than affecting all words at once.

We can thus speak of LEXICAL DIFFUSION⁴ as being a major mechanism in language change, with sound changes beginning in a relatively small number of words and later spreading to other words of the same basic phonological shape; the change is completed only when it has worked through the entire lexicon. If you were to examine a language at any point from the time after a sound change has begun and before it has completely worked through the lexicon, you would probably find that it is impossible to predict precisely which words will have undergone the change and which words will have so far remained unchanged. After a change has worked itself right through the lexicon, it will look as though it affected all words of the same basic phonological shape.⁵

Grammatical change may also spread through the lexicon in a similar sort of way as phonological change. Here is an example of a language in which a grammatical change appears to be taking place, by which a plural suffix derived from the English *-(e)s* suffix is coming to be added to nouns in the Tok Pisin of some speakers of this variety of Melanesian Pidgin in Papua New Guinea. Until relatively recently, the difference between singular and plural nouns in Tok Pisin was marked by adding the plural marker *ol* before the noun phrase, as in the following examples:

| | |
|--------------------------|-----------------------|
| man | 'man' |
| ol man | 'men' |
| liklik manggi | 'small boy' |
| ol liklik manggi | 'small boys' |
| traipela banana mau | 'big ripe banana' |
| ol traipela banana mau | 'big ripe bananas' |
| dispela haus | 'this house' |
| ol dispela haus | 'these houses' |
| dispela switpela popo | 'this tasty papaya' |
| ol dispela switpela popo | 'these tasty papayas' |

There is a change that is beginning to spread in the language, by which a plural suffix derived from English *-(e)s* is coming to be used along with the older plural marker *ol*. The plural suffix that is coming to be used in Tok Pisin has the form *-s* after nouns ending in vowels or in consonants other than *-s*, while the allomorph *-is* is used after nouns ending in *-s*. This change is commonly observed in the speech of people who have been well educated in English, while less well educated rural people tend to use only the preposed *ol* plural marker. The use of the suffix *-s/-is* is most widespread among university-educated Papua New Guineans, and it is the speech of this small elite group that I now examine.

It is interesting to note that some words seem to be more likely than others to add the new plural suffix *-s/-is*. The most likely words to take this suffix are words that we can call **NONCE BORROWINGS**, or words that are copied from English on an ad hoc basis but are not fully accepted as part of the ordinary lexicon of the language. These are the sorts of words that would probably not be understood or used by less-educated speakers of the language, and they would certainly not appear in any standard dictionary of the language. So, for instance, university-educated Papua New Guineans may use learned terminology such as the following with the English-derived plural suffix:

| | |
|-------------------------------------|---------------------------------|
| ol risos-is bilong yumi | 'our resources' |
| ol politikal divlopmen-s bilong nau | 'recent political developments' |
| ol staf-s bilong yunivesiti | 'university staff' |

Among the words that are accepted as genuine Tok Pisin words, there is not nearly as much use of the plural suffix *-s/-is* as there is in nouns that are copied from English on an ad hoc basis. However, it is possible to recognize a difference in behavior between nouns that are of English origin and those that are derived from languages other than English. Nouns that have an English source can take the plural suffix *-s/-is*, whereas those that are not derived from English cannot take this suffix; for example:

| | |
|----------------------------|---------------------------------|
| ol de(-s) | 'days' |
| ol hama(-s) | 'hammers' |
| ol plaua(-s) | 'flowers' |
| ol yia(-s) | 'years' |
| ol pekato/*ol pekato-s | 'sins' (from Latin) |
| ol diwai/*ol diwai-s | 'trees' (from a local language) |
| ol pikinini/*ol pikinini-s | 'children' (from Portuguese) |
| ol kanaka/*ol kanaka-s | 'bumpkins' (from Hawaiian) |

Even within the category of English-derived words, there are still some words that appear to be more likely than others to accept the new plural suffix *-s/-is*. Words that end in vowels are more free to behave in this way than those that end in consonants, and words that end in *-s* are the least likely of all to take the plural suffix. Thus:

| | |
|---------------------------|--------------|
| ol blanket/?*ol blanket-s | 'blankets' |
| ol naip/?*ol naip-s | 'knives' |
| ol tang/?*ol tang-s | 'tanks' |
| ol pes/?*ol pes-is | 'faces' |
| ol glas/?*ol glas-is | 'glasses' |
| ol bisnis/?*ol bisnis-is | 'businesses' |

These examples show that even a grammatical change can spread through a language gradually, diffusing through some parts of the lexicon before others. A new rule can apply to just a small part of the lexicon to begin with, and it can gradually extend to other parts of the vocabulary that belong to the same grammatical category.

In the case of the change that I have just described, it will be interesting to watch for any future developments. Perhaps the plural suffix will spread to more nouns in the language. Will the social prestige of the small elite who now use this suffix cause it to spread to lower socioeconomic groups? Will the lower classes react against the tendency of the educated elite to exhibit their level of education in the way they speak Tok Pisin, preferring, instead, to maintain the original situation in which plurals were marked only by preposing *ol* before the noun phrase? If the change spreads further among the educated classes but not to the lower classes, could a genuinely diglossic situation result, in which two quite different varieties of the same language emerge, with each being used in a specific set of social contexts? Or will the educated elite succumb to the pressure of the majority of the population and simply abandon the plural suffix, with the language reverting to its original pattern? All of these different outcomes represent plausible possibilities, and there is no way that we can be certain of any particular outcome at the moment.

Reading Guide Questions

1. Why did linguists traditionally regard language change as being unobservable?
2. What is indeterminacy in language, and how is it involved in the observation of language change?
3. What is variability in language, and in what way can this be seen as evidence of language change in progress?

4. How did linguists in the past deal with indeterminacy and variability?
5. What does the lower-middle-class crossover refer to? What is the importance of this phenomenon?
6. What is hypercorrection? What does the existence of this kind of behavior say about how changes spread in languages?
7. What is the basic error in the traditional view that sound changes operate with purely phonetic conditioning factors?
8. What is meant by lexical diffusion? How does this cause problems for application of the comparative method?

Exercises

1. The vowel [ʌ] in English is normally reflected in Tok Pisin as [a], while the vowel [æ] corresponds to either [a] or [e], with some words having only [a], some alternating between either [a] or [e], and with “learned” words nearly always having only [e], for example:

| | | | |
|-------|----------|-----------|--------------|
| namba | ‘number’ | kabis | ‘cabbage’ |
| bam | ‘bump’ | blak/blek | ‘black’ |
| san | ‘sun’ | fektori | ‘factory’ |
| tagj | ‘tank’ | menesmen | ‘management’ |
| man | ‘man’ | | |

Imagine a primary school-educated person who speaks Tok Pisin and a little English talking to somebody with a university education in Tok Pisin. Although the less-educated person knows that Tok Pisin has the expression /graun malmal/ (literally: ‘soft ground’) to refer to ‘mud’, she prefers to use the English word. But instead of pronouncing it in Tok Pisin /mat/, as you might predict, she pronounces it as /met/. What do you think is going on here?

2. In English, voiceless stops are generally aspirated, except in word-final position, after /s/ and before unstressed vowels, where they are unaspirated. The vowel /æ/ also tends to be phonetically quite long when there is a following voiced sound. Thus:

| | | |
|---------|-----------------------|----------|
| /stɒp/ | [stɒp] | ‘stop’ |
| /stæmp/ | [stæːmp] | ‘stamp’ |
| /bæd/ | [bæːd] | ‘bad’ |
| /ænd/ | [æːnd] | ‘and’ |
| /hæpi/ | [hæpi] | ‘happy’ |
| /pɹɪti/ | [p ^h ɹɪti] | ‘pretty’ |

In Papua New Guinea English, voiceless stops are generally unaspirated and vowels are generally short (with many distinctions of phonemic length being

lost; for example, *ship* and *sheep* are both pronounced [ʃip]). Some Papua New Guinean speakers of English, especially educated women, tend to produce forms such as the following, which do not normally occur in either Papua New Guinea English or in native-speaker varieties of English (such as Australian English):

- [st^hɒ:p^h] 'stop'
 [st^hæ:m p^h] 'stamp'
 [p^hji:t^hi] 'pretty'
 [hæ:p^hi] 'happy'

What is going on here?

3. An Australian is likely to pronounce the word *dance* as [dæns] and *transport* as [træns pɒt], respectively. Imagine yourself to be an Australian speaking before a New Zealand audience that likes to ridicule people with recognizably Australian accents, and you find yourself saying [ʌndəstænd] for *understand* instead of [ʌndəstænd], even though Australians and New Zealanders both say [ʌndəstænd]. Why might this happen?

Further Reading

- William Labov, *Sociolinguistic Patterns*, chapter 1, "The Social Motivation of a Sound Change," pp. 1–42.
 Jean Aitchison, *Language Change*, chapter 3, "Charting the Changes," pp. 47–60; chapter 4, "Spreading the Word," pp. 63–76; chapter 5, "Conflicting Loyalties," pp. 77–88; chapter 6, "Catching On and Taking Off," pp. 89–107.
 William Labov, *Principles of Linguistic Change*, vols. 1 and 2; especially volume 1, Part A, "Introduction and Methodology," pp. 7–112.
 Penelope Eckert and John Rickford, *Style and Sociolinguistic Variation*.
 Suzanne Romaine, *Language in Society*. This book covers in much more detail topics mentioned only briefly here. Chapter 5, "Language Change in Social Perspective," pp. 135–166, is most relevant to historical linguistics.

Language Contact

There are many bilingual and multilingual societies in the world. Among countries, Canada is officially bilingual, with both English and French functioning at the national level. Switzerland is officially quadrilingual, with German, French, Italian, and Romansh as official languages. Other nations are more complex in their linguistic makeup, such as the former Soviet Union, India, and Indonesia, where there are hundreds of separate languages spoken (although, of course, not all these languages have official status).¹ The most linguistically complex nations in the world are the small Melanesian countries. Papua New Guinea boasts over 800 distinct languages, spoken by a population of about 3.5 million people. Nearby Vanuatu has only 100 or so languages, but its population is much smaller, with the total number of people scarcely reaching 140,000. Some 90 percent of the world's languages are spoken by about 10 percent of its population.

Just because a society is multilingual or bilingual does not necessarily mean that there is a great deal of language contact, however. While Belgium recognizes both Flemish and French as official languages, there is relatively little language contact, as 85 percent of the population is monolingual in either Flemish or French and does not speak the language of the other group.

In world terms, monolingualism is relatively rare. This may come as a surprise to some people, especially to people from Western industrialized societies, particularly English-speaking societies such as the United States, the

United Kingdom, and Australia. There is a standard joke among migrants to Australia that goes like this:²

- Q. What is a person who speaks three languages?
 A. Trilingual.
 Q. What is a person who speaks two languages?
 A. Bilingual.
 Q. What is a person who speaks one language?
 A. Australian.

People from Vanuatu generally speak two, three, four, and sometimes even more languages fluently, and they often find it incomprehensible that the average Anglo-Celtic Australian or Pākehā New Zealander speaks only English. In this chapter, I will explore some of the linguistic consequences of language contact in societies such as those of Melanesia and elsewhere where multilingualism is a fact of everyday life.

Up to now in this volume, I have frequently referred in passing to the results of language contact, though this has almost always involved discussion of language change that has involved lexical change as a result of new words being copied into the lexicon from other languages. In this chapter, however, I look not so much at how languages can influence each other lexically but at how the whole phonological or grammatical system of a language can be influenced by that of another language.

14.1 CONVERGENCE

When you hear somebody speaking and their first language is not English, it is generally easy to recognize that he or she is not a native speaker of English. There are usually a number of telltale signs that indicate not only that the person is not a native speaker of English but also what that person's first language might be. By this I mean that it is often possible to recognize from the way somebody speaks English whether he or she is a speaker of French, German, Italian, Chinese, Japanese, Russian, or whatever other language. Typically, people carry over features from their first language into another language that they learn later in life, and we hear this at the phonological level as a foreign accent and at the grammatical level as learner errors. However, it is not just among people who are learning a second language that one language can influence another. Even among people who can be considered to be fluently bilingual—that is, people who have been speaking two languages regularly and fluently from early childhood—we find that features of one language can cross over into the way that person uses the other language. The influence of one of the linguistic systems of an individual

on the other linguistic system of that individual is referred to in general as INTERFERENCE.

Interference can occur in the phonological system of a language, in its semantics, or in its grammar. Phonological interference simply means the carrying over of the phonological features of one language into the other as an accent of some kind. This might involve the incorrect transfer of the distribution of the allophones of a particular phoneme into the other language in such a way that the phonological system of that language is violated. For example, the English of a Japanese-English bilingual who says *rots of ruck* instead of *lots of luck* has been influenced by interference from the fact that in Japanese there is no phonemic contrast between /l/ and /r/ as there is in English.

To illustrate grammatical interference, examine the following sentence, which contains a relative clause. Sentences such as these are often produced by schoolchildren in Vanuatu who are learning English:

This is the book which I read it yesterday.

To a native speaker of English, this sentence contains an obvious error—namely, the use of the pronoun *it* after the verb *read* in the relative clause. English grammar contains a general rule that deletes any reference to noun phrases in a relative clause that have already been mentioned in the sentence. In just about all relative clauses, it's not grammatical—according to the rules of English grammar—to leave a pronoun there. Instead, the English sentence is:

This is the book which I read yesterday.

However, relative clauses in the first languages of children in Vanuatu schools typically require that the noun phrase be mentioned again in sentences such as these by means of some kind of pronominal copy after the verb. To illustrate this kind of construction, I give an example from one of these languages, Paamese:

Tu:s keke na-les-i naŋaneh keiek.
 book which I-read-it yesterday this
 'This is the book which I read yesterday.'

In this example, you can see that in Paamese it is necessary to include an object pronoun referring to the book after the verb (in the form of the pronominal suffix /-i/). A speaker of Paamese who fails to delete the pronoun in sentences such as these in English is engaging in grammatical interference from his or her first language.

We also find cases where the meaning of words have been transferred from one language to another. SEMANTIC INTERFERENCE can also be referred

to as "semantic copying," as "loan translation," or as "calquing." A CALQUE is when we do not copy a lexical item as such from one language into another, but just the meanings are transferred, while at the same time we use the corresponding forms of the original language. The term *hot dog* as a name for a kind of fast food originated in English, but in French in Québec the same thing is referred to as a *chien chaud*. *Chien*, of course, is the French word for 'dog' and *chaud* is the word for 'hot'. Thus, we can say that *chien chaud* is a calque based on English 'hot dog'. French speakers speaking English occasionally say that they will 'toast a CD' rather than 'burn' it; this is another calque.

As I said at the beginning of this chapter, I do not plan to enter into a great deal of discussion about lexical interference (or lexical copying, or borrowing) between languages, as this is covered elsewhere in this volume (most notably in sec. 11 and sec. 8.2). However, I would like to mention at this point that, while the introduction of lexical items from one language into another does not necessarily affect the structure of the language that is receiving the new material, it is also possible that introduced lexical items can affect the phonology and the grammatical system. In chapter 4, I show how words originating from English that have been introduced into the Motu language of Papua New Guinea now show signs of disrupting the previous complementary distribution between [t] and [s] and are, in fact, causing a phonemic split to take place in the modern language. It is also possible for completely new sounds to be introduced into a language via words copied from other languages. Bahasa Indonesia originally had no voiced velar fricative at all, either as a separate phoneme or as an allophone of some other phoneme. However, with the introduction of words of Arabic origin into the everyday vocabulary, we can now show evidence of phonemic contrast between /g/ and /ɣ/ in this language.

It is also possible for words from other languages to introduce new grammatical patterns into a language. To a very minor extent this has happened in English, as some words of foreign origin have kept their original plurals. Table 14.1 gives some examples. It is rare for bound morphemes to be incorporated into the general grammar of another language, so it is unlikely that any of these patterns for the formation of plurals will spread beyond the words that originally introduced the patterns in the first place. In fact, most nouns of foreign origin are quickly adapted to the rules of the language anyway. So, the plural of *atlas* in English is now *atlases*, not *atlantes* as we might have expected on the basis of the morphological behavior of the word in its original Greek. Most of the words in Table 14.1 also have variant plurals with the regular English plural marker, such as *indexes* and *cactuses*.

While the example that I just gave involved the influence of one language on another in the area of morphology, it is possible for lexical copying to influence higher levels of grammar as well. In Paamese, all verbs are required to carry prefixes that indicate the pronominal category of the subject, as well

TABLE 14.1 Words of Foreign Origin in English with Irregular Plurals

| | Singular | Plural |
|---------|--------------------------|--------------------------|
| Greek | phenomenon criterion | phenomena criteria |
| Latin | datum index cactus | data indices cacti |
| Italian | lingua franca | lingue franche |
| Hebrew | kibbutz | kibbutzim |

as a variety of tense and mood categories. So, from the root *loh* 'run' (which cannot occur without any prefixes), we can derive the following inflected forms (among many others):

naloh 'I ran'
 ni-loh 'I will run'
 ko-loh 'you ran'
 ki-loh 'you will run'
 a-loh 'they ran'

However, verbs such as *sta:t* 'start', *ra:u* 'argue' (from 'row'), and *ri:t* 'read' that are borrowed from English are not permitted to carry any prefixes, and so a new grammatical construction evolved just to handle these new forms. There is a verb of the form *vi:* in Paamese which functions as a copula in sentences such as the following:

Inau na-*vi:* meahos.
 I I-am man
 'I am a man.'

The only kinds of words that could originally follow the verb *vi:* in Paamese were nouns in equational sentences such as the above. However, in the modern language, verbs introduced from English have also been incorporated into the same grammatical construction, and the prefixes that ordinarily would have been attached directly to the verb root are now attached instead after the preceding copula, as in the following examples:

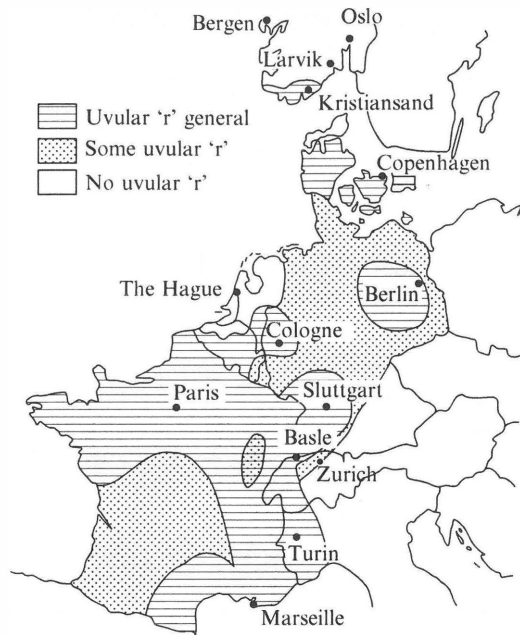
na-*vi:* sta:t 'I started'
 ko-*vi:* ra:u 'you argued'
 a-*vi:* ri:t 'they read'

In these examples, while a new pattern in Paamese grammar has emerged as a result of new words coming into the language, this pattern has not come from English. It is in fact a brand new pattern that has emerged out of the existing structural resources of Paamese as a way of coping with introduced vocabulary that speakers somehow felt did not "fit" the language properly.³

It is absolutely clear that languages can influence each other lexically (and, through lexical introductions, also to some extent grammatically), and it is just as clear that a speaker's first language can influence the way he or she speaks another language at all levels of language (i.e., in the phonology, the grammar, and the semantic system). However, there has been considerable debate in recent years on the question of whether one language as a whole can really influence another language as a whole (as against individual speakers of the language). Some have argued that only words can be borrowed and that apparent influences in grammar and morphology actually follow only from the borrowing of lexical items. Others point to examples where there is influence in the absence of lexical borrowing.⁴

A significant body of literature on the subject of linguistic diffusion and convergence is based on the assumption that languages can and do influence each other in patterns, as well as in vocabulary. The term *DIFFUSION* is used to refer to the spread of a particular linguistic feature from one language to another (or, indeed, to several other languages). One example of diffusion that is often referred to is the spread of the uvular [ʁ] in the languages of Europe. This is the kind of sound that you are taught to produce when you are learning to pronounce French words such as *rare* 'rare', *rire* 'laugh', and so on. Originally, these words were pronounced in French with an alveolar trill, and this is preserved today in languages like Italian. However, it appears that in the 1600s, speakers of French in Paris began to pronounce their 'r' sounds as uvulars rather than as alveolars. This change then spread to other language areas in Europe, and people in Copenhagen (in Denmark) were apparently doing the same thing in Danish by about 1780. The uvular pronunciation of r is now common in French, German, and Danish, and it is also used in some areas where Dutch, Norwegian, and Swedish are spoken. Map 14.1 suggests that the uvular r has hopped from city to city and then radiated out from the cities to the surrounding rural areas.

Such examples are not isolated. Albanian, Bulgarian, Romanian, and Greek, all spoken in the Balkans area of Europe, are only fairly distantly related to each other within the Indo-European language family. However, these languages share certain grammatical features that do not appear to be derived from their respective protolanguages. One of these features is the use of a special complex sentence construction instead of the infinitive construction to express meanings such as 'I want to leave'. All of these languages express this meaning instead by a construction that translates literally as something like 'I want that I should leave'. The following examples show that while the words used to express this meaning are quite different in these



MAP 14.1 Spread of uvular 'r' in France

four distantly related languages, the grammatical construction is basically the same:

| | | | |
|-----------|--------------------|-----------|----------------|
| Albanian | <i>Due</i> | <i>te</i> | <i>shkue.</i> |
| Bulgarian | <i>Iskam</i> | <i>da</i> | <i>otida.</i> |
| Romanian | <i>Veau</i> | <i>sa</i> | <i>plec.</i> |
| Greek | <i>Thelo</i> | <i>na</i> | <i>pao.</i> |
| | I-want | that | I-should-leave |
| | 'I want to leave.' | | |

This similarity between these four languages is not something that we would have predicted from Proto-Indo-European, and the suggestion is that these four languages have *converged*, or come to resemble each other structurally as a result of a long period of linguistic contact and mutual interference.

Languages that have come to resemble each other as a result of linguistic convergence in this way are said to belong to LINGUISTIC AREAS, and the features that have diffused among the languages that belong to such an area are called AREAL FEATURES. Thus, in the case of the languages that I have just described, we could refer to the Balkans as a linguistic area (or sometimes as a SPRACHBUND, to use a word of German origin), and the special construction

that I illustrated would be called an areal feature. Linguistic areas can be recognized in a number of different parts of the world. Chinese, Thai, and Vietnamese all belong to a linguistic area, as all have developed phonemic tone distinctions. The Indo-European and the Dravidian languages of the Indian subcontinent have developed widespread retroflex consonants, which set them apart as a linguistic area, and a number of Bantu languages and Kalahari languages in southern Africa also constitute a linguistic area characterized by the presence of rather unusual click consonants.

A linguistic area can be characterized by shared phonological features, as well as grammatical features, as illustrated by the example given above of the construction in the Balkans linguistic area. In sec. 12.3, I refer to the possibility that SOV word order has diffused from some non-Austronesian languages in Central Province in Papua New Guinea to the Austronesian languages, resulting in a linguistic area characterized by SOV syntax. Some scholars who have described both the Austronesian and the non-Austronesian languages of parts of the West New Britain province of Papua New Guinea have argued that syntactic convergence among these languages has been even more thorough than this, involving quite a number of different syntactic constructions. For many sentences, it seems that speakers of a number of different Austronesian and non-Austronesian languages in this area map their own words onto grammatical constructions that are almost identical. In fact, the same constructions are also found in Tok Pisin, even though this language is lexically derived mostly from English:

Non-Austronesian

Anêm *Ezim* *o-mên* *da-kîn*

Austronesian

Mouk *Eliep* *max* *na-nas*
 Aria *Bile* *me* *ne-nenes*
 Tourai *Bile* *me* *na-nes*
 Lamogai *Bile* *me* *ne-nes*
 Lusi *Vua* *i-nama* *na-sono*
 Kove *Vua* *i-nama* *na-sono*
 Kabana *Bua* *i-nam* *na-sono*
 Kilenge *Vua* *i-mai* *na-sono*
 Amara *Eilep* *i-me* *a-nas*
 Tok Pisin *Buai* *i kam* *mi kaikai*

betel nut it-come I-chew

'Hand me some betel nut to chew.'

The diffusion of grammatical features in this way has caused some linguists to question further the validity and basic assumptions of the whole comparative method. Some languages appear to have undergone so much diffusion in the lexicon and the grammar that it can be difficult to decide which protolanguage they are derived from. According to the comparative method as I have described it in this volume, it is possible for a language to be derived from only a single protolanguage, yet some linguists have found it necessary to speak of mixed languages, which seem to derive from two different protolanguages at once.

Linguists tend to be thankful that such cases appear to be fairly rare. However, where such languages exist, they often produce much heated discussion as different scholars come down in support of undeniable membership in one language family or another, and yet others argue that such either/or conclusions do not accurately reflect the genuinely indeterminate nature of the language. One example of such a situation involves the languages of the Reef-Santa Cruz islands in the Solomon Islands of Melanesia, where there has been debate as to whether these are basically Austronesian languages that have been heavily influenced by non-Austronesian languages, or whether they are non-Austronesian languages that have been heavily influenced by Austronesian languages.

Despite the fact that areal studies of languages frequently refer to linguistic convergence, and scholars often speak of the "borrowing" of features at all levels of language, some linguists are reluctant to accept the possibility of syntactic copying between languages. While accepting the obvious fact that lexical copying occurs, as well as the possibility that individual words can bring certain morphological characteristics with them into another language, some linguists argue that grammatical patterns as such cannot be copied or, if they are, that this happens only in the rarest of circumstances. Facts which are often quoted as evidence of syntactic copying, these scholars argue, often turn out to have quite different explanations.

For instance, it is often stated that Québec French is changing not only lexically but also syntactically in the direction of the dominant English language, and this tendency is widely condemned by purist Québécois. In English, it is possible to end a sentence with a preposition (despite the claims of the prescriptive grammarians among us), as in the following:

That's the girl I go out with.

French differs from English in that it is not possible to end a sentence with the corresponding preposition *avec* 'with', and in order to express the same meaning, the sentence has a different word order. The sentence with word order corresponding to the English one is ungrammatical:

C'est la fille avec qui je sors.

that-is the girl with who I go-out

'That's the girl with whom I go out.'

*C'est la fille qui je sors avec.
that-is the girl who I go-out with

In the French that is spoken in Québec, however, sentences of the following type, which closely parallel the English construction, are frequently heard:

C'est la fille que je sors avec.
that-is the girl that I go-out with
'That's the girl I go out with.'

Despite the close structural similarity between the English and the French patterns in those examples, we cannot assume that, merely because there are structural similarities between the two languages, one is necessarily derived from the other. Historical research reveals that, in fact, there is written evidence of the stranding of *avec* without a following pronoun in French going back about 600 years (which was well before French and English came into contact in Québec). The same pattern is apparently still preserved in some French dialects in France that have not been in contact with English, and even in some other Romance languages, which suggests that the pattern goes back even further in time. Another point to consider is that, while we can strand any preposition in English without a following pronoun, this is possible in French only with the longer prepositions. With very short prepositions such as *à* 'to', this construction never occurs. So, note that the following is not possible in French:

*C'est la fille que j'ai parlé à.
that-is the girl that I-past speak to
'That's the girl I spoke to.'

14.2 LANGUAGE GENESIS: PIDGINS AND CREOLES

14.2.1 Some Definitions

According to the model of language change that I present in this volume, every language is derived as a result of (more or less gradual) change from a single language that was spoken in the past. However, there is one category of languages that appears to have evolved under rather special circumstances—the languages known as pidgin languages and creole languages. When speakers of several different languages come into contact in a situation where there is an urgent need to communicate and there is little social opportunity to learn whatever happens to be the dominant language, and where no other language

predominates in terms of numbers of speakers, what often happens is that a PIDGIN language develops. The pidgin that forms has a vocabulary that derives largely from the dominant language, but the vocabulary is very much reduced in size. The grammar of a pidgin language is radically different from that of the dominant language and typically involves much greater regularity than the grammar of the dominant language, as well as less redundancy. A pidgin language also tends to have only free morphemes with very few bound morphemes. In addition to these purely linguistic features, a pidgin language is used only as a second language by all of its speakers.

Pidgin languages have evolved frequently and in many different parts of the world when the contact circumstances have been ripe for their formation. When Melanesian laborers were taken by English-speaking Europeans from what are now Vanuatu, Solomon Islands, and Papua New Guinea in the nineteenth century to work on sugarcane plantations in Queensland and Samoa, the circumstances for the formation of a pidgin based on English vocabulary were ideal. There were speakers of large numbers of different languages working together under European overseers. Very rapidly a new language came into existence.

A CREOLE language has some things in common with pidgin, but in other ways creoles are quite different. Like pidgins, creoles emerge in situations where speakers of different languages are in contact. Like pidgins, they tend to have a majority of their vocabulary from one language, and they tend to have little bound morphology. However, unlike pidgins, creoles are not only used as second languages; they are also first languages with the full communicative function and lexicon of any other language that has first-language speakers.⁵

Confusingly, some languages that began as trade pidgins are now creoles but are still called “pidgin English”. One such language is Melanesian Pidgin. This language is still spoken in slightly different forms in Papua New Guinea (where it is known as Tok Pisin), Solomon Islands (where it is known as Pijin), and Vanuatu (where it is known as Bislama). Although between 80 and 90 percent of the vocabulary is derived from English, there is also a sizeable proportion of words that come from a variety of different local languages. Some words of German origin have also found their way into Tok Pisin, while a significant number of words of French origin are found in Bislama. A fluent speaker of Melanesian Pidgin (which is how we can refer generically to these three dialects) cannot be understood by someone who speaks only English, and Melanesians who speak their variety of Pidgin cannot understand speakers of English unless they learn it in school. By all criteria, therefore, Melanesian Pidgin is a new and distinct language with its own phonology, grammar, and lexicon.

14.2.2 Case Study 1: Tok Pisin

As an illustration of what a pidgin language is like, I refer here to Tok Pisin. As I have already indicated, the vocabulary of this language is largely of

English origin, in this case about 80 percent, though the words have been phonologically restructured to fit Melanesian sound systems. Here are some examples:

| | |
|---------|------------|
| dok | 'dog' |
| aus | 'house' |
| rot | 'road' |
| ren | 'rain' |
| trausis | 'trousers' |

Of the remaining 20 percent of the lexicon, most comes from the languages of the New Britain and New Ireland people who were the original laborers on the Samoan plantations. So, we find words such as the following:

| | |
|---------|--------------------|
| kakaruk | 'chicken' |
| kiau | 'egg' |
| buai | 'betel nut' |
| kunai | 'long grass' |
| kulau | 'drinking coconut' |

The small number of remaining words in Tok Pisin do not come from English or from local languages, but from a variety of other sources. Such words include the following:

| | | |
|----------|------------|---|
| rausim | 'take out' | From German <i>heraus</i> 'get out' |
| beten | 'pray' | From German <i>beten</i> 'pray' |
| pater | 'priest' | From Latin <i>pater</i> 'father' |
| binatag | 'insect' | From Malay <i>binatang</i> 'animal' |
| pikinini | 'child' | From Portuguese <i>pequenho</i> 'small' |
| kanaka | 'bumpkin' | From Hawaiian <i>kanaka</i> 'man' |
| kaikai | 'eat' | From Māori (or other Polynesian) <i>kai</i> 'eat' |

The vocabulary of Tok Pisin is also clearly "reduced" with respect to that of English, as well as that of Melanesian languages. This language lacks the vocabulary that we have in English to discuss many concepts in law, science, and technology, and it also lacks much of the vocabulary that is present in Melanesian languages to name different parts of the natural environment, especially some of the rarer flora and fauna, as well as cultural practices.

Grammatically, if you compare Tok Pisin with English, you will find that Tok Pisin is simpler in its structure, in that it is much more regular. For example, while English has many unpredictable past tense forms for verbs,

Tok Pisin verbs are the same in all their forms. So, while in English we have to learn the past-tense forms of the following verbs separately, verbs in Tok Pisin exist in only a single invariant form:

| Present | Past |
|---------|---------|
| bring | brought |
| ring | rang |
| string | strung |
| ping | pinged |

Differences in tense and aspect, which are sometimes marked in English by suffixes to the verb, are marked in Tok Pisin by independent grammatical words, for example:

Em i toktok.
 he/she predicate talk
 'He/she talks.'

m i bin toktok.
 he/she predicate past talk
 'He/she talked.'

Tok Pisin grammar also differs from that of English in that it has far less redundancy built into its grammatical system. For example, in English, plural marking is expressed in a variety of different ways in a sentence, often in more than one way at once. For instance, it can be marked in the following ways:

1. By a separate form of the noun: i.e., *dog* versus *dogs*, *child* versus *children*, *man* versus *men*, *woman* versus *women*
2. By a difference in the form of a preceding demonstrative: i.e., *this* versus *these*, *that* versus *those*
3. By a separate form of the verb: i.e., *am* versus *are*, *is* versus *are*, *does* versus *do*

So, in the following sentence, the idea of plural is expressed in three separate places, as shown by the contrasting singular form:

Those women are singing.
 This woman is singing.

In Tok Pisin, however, the idea of plural is expressed only once in the sentence, and even then it is optional. We can say the following to refer to one woman or to many women:

Dispela meri i singsing i stap.
 this/these woman/women predicate sing predicate continuous
 'This/these woman/women is/are singing.'

If you specifically want to mark the fact that there is more than one woman involved, you can use the plural marker *ol* at the front of the noun phrase, but you will note that none of the other words in the sentence are marked in any way:

Ol dispela meri i singsing i stap.
 plural these women predicate sing predicate continuous
 'These women are singing.'

14.2.3 Case Study 2: Motu

Pidgin languages can be formed in any situation where the contact circumstances are right. There are pidgin languages in which the lexicon is derived predominantly from Spanish, French, Portuguese, and Dutch in various parts of the world. It is not necessary that the lexicon of a pidgin should be derived only from European languages, as there also are cases where pidgins have been formed out of non-European languages. In the Pacific, for instance, we find Hiri Motu, which is widely spoken in Papua today, and this language is based on the vocabulary of the vernacular Motu language of the Port Moresby area. When outside laborers were introduced into Fiji, the resultant pidgin was not based on the vocabulary of English but that of Fijian. In Australia, there was a pidgin based on the Ngarluma language of western Australia.

I mentioned at the beginning of this section that pidgin and creole languages tend to avoid bound morphemes, but the Tok Pisin examples do not illustrate this very well because English is a language that has relatively few prefixes and suffixes, at least when compared with many other languages of the world. To illustrate this point, and also to illustrate what a pidgin that is derived from a non-Indo-European language looks like, I now give some examples from Hiri Motu and compare them with the vernacular Motu from which it is lexically derived. Some of the differences between these two languages involve the following points:

1. Objects to verbs in vernacular Motu are expressed as suffixes to the verb, and these have the following shapes:

| | Singular | | Plural |
|--------|----------|-----------|--------|
| First | -gu | Inclusive | -da |
| | | Exclusive | -mai |
| Second | -mu | | -mui |
| Third | -(i)a | | -dia |

In pidgin Motu (or Hiri Motu, as it's also called⁶), objects are expressed by full-form pronouns that have the same form as the subject pronouns. The grammatical difference between subject and object is shown by the position of the form in the sentence. The full-form pronouns are the same in both vernacular and pidgin Motu:

| | Singular | | Plural |
|--------|----------|-----------|--------|
| First | lau | Inclusive | ita |
| | | Exclusive | ai |
| Second | oi | | umui |
| Third | ia | | idia |

2. Subjects to verbs are marked in vernacular Motu as prefixes to the verb, and the forms of these prefixes are as follows:

| | Singular | | Plural |
|--------|----------|-----------|--------|
| First | na- | Inclusive | ta- |
| | | Exclusive | a- |
| Second | o- | | o- |
| Third | e- | | e- |

In pidgin Motu, subjects are expressed by placing the full pronoun in the subject position of the sentence, and there is no further subject marking on the verb.

3. To make a verb negative in vernacular Motu, there is a different set of subject markers from those that are used in the affirmative, as given above. The negative prefixes are as follows:

| | Singular | | Plural |
|--------|----------|-----------|--------|
| First | asina- | Inclusive | asita- |
| | | Exclusive | asia- |
| Second | to- | | asio- |
| Third | se- | | asie- |

In pidgin Motu, negation is marked by placing the free form /lasi/ after the verb phrase. The word /lasi/ also occurs in vernacular Motu, where it is a word meaning 'no'.

The following examples are presented to show the difference between vernacular Motu and pidgin Motu. The two languages are not mutually intelligible, even though most of the words that occur in pidgin Motu are derived directly from roots that are used in vernacular Motu:

Vernacular Motu

Ia e-ita-mu.
 he/she he/she-see-you
 'He/she saw you.'

Asi-na-rakatani-mu.
 not-I-leave-you
 'I didn't leave you.'

Pidgin Motu

Oi ia itaia.
 you he/she see
 'He/she saw you.'

Oi lau rakatania lasi.
 you I leave not
 'I didn't leave you.'

14.2.4 Research on Pidgins and Creoles

Linguists have drawn a distinction in the past between pidgins and creoles because they have argued that there are structural differences between the two. Being only a language used in sporadic contact, a pidgin has generally been seen as a very basic sort of language indeed, with the smallest possible lexicon, as well as a rudimentary grammar. However, once a pidgin becomes the mother tongue of a community, it is generally assumed that it undergoes rapid lexical and structural expansion in order to meet the normal needs of a community of native speakers.⁷

Pidgin and creole languages have aroused a great deal of interest because linguists are keen to find out how these languages acquire their structures. You may have noticed that up to this point I have spoken about pidgins and creoles having a predominantly English (or French, or Spanish, or Motu) vocabulary, yet they are still mutually unintelligible with the languages from

which their vocabularies are derived. This suggests that pidgins and creoles are structurally very different from their lexifier languages (i.e., the languages from which their vocabularies are derived), and this is a point that I think you will appreciate from the examples of the Tok Pisin structure that I presented earlier in this section (as well as in earlier chapters of this volume).

Many linguists have been struck by the fact that pidgin and creole languages often show strong parallels in their structure with their substrate languages rather than their superstrate languages. The term *SUPERSTRATE* (or *superordinate language*) is used to refer to the dominant language in the contact situations in which a pidgin or creole language develops. In the case of Tok Pisin, for example, English is clearly the superstrate language. The *SUBSTRATE*, in contrast, refers to the vernaculars of the people who actually develop a pidgin or creole. In the case of Tok Pisin, the substrate languages would be the various vernaculars of the New Britain and New Ireland laborers who were originally taken to work in Samoa and Queensland in the nineteenth century. While the grammar of Tok Pisin is clearly different from that of English, it seems that when we examine many of the points of difference between English and Tok Pisin, we can find structural parallels with the substrate languages. For instance, the form *i* that occurs in the examples above as a predicate marker corresponds roughly in shape and in function to a morpheme *i* that is found in Tolai (and many other of the substrate languages), for example:

Tolai

To Pipira i vana.

article Pipira predicate go

'Pipira is going.'

Tok Pisin

Pipira i go.

Pipira predicate go

'Pipira is going.'

The existence of two separate forms of the first person nonsingular pronoun in Melanesian vernaculars is also paralleled in the structure of the Melanesian Pidgin pronoun system but not in that of English. In Tok Pisin, there are two separate pronouns corresponding to the single form 'we' in English. First, there is *yumi*, which means 'we' when you are including the person you are speaking to (i.e., the so-called inclusive pronoun). Second, there is the form *mipela*, which means 'we' when you are excluding the person you are speaking to (i.e., the so-called exclusive pronoun). This distinction is widespread in the substrate languages for Melanesian Pidgin, but English

grammar does not make the distinction (and sometimes English speakers even find it hard to use the pronouns *yumi* and *mipela* correctly).

The existence of such structural parallels between pidgins and creoles and their substrate languages has led many scholars to argue that pidgins and creoles are mixed languages in the sense that they derive their lexicons from the superstrate, while their grammars come predominantly from the substrate. If this interpretation is correct, then pidgin and creole languages differ dramatically in their genesis from other languages as they have multiple ancestors rather than a single ancestor. According to such a view, it would be impossible to classify Tok Pisin either as an Austronesian language or as an Indo-European language because it contains significant elements from both language families. (I referred to mixed languages in sec. 14.1.)

Remember from the preceding section that some scholars today do not accept that languages can easily influence each other structurally. Linguists who hold this point of view sometimes extend this even to pidgin and creole languages, arguing that the existence of parallels in structure between pidgins and creoles and their substrate languages is not necessarily evidence that a pidgin has been structurally influenced by the substrate, and they argue that other factors may also be involved. For instance, it could be equally argued that the 'predicate marker' *i* that I described earlier in Tok Pisin does not derive from the substrate at all, but from the English pronoun *he*, which may have been repeated after the subject noun phrase. Thus, *?ipira i go* 'Pipira is going' is not necessarily derived from the Tolai construction but from a pidgin "broken English" sentence of the form *?ipira he goes* (and sentences of this type do sometimes occur when people are learning English as a second language).

Scholars who deny any significant effect of substrate structural patterns in the development of a pidgin or a creole language tend to point, instead, to what they see as the remarkable structural similarities between pidgin and creole languages that have radically different histories and even different lexical source languages. For instance, if you compare the grammatical structure of a simple intransitive sentence in Tok Pisin with the corresponding sentence in Haitian Creole spoken in the Caribbean (which has French as its lexifier language), you find that there are remarkable similarities between the two. Compare the following two sentences in these two languages:

| | | | | |
|----------------|------------------------|-----|------|-------|
| Tok Pisin | Em | no | bin | save. |
| Haitian Creole | Li | pa | te | koné |
| | he/she | not | past | know |
| | 'He/she did not know.' | | | |

You can see that, although the words in these two languages are quite different in their shape, reflecting their different origins, the order in which the words occur is exactly the same.

This becomes even more significant if we compare the corresponding sentences in their respective lexifier languages. In English, the structure of the sentence *Hel/she did not know* involves the following features:

The first element is the subject pronoun.

The second element is the verb *do*, which is put there to carry the tense marking. In this case, the tense is past, so the verb appears in the form *did*.

The third element is the negative marker *not* (which optionally appears reduced in form to the suffix *-n't*).

The fourth element is the verb *know*, which occurs in the infinitive form; that is, it does not take any suffixes for tense, as this is already in the form *did*.

The corresponding French phrase *Il/elle ne connaissait pas*, however, has the following, quite different structure:

The first element is again a subject pronoun, of the form *il* 'he' or *elle* 'she'.

The second element is the form *ne*, which marks the verb as being negative.

The third element is the verb root *connaiss-* 'know'.

Attached to this verb is the fourth element, the suffix *-ait*, which marks the verb as being in the past tense, as well as agreeing with the subject *il/elle*.

The fifth and final element is the form *pas*, which, in conjunction with *ne* before the verb, also marks the negative.

Thus, the structures of the English and French sentences can be summarized as follows:

| | | | | |
|---------|---------|----------|------------|----------|
| English | Subject | Do+Tense | Negative | Verb |
| French | Subject | Negative | Verb+Tense | Negative |

The question that we need to ask ourselves now is this: If the structures of English and French are so different, how is it that the structures of the two pidgin and creole languages that are derived from them are so similar? Both Tok Pisin and Haitian Creole share the following basic structure in these sentences:

Subject Negative Tense Verb

The two pidgin languages are closer in structure to each other than either is to French or to English. Clearly, this cannot be because of the influence of the superstrate languages, as English and French are quite different from each other. We cannot put this down to similarities in the substrate languages,

either, as these are the languages of New Britain and New Ireland in the case of Tok Pisin, and West African languages in the case of Haitian Creole, and these languages are quite different from each other.

One explanation that has been proposed in the past to explain facts such as these was that speakers of all languages are born with some kind of basic idea about how to simplify their language in situations where it is necessary, typically in language-contact situations. This means that we all have some kind of readymade instructions in our heads that tell us how to simplify our languages and to speak a kind of basic, understandable language where all we have to learn is the vocabulary. The reason Tok Pisin and Haitian Creole exhibit such similarities is that people in both places share this basic set of instructions about how to simplify language.

Despite the existence of similarities such as this between Tok Pisin and Haitian Creole, it has become apparent that pidgin languages exhibit many differences, as well as similarities. The apparently remarkable similarity between these two languages that you have just seen may, in fact, not be as significant as it appears. If we accept that both English and French structures are going to have their bound forms eliminated, as well as the grammatical redundancy reduced, it is almost certain that we will end up with four morphemes in whatever pidgin emerges in order to express this meaning. Given that the basic word order in English, French, and the two sets of substrate languages is SVO, it is again predictable that the subject pronoun would end up coming before the verb. The verb in both English and French is the final element in the verb phrase in these clauses, so again it should not be a great surprise to find the other morphemes marking negation and tense occurring before it. The only real surprise is the relative ordering of the negative and tense marker in Tok Pisin and Haitian Creole, but with just this single similarity, we could suggest that this is due to mere chance.

Attempts to find shared structural characteristics among all pidgins and creoles have failed to reveal anything that is absolutely consistent for every case, and attention has since turned specifically to creoles. Pidgins, it is now felt, are less likely to show any kinds of features common to all languages, because pidgins are by definition nobody's mother tongue. This means that there is always the possibility that substrate patterns could interfere with patterns derived from features that might be common to all languages. If parallel features develop among creoles, however, presumably this cannot be due to substrate interference as the speakers of such languages do not know any other languages. The prediction is that as a pidgin becomes a creole, it will expand structurally (as well as lexically). Although we should therefore be able to examine the structures of creoles in order to find out how it is that languages worldwide undergo creolization, initial studies of the process of creolization produced disappointing results. In comparisons between people who speak Tok Pisin as a second language and the increasing number who are growing up speaking it as their first language, it has turned out, that there are very few real differences between the two groups.

The terms “pidgin” and “creole” are actually quite difficult to apply to particular situations when they are defined as they are in this section (even though these are the definitions that are given in almost all standard textbooks on the subject). As I have just indicated, in Papua New Guinea today, Tok Pisin is spoken as a second (or third, or fourth) language by the majority of the population, but a sizeable minority of urban Papua New Guineans—typically those whose parents come from different parts of the country and who speak different vernaculars—are now growing up speaking Tok Pisin as their first language. Do we say that Tok Pisin is a pidgin, or a creole, or a pidgin that is becoming a creole? The fact that there are no major differences in the speech of those who speak it as a “pidgin” and those who speak it as a “creole” makes the distinction seem almost pointless.

There is no agreement on the issue of how pidgins and creoles should be handled in a family-tree model of language change. Pidginization is generally regarded as a somewhat exceptional case in the evolution of languages. “Normal” languages can be said to be descended from another language that is clearly recognizable as its ancestor. French, for example, is descended from Latin, and Samoan is descended from Proto-Polynesian. But where do pidgin languages fit into the comparative method? Melanesian Pidgin does not have an ancestor in the same sense in which French has Latin as its ancestor. In 1840, Melanesian Pidgin did not exist, but by the 1860s, it was widely spoken in some parts of Melanesia and had already spread to other areas where Melanesians had been taken as laborers, such as parts of Queensland in Australia.

What language is Melanesian Pidgin descended from? The family-tree model breaks down when it comes to creole languages, because in a sense they spring out of nowhere! Some linguists might be tempted to classify Melanesian Pidgin as a Germanic language and to place it in a subgroup along with English (as a kind of daughter language of English). Certainly, the lexicon of Melanesian Pidgin is largely derived from English, but it is much harder to say that its grammar is derived from English grammar. Although many features of the grammar of Melanesian Pidgin seem to derive from Austronesian languages, few linguists would go as far as to draw a family tree of the Austronesian languages with Melanesian Pidgin as one of the branches. After all, there are no systematic sound correspondences between Melanesian creoles and the Austronesian languages, as its lexicon is largely derived from English.

Therefore, creoles tend to be either ignored or placed in the “too hard” basket by traditional comparative linguists. Our models for representing linguistic relationships do not provide a way to model creoles, because the models assume incremental change within a continuous speech population. Furthermore, family-tree models assume that languages have a single parent. However, as we’ve seen with creoles, they tend to have multiple inputs: pidgins, the lexifier or superstrate language, and one or more substrate languages.

As such, the different type of language transmission we see in creoles falls outside this type of modeling.

That does not mean, however, that pidgins and creoles can't (or shouldn't) be studied in historical linguistics. On the contrary! It means that they can't (and shouldn't) be represented on family trees, but relationship modeling is only one part of historical linguistics. We also study the types of changes that languages undergo, the processes in society that lead to differentiation in language, and what happens when languages come into contact.

Are there significant differences between creoles and non-creole languages? Yes and no. The main difference is in their formation, although it shouldn't be surprising that we can see some similarities between processes in creolization and in language change across generations of speakers of the same language. Furthermore, once creoles are formed, they also exhibit the same types of sound and grammatical change that we see elsewhere (although most of the creoles studied by linguists are colonial phenomena and are therefore rather recent in world terms). Some linguists (for example, Mufwene 2001) argue that there is little meaningful difference between processes of pidginization and creolization and other types of language change; we can find non-creole features in creole languages, and creole-like features in the histories of non-creole languages. For example, if you compare the grammars of modern Dutch and modern Afrikaans, it is tempting to describe Afrikaans as pidginized Dutch, as its grammar is certainly simplified and more regular than that of Dutch. The term "creole" can at times be as much a political label as a linguistic one.

14.3 MIXED LANGUAGES

Contact phenomena are widespread in many languages and societies, and pidgins and creoles are not the only types of contact languages. There is a third type, which seems to be rarer than creoles, called MIXED LANGUAGES. They have been recognized only relatively recently in the literature, but they are found in many parts of the world, including North America (Michif), South America (Media Lengua), Africa (Sango), and Australia (young people's Gurindji). Mixed languages have parts of their grammar from different donor languages. Here I show features of mixed languages by taking a case study from Michif.⁸

Michif is the language of the Métis of Canada, the descendants of Cree First Nations and French settlers, but now an independent group in their own right. Most of Michif noun morphology is based on French, but there have been some changes. For example, some vowel-initial words in French have been reanalyzed as starting with the consonant of the French article:

| Michif | | French | |
|--------|------------|--------------|-------------|
| zaʔɛr | 'business' | les affaires | |
| lɔm | 'man' | l'homme | |
| laʒ | 'age' | l'age | |
| za:br | 'tree' | les arbres | 'the trees' |
| nil | 'island' | une île | 'an island' |
| larʒã | 'money' | l'argent | 'the money' |

The verbs, on the other hand, are close to those found in Cree. Cree is an Algonquian language, and all Algonquian languages are polysynthetic and have a great deal of agreement. There can be up to seven affixes on a Michif or Cree verb root. Here is an example of a Michif sentence with the origins of each word given:

| | | | | |
|---|--------|-----------|--------|--------------------------------|
| gi:-fa:pu-st-a:na:n | li | rũ-d | pɔrt | ʃi-pi:stukwe:-ja:hk |
| 1 past-passive-go-1 plural.excl | the | circle-of | door | completive-enter-1 plural.excl |
| Cree | French | French | French | Cree |
| 'We [but not you] walked through the archway to come in.' | | | | |

In effect, Michif has two phonological systems: one for the Cree portion of the vocabulary and one for the French portion. The French vocabulary has nine oral vowels (*i, ɛ, y, œ, ɨ, a, u, ɔ, ɑ*) and four nasalized vowels (*ãe, õẽ, ã, õ*); the Cree vocabulary has a set of seven long and short oral vowels (*i, i:, e:, a, a:, u, u:*) and three nasalized vowels (*ɨ, õẽ, õ*). The French vocabulary shows a voicing contrast, but the Cree portion does not.

Thus Michif has one part of the language from French and another part from Cree. Michif is not a creole; we saw in the preceding section that creoles have most of their vocabulary from a single (superstrate) language and may have grammatical features that are similar to the substrate language or languages. Michif is not like that; rather, there is morphology and syntax from both contributing languages (and some additional features that are found in neither of the contributing languages).

There seems to be one thing in common among all the groups with mixed languages. They all appear to have been formed as a type of "in-group" language, where minority speakers wished to differentiate themselves linguistically from their neighbors. This commonality does not provide us with an explanation for how mixed languages form, but it does provide us with an explanation for why they might exist.

14.4 ESOTEROGENY AND EXOTEROGENY

Creole languages are often described as being structurally more simple than other languages. While all languages are complex systems, different languages distribute their complexity in different ways. Some languages have lots of distinctive sounds, while others have fewer (and are therefore more "simple" in that dimension), but having fewer distinctive sounds means either that the words in the language will tend to be longer or that there will be more homonyms (thus increasing complexity in other dimensions).

The linguist Thurston (1987) introduced the terms "esoteric" and "exoteric" languages as a way of explaining this difference. An ESOTERIC language is one that is used primarily for intragroup communication and which sets a group off from surrounding groups. Such languages tend to become increasingly complex as they are transmitted from generation to generation as they are subject to a number of functional pressures. Phonological efficiency is developed at the expense of morphological transparency, which means that there is likely to be a greater number of portmanteau morphemes and a greater amount of allomorphic variation. Such languages typically develop suppletive morphological marking, and the lexicon makes an increasingly fine set of semantic distinctions. Originally optionally marked categories become grammaticalized. Outsiders typically find an esoteric language difficult to learn, which means that it functions even more efficiently as a marker of identity.

An EXOTERIC language, in contrast, is one that is used also for intergroup communication. Given the kinds of circumstances in which such languages are used, there will be many people for whom intelligibility rather than grammaticality is the primary concern. Such languages tend to develop in ways that make them easier to learn. Changes in exoteric languages are therefore likely to be in the opposite direction to those that are characteristic of developments in esoteric languages.

Of course, if we are going to allow that languages can differ in their degrees of complexity, we need to offer some kind of absolute definition of what constitutes linguistic simplicity, as vague feelings are not going to be enough to go by. We also have to avoid the possibility that a particular pattern in Polish may be relatively complex for me as, say, an English speaker, though a speaker of Russian may find it quite unchallenging, simply because the two languages are structurally similar to begin with. According to Thurston, a language that approximates the following characteristics can be described as simple:

1. There is an approximation of a one-to-one correspondence between form and meaning.
2. There is little stylistic or sociolinguistic variation.
3. There are relatively few grammatically marked distinctions.

4. There are relatively few bound morphemes, and these morphemes exhibit little allomorphic variation or suppletion.
5. Grammatical constructions would approximate one function per pattern, and the patterns would apply regularly—that is, without special exceptions.
6. At the level of the lexicon, there would be few opaque idioms.

Whether a language is simple or complex is obviously relative rather than absolute. This means that it is possible to place languages on a continuum between two extremes. Comparing closely related German, Dutch, and Afrikaans, for example, it is clear that in most respects these languages are structurally similar, except that the Dutch inflectional system is simpler than that of German, while that of Afrikaans is simpler than that of Dutch. In this case, then, we clearly have a cline of structural simplicity: German > Dutch > Afrikaans.

The evolution of new languages can therefore be said to involve two different kinds of processes: exoterogeny and esoterogeny. EXOTEROGENY results in the development of a new exoteric language. In the most extreme example of this kind of process, words are simply taken from another language and mapped in the simplest possible way onto the phonological, syntactic, and semantic patterns of the language that the community already speaks. ESOTEROGENY, on the other hand, involves the development of a new esoteric language, in which diversification proceeds in the direction of greater complexity.

While the correspondence between local emblematicity and esoterogeny, on the one hand, and use as a lingua franca and exoterogeny, on the other, is appealing, and many examples can be presented to make the correspondence appear convincing, we have a long way to go before we can say that we have explained all instances of linguistic diversification according to this model, and there are counterexamples. For instance, while it may be possible to invoke such explanations in the exogenetic development of Afrikaans out of Dutch—given what we know of the multilingual situation of the early Dutch settlers in South Africa—it would be much more difficult to account for the apparently exogenetic development of Dutch out of an earlier pattern that was more like that which we find in German.⁹

14.5 LANGUAGE DEATH AND LANGUAGE SHIFT

In chapter 1, I refer to the fact that a language can die. Language death is almost always associated with language contact. The only situation in which a language may die without language contact taking place is in the comparatively rare situation in which an entire speech community is wiped out by a massive calamity such as a volcanic eruption, a tsunami, a military slaughter, or an epidemic. Such things have unfortunately happened in the

past. Oral tradition in central Vanuatu tells of the once large island of Kuwae which was shattered by a volcanic cataclysm into the much smaller present-day islands of Tongoa and the Shepherd Islands. This massive eruption must have killed large numbers of people. Oral tradition records that, although a small number of people from Kuwae survived this holocaust, when the new, smaller islands were resettled by people from the nearby larger island of Efate, they brought with them their own language, which explains why the people from these islands speak a dialect of the Efate language to this day. Presumably, the original language of Kuwae disappeared with the death of the last survivors of the eruption. The history of Aboriginal Australia is full of accounts of the extermination of whole communities of Aboriginal people by European settlers, often by the most inhumane methods such as the deliberate introduction of smallpox or by vicious shooting sprees. In some cases, epidemics preceded the settlers. Again, unknown numbers of languages disappeared from the record with the disappearance of their speakers.

Language death typically occurs in much less catastrophic circumstances and arises as a result of language contact over an extended period of time. When speakers of two languages come into contact and speakers of one of the two languages have power over speakers of the other language, either by force of social prestige or by demographic dominance, it is possible for speakers of the socially weaker language to abandon their language in favor of the dominant language. This has taken place in many parts of the world in the past and is probably accelerating today as languages like English, French, and Spanish become increasingly dominant worldwide through the power of education, government, and the mass media.

Many Australian languages have disappeared, not because their speakers were exterminated but because the generations of the past either chose to or were forced to speak to their children in English. Only about 1 percent of Hawaiians today speak Hawaiian, the remainder having shifted to English, and Māori in New Zealand has shown signs of going the same way, with only about 10 percent of Māori people today speaking the ancestral language. The languages in some parts of Papua New Guinea (especially in the Sepik) are under pressure, not from English but from Tok Pisin. In Europe, also, minority languages are under pressure from larger languages. Irish, Scots Gaelic, and Welsh are all under pressure from English; Friesian is under pressure from Dutch; and Breton is under pressure from French. In East Africa, many languages are losing out to Swahili.

14.5.1 Causes of Language Death

A description of the social circumstances surrounding the death of a language belongs in a volume on sociolinguistics, so in this book I concentrate not on what causes a language to die but on what happens to the language itself as it dies. Before I can do this, however, some discussion of what causes a language

to die is necessary. Here I outline what has happened in the history of Māori in New Zealand.

From the time of the original settlement of Aotearoa (known to the outside world as New Zealand) about 1,000 years ago, the Māori had uncontested control over their territory, and their language functioned as part of their flourishing culture. In the beginning of the nineteenth century, European settlers began arriving, initially in small numbers and, from the second half of the nineteenth century, in an increasing flood. The Māori lost much of their land to the settlers and quickly came under their military control, and later also under their political and economic control. However, the Māori remained a largely rural rather than an urban people, living together in communities, and their language continued to flourish, even though their children learned English at school.

A major social change occurred after World War II as many Māori began moving from the rural areas to the cities and towns to get jobs. Without a fluent command of English, it was difficult to get jobs, and parents saw it as benefiting their children if they refused to speak to them in Māori and insisted on only English in the home. The next generation that grew up in towns therefore tended to learn only a little Māori (possibly from their grandparents who spoke little English) or none at all. Of course, the children of this generation who are today's teenagers and young adults have also grown up speaking nothing but English. It is probably not completely accurate to say that the Māori language began to die; rather, it began to commit suicide. The result is that today, about 90 percent of Māori speak only the language of their original conquerors.¹⁰

14.5.2 Young People's Varieties: Structure Change during Language Shift

In communities where it is recognized that a language is in a precarious situation, the remaining fluent speakers frequently comment on the fact that younger generations no longer speak the language "properly." Fluent speakers of Māori, for example, point to overwhelming lexical interference from the dominant language, confusion of grammatical distinctions, and poor command of the stylistic repertoire. The same sorts of changes are found all over the world in situations where languages are showing signs of being replaced by other languages. Sometimes the older generations will attempt to correct the mistakes of younger partial speakers (i.e., those whose command of the language has suffered as a result of language shift taking place). This can, of course, cause partial speakers to become embarrassed and to avoid using the language with older people for fear of further correction. Rather than improving the chances of the language surviving, this may make it even less likely that it will survive, especially if the language is very small one.

There has been a growing move in recent years to study structural change under language pressure. One type of language goes under the name "young people's varieties": in such cases, the children of a speech community speak rather differently from their parents' and grandparents' generation. Let us look at a case study.

One language that is recognized as being near to extinction is the Dyirbal language of the coast of northern Queensland in Australia. While the older people are recognized as able to speak the language "correctly," the younger generations have grown up either speaking no Dyirbal at all (using only English) or speaking a kind of Dyirbal that everybody considers to be "corrupted" in some way. At the simplest level, this involves the frequent use of words (and phrases) of English (or pidgin) origin for which there are established Dyirbal words. Sometimes the younger people may have forgotten the original Dyirbal word, though in other cases they may use an English word even though they do know the corresponding word in Dyirbal.

The use of words of foreign origin is not necessarily a sign of imminent language death. If it were, then English with its huge number of borrowed words should be a prime example of a dying language. Instead, the enthusiasm with which English has accepted new vocabulary is generally taken as a sign of its extreme vitality. However, the speech of younger people in the Dyirbal community is also grammatically quite different from that of the older people. Younger speakers are reducing the morphological complexity of the language by eliminating some suffixes. The grammatical functions that were originally expressed by these suffixes are now often expressed by free forms that are derived from English, as in the following examples:

Old People's Dyirbal

| | | | |
|----------|----------|-----------------|--------------------|
| Ban | jugumbil | pina- <i>pu</i> | jugu- <i>ŋga</i> . |
| feminine | woman | sit-nonfuture | log-on |

Young People's Dyirbal

| | | | | |
|----------|----------|-----------------|----|-------|
| Ban | jugumbil | pina- <i>pu</i> | on | jugu. |
| feminine | woman | sit-nonfuture | on | log |

'The woman sat on a log.'

Grammatical constructions that are different from those of English are also particularly subject to change. The Dyirbal that is spoken by the older people has a free word order, similar to what I describe for Latin in sec. 12.2.1 This is possible because all noun phrases in Dyirbal are obligatorily marked by suffixes that clearly indicate which is the subject

and which is the object. However, younger speakers tend to leave the ergative suffix off nouns that function as the subjects of transitive verbs and distinguish the subject and object noun phrases by using a fixed SVO word order as in English. (For more detailed discussion of ergativity, see sec.12.2.2.) Instead of using the ergative form /buliman-du/ 'policeman (ergative)', a younger person might produce a sentence such as the following, with no suffix at all on the subject—which an older person would judge as ungrammatical:

Buliman ŋanba-n ban bulaji.
 policeman ask-nonfuture feminine two
 'The policeman asked those two (women).'

As the language comes under increasing pressure from English, we can expect that there will be greater influence of English vocabulary and structural patterns on the language. Some speakers are already producing sentences that are basically English, even though they still contain fragments of Dyrirbal, such as the following:

They bin gunimariju but they never bin find-im.
 'They looked for him but they didn't find him.'

The Dyrirbal verb *gunimariju* 'look for' occurs in a sentence with an English subject and with the past tense marker /bin/ that derives from the earlier Aboriginal pidgin. The object suffix *-im* on the verb *find* also derives from the pidgin and is part of many forms of Aboriginal English.

The question of just what happens to the grammar as a language dies has begun to arouse considerable interest among scholars of language change. In one sense, what has happened to the vocabulary and the grammar of Dyrirbal is quite unexceptional. The incorporation of vocabulary from one language into another is, as we have seen, a perfectly normal aspect of language change. The kinds of grammatical changes that are taking place are also not radically different in nature from the kinds of changes that take place in situations of ordinary language change. In chapter 12, I indicate that inflecting and agglutinating languages often evolve into isolating languages and that morphological irregularities in languages tend to be eliminated. Just as an accusative language can, over time, acquire an ergative structure, so the shift of Dyrirbal from an ergative structure to an accusative structure marked by word order rather than by case suffixes is again perfectly within the bounds of normal language change.

What is exceptional in the case of Dyrirbal is that the changes are happening on such a massive scale and in such a short period of time. These structural changes have all taken place within the space of 25 years. English began to undergo similar sorts of changes from around the time of contact

after the Viking and Norman invasions, but it took several centuries for this to happen. Another difference between the kinds of changes that are taking place with Dyirbal and those which happened in the history of English is that in Dyirbal, a change that results in the loss of some aspect of the grammar (such as the loss of the locative suffix /-ŋga/ in one of the examples above) has not been compensated for by a corresponding development somewhere else in the language. What has happened is that speakers have simply taken over the corresponding English form to express this function—that is, the preposition *on*.

This has led some scholars to suggest that what happens when a language dies is somehow similar to what happens when a pidgin language comes into existence, except that events take place in the opposite order. Just as a pidgin in the early stages of its formation involves grammatical reduction and both structural and lexical variability, so, too, does a dying language. A pidgin also has a reduced stylistic repertoire compared with a “normal” language, and we find the same thing with a language that is dying. Such parallels don’t always work. We also find cases of language death where complexifying changes have occurred or where changes have occurred that do not bring the endangered language closer to the language that speakers are shifting to.

14.5.3 Speed of Language Death

Although it need only take a break in transmission between a single generation for a language to be doomed, it is possible for features of an old language to be maintained over a relatively long period. In sec. 9.4, I talk about how languages can sometimes change to allow a local group to mark its separate identity in some way. Although the examples that I give there came from a Papua New Guinea context, this is not something that is restricted just to these languages. Often when an ethnic group switches from one language to another, people develop ways of marking their ethnicity through their new language. There is a variety of English in America that is typically associated with Blacks as against Whites. Māori in New Zealand can often be distinguished from Pākehā by the way they speak English. Books about the history of Tasmanian Aborigines point out that the last fully-descended Tasmanian died in 1876, and her language died with her, yet some of the 4,000 or so people in Tasmania today who are of Aboriginal descent (and who proudly identify themselves as Tasmanian Aborigines) still use the occasional word of Aboriginal origin in their speech. We can expect that while Dyirbal is doomed as a distinct language, the succeeding generations of people who belong to this community will continue to sprinkle their English with individual words of Dyirbal origin, even though there will be little evidence of Dyirbal grammatical structures.

Reading Guide Questions

1. What is interference as distinct from diffusion?
2. What is calquing?
3. Can phonemes be copied from one language to another?
4. How can morphemes from one language enter another?
5. What is the difference between convergence and diffusion?
6. What is a linguistic area?
7. To what extent is syntactic copying possible?
8. What are the characteristics of a pidgin?
9. What is a creole?
10. What is the difference between a superordinate/superstrate and a substrate language?
11. What is meant by language death?
12. What are some of the changes that can happen when a language dies?

Exercises

1. Examine the following data from two different languages, one of which is vernacular Fijian and the other the pidginized form of Fijian that emerged on plantations in Fiji during the nineteenth century:

| Language A | Language B | |
|--------------------|-------------------|---------------------------------|
| na noqu vale | na vale koyau | 'my house' |
| na nomu veiniu | na veiniu koiko | 'your plantation' |
| na nona koro | na koro kokoya | 'his/her village' |
| na nodra vale | na vale koratou | 'their house' |
| na nodratou veiniu | na veiniu koratou | 'their (three) plantation' |
| na nodrau koro | na koro koratou | 'their (two) village' |
| na nomudrau bilo | na bilo kemudou | 'your (two) cup' |
| na nomuni vale | na vale kemudou | 'your (many) house' |
| na nomudou veiniu | na veiniu kemudou | 'your (three) plantation' |
| na noda vosa | na vosa keitou | 'our (many inclusive) language' |
| na neimami vosa | na vosa keitou | 'our (many exclusive) language' |
| na meirau wai | na wai keitou | 'our (two exclusive) water' |
| na meitou bia | na bia keitou | 'our (three exclusive) beer' |
| na meimami bia | na bia keitou | 'our (many exclusive) beer' |
| na medaru bia | na bia keitou | 'our (two inclusive) beer' |
| na medatou wisiki | na wisiki keitou | 'our (three inclusive) whisky' |

| | | |
|----------------|-------------------|-------------------------------|
| na meda wisiki | na wisiki keitou | 'our (many inclusive) whisky' |
| na tamaqu | na tamana koyau | 'my father' |
| na tamamu | na tamana koiko | 'your father' |
| na tamana | na tamana kokoya | 'his/her father' |
| na ligada | na ligana keitou | 'our (many inclusive) hands' |
| na ligadra | na ligana koratou | 'their hands' |

(In the Fijian orthography that is used in these examples, the symbol *ŋ* is used to represent a prenasalized voiced velar stop, which is phonetically [ŋg].)

- Which of these two languages is vernacular Fijian, and which is the pidgin form of Fijian? What are the structural features that enable you to say this?
- Give the equivalents of the following phrases in both vernacular Fijian and pidgin Fijian:

| | |
|--------------------|---------------------------------|
| your hand | their (three) beer |
| your (two) father | our (many exclusive) house |
| your (many) father | their language |
| his/her water | our (many inclusive) plantation |
| their (two) whisky | my cup |

- Consider the following additional forms in Language A:

| | |
|-----------|------------------|
| na tinamu | 'your mother' |
| na tinana | 'his/her mother' |

On the basis of this information, give the following in both vernacular Fijian and pidgin Fijian:

| | |
|-----------------------------|------------------------------|
| our (many inclusive) mother | your (many) mother |
| their mother | our (three exclusive) mother |
| your (two) mother | |

- Examine the following forms in Haitian Creole and in French:

| French | Haitian Creole | |
|--------------------|----------------|----------------------------------|
| je suis malade | m malad | 'I am sick' |
| ils sont malades | yo malad | 'they (<i>masc.</i>) are sick' |
| elles sont malades | yo malad | 'they (<i>fem.</i>) are sick' |

| | | |
|---------------------|-------------|--------------------------------|
| j'étais malade | m te malad | 'I was sick' |
| ils étaient malades | yo te malad | 'they were sick' |
| nous achèterons | n ap achte | 'we will buy' |
| je vais | m ale | 'I am going' |
| vous irez | u ap ale | 'you (<i>pl.</i>) will go' |
| tu iras | u ap ale | 'you (<i>sing.</i>) will go' |
| il a couru | li te kuri | 'he has run' |
| il est bon | li bien | 'he is good' |
| elle va | li ale | 'she is going' |
| je suis allé | me te ale | 'I have gone' |
| il a acheté | li te achte | 'he has bought' |

- a. What are the features of Haitian Creole by which we can recognize that it has undergone the process of pidginization?
- b. How would you express the following in French?

| | |
|-------------|------------|
| yo ap malad | yo te bien |
| m achte | u ap ale |
| n te kuri | |

3. Compare the following forms from the Bandjatang language of northern New South Wales (in Australia) as it was spoken by people who learned the language in the early twentieth century and people who learned to speak it in the late nineteenth century. Describe the system of plural formation in the older language and how it has changed in the modern language. In what way are the changes that have taken place similar to the changes involved in the formation of pidgins and creoles?

| Older People's Bandjatang | Later Generation Bandjatang | |
|------------------------------|--------------------------------|-------------------------|
| gala gibirga: | gala gibirga: | 'mahogany tree' |
| ga:ɲu gibi:ɲbilga: | ga:ɲu gibirga: | 'mahogany trees' |
| gala bunawga: | gala bunawga: | 'bloodwood tree' |
| ga:ɲu buna:ɲbilga: | ga:ɲu bunawga: | 'bloodwood trees' |
| gala barbamga: | gala barbamga: | 'spotted gum tree' |
| ga:ɲu barba:ɲbilga: | ga:ɲu barbamga: | 'spotted gum trees' |
| gala buje:ga | gala buje:ga | 'Moreton Bay fig tree' |
| ga:ɲu buje:ɲbilga: | ga:ɲu buje:ga | 'Moreton Bay fig trees' |

| | | |
|----------------------|-------------------|---------------------|
| gala bilanga: | gala bilanga: | 'oak tree' |
| ga:pu bila:ŋbɪŋga: | ga:pu bilanga: | 'oak trees' |
| gala ŋarulga: | gala ŋarulga: | 'box tree' |
| ga:pu ŋaru:ŋbɪŋga: | ga:pu ŋarulga: | 'box trees' |
| gala jigam | gala jigam | 'piece of meat' |
| ga:pu jigambil | ga:pu jigam | 'pieces of meat' |
| gala wura:ŋ | gala wura:ŋ | 'leaf' |
| ga:pu wura:ŋbil | ga:pu wura:ŋ | 'leaves' |
| galajinaŋ | galajinaŋ | 'foot' |
| ga:pujinaŋbil | ga:pujinaŋ | 'feet' |
| gala deberdebe:r | gala deberdebe:r | 'plover' |
| ga:pu deberdebe:rgan | ga:pu deberdebe:r | 'plovers' |
| gala bagawaŋ | gala bagawaŋ | 'leatherhead bird' |
| ga:pu bagawaŋgan | ga:pu bagawaŋ | 'leatherhead birds' |
| gala muʒumʒar | gala muʒumʒar | 'son' |
| ga:pu muʒumʒir | ga:pu muʒumʒar | 'sons' |
| gala baniʒar | gala baniʒar | 'father' |
| ga:pu baniʒir | ga:pu baniʒar | 'fathers' |
| gala balun | gala balun | 'river' |
| ga:pu balungali | ga:pu balun | 'rivers' |
| gala bagul | gala bagul | 'canoe' |
| ga:pu bagulgali | ga:pu bagul | 'canoes' |
| gala daba:j | gala daba:j | 'dog' |
| ga:pu daba:jgali | ga:pu daba:j | 'dogs' |
| gala muru | gala muru | 'nose' |
| ga:pu murugali | ga:pu muru | 'noses' |
| gala dubaj | gala dubaj | 'woman' |
| ga:pu dubaymir | ga:pu dubaj | 'women' |
| gala wagap | gala wagap | 'catfish' |
| ga:pu wagapmir | ga:pu wagap | 'catfish' (pl.) |
| gala bajgal | gala bajgal | 'man' |
| ga:pu bajgalbajga:l | ga:pu bajgal | 'men' |
| gala dugun | gala dugun | 'mountain' |
| ga:pu dugundugu:n | ga:pu dugun | 'mountains' |
| gala bargan | gala bargan | 'boomerang' |
| ga:pu bargan | ga:pu bargan | 'boomerangs' |

| | | |
|--------------|--------------|-----------------|
| gala mundu | gala mundu | 'stomach' |
| ga:ɲu mundu | ga:ɲu mundu | 'stomachs' |
| gala jamba: | gala jamba: | 'carpet snake' |
| ga:ɲu jamba: | ga:ɲu jamba: | 'carpet snakes' |
| gala ɲa:wun | gala ɲa:wun | 'wood duck' |
| ga:ɲu ɲa:wun | ga:ɲu ɲa:wun | 'wood ducks' |

Further Reading

- Robert J. Jeffers and Ilse Lehiste, *Principles and Methods for Historical Linguistics*, chapter 9, "Language Contact and Linguistic Change," pp. 138–59.
- Theodora Bynon, *Historical Linguistics*, chapter 6, "Contact between Languages," pp. 216–61.
- René Appel and Pieter Muysken, *Language Contact and Bilingualism*.
- Sarah Thomason and Terrence Kaufman, *Language Contact, Creolization, and Genetic Linguistics*.
- Suzanne Romaine, *Pidgin and Creole Languages*.
- Martin Maiden, "Into the Past: Morphological Change in the Dying Years of Dalmatian."
- Annette Schmidt, *Young People's Dyirbal: An Example of Language Death in Australia*. (There are quite a few books on "young people's varieties" in Australia; two others are Lee, *Tivi Today*, and Langlois, *Alive and Kicking*.)
- Hans Henrich Hock, *Principles of Historical Linguistics*, chapter 16, "Linguistic Contact: Koinés, Convergence, Pidgins, Creoles, Language Death," pp. 472–531.
- Sarah Thomason, *Language Contact: An Introduction*, chapter 4, "Contact-induced change: results" pp. 59–98; and chapter 6, "Contact-induced language change: language mechanisms," pp. 29–156.
- The *Journal of Pidgin and Creole Languages* has many articles of interest to historical linguistics, as well as articles on the synchronic structure of pidgin and creole languages.
- David Crystal, *Language Death*, chapter 1, "What is language death", pp 1–26.
- K. David Harrison, *When Languages Die*, chapter 3, "Many Moars Ago," pp. 61–94.
- Malcolm D. Ross, "Calquing and Metatypy."

Cultural Reconstruction

Different people who practice historical linguistics have their own particular reasons for their interest in this field. Some enjoy the intellectual challenge of applying a difficult technique to “dig” into the past and find out about things that we could not know about otherwise. Some may be looking for “universal” features of language and how languages change. And others may study historical linguistics in an effort to use the information it can provide to tell us something about the nonlinguistic history of the people who speak the language. There are many different methods we can use to find out about the past, and looking at language change is only one of them. This chapter is all about how to use language as part of the reconstruction of culture and other nonlinguistic aspects of prehistory. We will see how linguistic prehistory ties in with other ways of looking at culture change.

15.1 ARCHAEOLOGY

Once we start considering the question of cultural reconstruction, there are various ways in which we can tackle this problem. Archaeology is one of them. Archaeologists attempt to reconstruct cultures on the basis of the material remains left by people of the past. They uncover material that has been buried by natural processes of soil movement, and they can use a variety of scientific methods to provide actual dates for the existence of particular cultural features and changes in cultures in the past, as long as there is some

material left over. Of course, if nothing of the society is preserved, we cannot draw any conclusions.

For instance, archaeologists are able to tell us that there have been people living in the land that is now Australia and Papua New Guinea for at least as long as there have been people living in what is now Europe. They can tell us with a fair degree of certainty that people were living 40,000 years ago on what was then a single huge land mass. Human burials were uncovered originally by erosion in one part of Australia that have been dated as very ancient, and in the Huon Gulf of Papua New Guinea, stone axeheads that are similar to stone axeheads found in Australia and other parts of Southeast Asia have recently been uncovered in soil layers that are well over 40,000 years old. However, there are now suggestions coming out that this period is too short, and archaeologists are expecting to find evidence that there has been human occupation in this area for 55,000 years or more.

We can also use archaeology to reconstruct ancient trade routes. In the Pacific, goods such as valuable shells, clay pots, and obsidian (a kind of natural glass of volcanic origin) were traded over huge distances for long periods of time. Analysis of the results can often suggest who were the economically dominant partners in these trading networks.

Archaeologists can also tell us something about population movements and other kinds of cultural contacts between people. For instance, they can tell us that the Australian Aborigines did not have dogs until about 4,000 or 5,000 years ago. By the time that the first Europeans set foot in Australia in the 1600s and 1700s, the dog was well established throughout Australia (except in Tasmania). Presumably, the dog was introduced to the mainland of Australia as a result of some cultural contact, either in the form of trading visits or in the form of a migration by some outside group into Australia, or maybe even by people who were blown off course and were stranded there. Who these people were we do not know, but we can say with certainty that once the Australian Aborigines arrived in their new home, they were not completely cut off from changes and developments that took place in other parts of the world.

Although there is much that archaeologists can tell us, there are many other things that they cannot. They cannot tell us about a people's oral literature, for instance, nor can they tell us much about a people's kinship system. And although archaeologists can often tell us that there have been cultural contacts between one group and another in the past, they cannot say exactly who the two groups were. And, obviously, an archaeologist cannot tell us what language was spoken in the past by a group of people unless they left written records of the language.

15.2 ORAL HISTORY

Another way by which we can attempt to reconstruct a culture is to look at a people's oral history. Eyewitness accounts of events are often passed on

from one generation to another. In particular, oral histories are important for recording genealogies (or extended family histories). For instance, oral historians are sometimes able to tell us the approximate time that a particular village was established, with time being measured by counting generations back from the present. Sometimes oral tradition will record where the people of that village originally lived, who was their leader at the time they moved, and why the move actually took place. In parts of West Africa, everybody can recite their family tree back six generations or more.

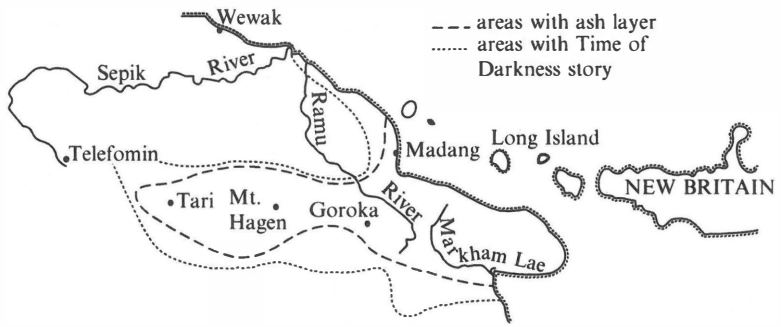
While many facts can be recorded in oral history, there are usually problems in interpretation. Some stories that are passed on from generation to generation are just myths (or legends) that reflect the religious and social system of a community and provide the basis for its religious and social organization. While such stories may have begun as oral history, over the years they have been expanded and changed so much that we no longer know what is truth and what has been added to the story of the time. It is common for people who see something strange to make up a story that explains it. For example, there are stories about times when people could walk between islands in the Pacific, even though these islands were never joined, no matter how low the sea level.

One interesting example is a story of a Time of Darkness that is told in many societies in the Madang and Morobe provinces of Papua New Guinea, as well as in all of the Highlands provinces. Although these stories differ in detail, according to where the story is told or who the particular storyteller is, there is still a remarkable degree of agreement in the stories as they are told over this whole area. The story, in its basic form, goes something like this:

The people heard a loud noise [or sometimes felt an earthquake, or both], and felt that something awful was going to happen. Black clouds started to build up, and they eventually blocked out the sun. Very quickly, the whole place was in darkness like the darkest night. People went into their homes to hide, and they heard something falling from the sky onto their roofs. When they looked out, they saw that it was raining ash. The darkness lasted for three or four days. When the sun reappeared, people found that the whole countryside was covered, and many food gardens had been destroyed, and many houses had collapsed.

This story sounds as though it is describing quite a disaster. However, while people from the Enga Province in the Highlands who tell the story believe that the Time of Darkness was a terrible thing, they also believe that it was the beginning of better times afterward. After this tragedy, the sweet potato grew better in the gardens, and people had more wealth (in the form of pigs) for exchange. Many cultural developments were said to have followed directly from the Time of Darkness.

What this story sounds like is a description of a distant volcanic eruption, even down to the details about the long-term benefits mentioned in the Enga version of the story. This is presumably compatible with the enriching of the



MAP 15.1 Areas of the Time of Darkness in Enga Province, Papua New Guinea

soil from fertile volcanic ash. In fact, this interpretation of the story can be checked out according to scientific methods. There are deposits of volcanic ash in a layer at least 2.5 centimeters thick in the area indicated in map 15.1. The areas where stories about the Time of Darkness are reported are also marked on this map. These two areas coincide very closely, so presumably the story and the layer of volcanic ash are connected historically. Geologists are able to locate the source of the ash as the volcano on Long Island (which is also marked on the map), and they suggest that there was probably a major eruption from this volcano sometime between 1640 and 1820. The people who tell the Time of Darkness story claim definitely that the story is “historical” rather than “legendary,” and the date that we can arrive at for the event is somewhere between 1820 and 1860, based on a count of generations from the present. The version of the scientist and the version of the oral historian are therefore clearly compatible. Even the actual dates overlap, suggesting that perhaps the eruption took place closer to 1820 than any of the other possible dates.

Many other examples of oral tradition are confirmed by archaeological evidence, for which we can also find supporting linguistic evidence. When Europeans first came to settle in New Zealand in the first half of the nineteenth century, they discovered that Māori oral traditions referred to large flightless birds that they called *moa*. These birds might have remained the stuff of myth and fantasy, as have dragons and unicorns,¹ except that from the 1860s onward, the bodies of birds with huge bones, and sometimes even feathers and mummified flesh, began to come to light. It rapidly emerged that these bones came from birds that looked a lot like emus or ostriches but were much bigger, some being at least twice the height of a human from head to foot. This meant that the *moa* was not mythical at all, but a real bird.

When the Māori first arrived in New Zealand (or Aotearoa, as it is referred to in the Māori language), they found a number of different species of this bird, and they developed a culture based in part on *moa*-hunting. We know that the Māori used to hunt the *moa* because archaeologists have discovered charred *moa* bones in fireplaces where food was cooked and

artifacts made of *moa* bone. However, within a few hundred years, it seems, the *moa* had been hunted to extinction, and all that was left was the name as it was recorded in oral tradition. The word /*moa*/ is found in most of the other Polynesian languages, though in these languages it means 'chicken'. Presumably when the first Māori arrived in Aotearoa and they came across these birds, they quickly abandoned their tiny chickens and took advantage of the much more plentiful amounts of meat that the *moa* could provide; in doing this, they transferred the name of their earlier source of protein to these newly discovered game birds. (Another possible explanation for the lack of chickens in Aotearoa is that the Māori had not brought chickens with them in the first place—or perhaps they did, but they drowned or were eaten on the long journey there.)

Other fascinating questions to examine are the stories told by the Māori about their origins. Māori oral traditions tell of canoe voyages from the distant land of Hawaiki, many many generations ago. The stories record the names of particular canoes that came ashore at different locations along the coastline of the new land that they called Aotearoa, and modern Māori groups speak of their descent from one or another of these founding /*waka*/, or canoes. The name Hawaiki has the same origin as the name of the biggest island in Hawaii, which is phonemically *hawaiʻi*, as well as the name of the largest Samoan island, *savaiʻi*. Many of the people whose names are recorded as having traveled on these first canoes are then said to have engaged in heroic voyages of discovery in Aotearoa itself, resulting in the formation of many of the rivers, lakes, mountains, and volcanoes that are found in New Zealand today.

15.3 COMPARATIVE CULTURE

You have seen that, by comparing a number of languages with certain similarities, it is possible to reconstruct a protolanguage as the ancestor of these languages. If we regard culture as involving a system of interrelated facts in the same way as a language is a system of interrelated facts, then it should logically be just as possible to reconstruct protocultures in the same way as we reconstruct protolanguages.

Obviously, any method of cultural reconstruction based on a method of comparative culture like this would not produce results with the same degree of likelihood as we have been able to produce for phonological reconstruction of languages, as our approach would have to involve the less well refined methods that we have for grammatical or semantic reconstruction. In fact, the actual units of a cultural system and the precise nature of the relationships between these units are probably going to be even more difficult to define than the interrelationships of units in grammar and semantics. (Anthropologists have long been envious of the techniques that linguists have developed for scientifically describing language, and they have attempted to imitate these

techniques to describe culture.) The range of “possible” changes in culture is probably even harder to define than the range of possible changes in grammar and semantics, which again makes cultural reconstruction harder. So, while cultural reconstruction by means of an adaptation of the comparative method is presumably possible, any conclusions that we reach in this way must be regarded as shaky.

Let us now look at an example of what I mean by comparative culture, to see how we might use the evidence of a variety of cultures to reconstruct an earlier cultural system. In Samoa, we find that there is an institution called the *fono*, which is a kind of meetinghouse. Most *fono* are oval in structure, with a series of posts in the ground and around the actual building. The *fono* has “members” who come from certain groups in society, and membership in these groups is passed on from father to son. The members select from among their number the person they regard as the most capable to represent them in the *fono*. Such a person is called a *matai*, and he sits inside the *fono* during the meetings, while the people he represents sit outside. In the meetings in the *fono*, all decisions are arrived at by consensus. In Kiribati, communities have a large rectangular meetinghouse called the *maneaba*. In each community there are various groups who have rights to sit in the */maneaba/*, while others sit underneath the building during meetings. Decisions are arrived at by consensus. The similarities between the Samoan *fono* and the Kiribati *maneaba* are so great that we would almost certainly want to say that these are two cognate systems, and that they derive from the same source.

Let us now go to yet another society, this time that of the island of Malakula in Vanuatu. In this society, each community has a large rectangular building (which in Bislama is known as *nakamal*, but which has different names in each of the local languages). The *nakamal* is partitioned off into areas that are regarded as progressively more sacred as one goes toward the back. Men can only go inside as far as the “grade” to which they have been initiated, and initiation into the highest grade requires enormous payments of pigs and other traditional forms of wealth. Outside, a series of carved images are placed in memory of dead people, and spirits of the dead return to these images for certain ceremonies.

This system from Vanuatu, while on the surface it is apparently quite different from the systems of Samoa and Kiribati, still shares some basic features with these other systems. There is still a central meetinghouse, access to which is restricted in certain ways. Decisions are reached by consensus in each case. Outside the meetinghouse are representatives—in one case, living; in the other case, dead. It is therefore not too difficult to imagine all of these features as having been present in the protoculture from which all of these different cultures have evolved. The problem, of course, is in deciding exactly what kind of protoculture we should reconstruct. It seems reasonable to reconstruct some kind of central meetinghouse with some kind of restricted access—but should we reconstruct it as oval (as in the case of Samoa) or as rectangular (as in Kiribati and Vanuatu)? Perhaps all we can say is that it

was probably longer than it was wide. And was access restricted by birthright (as in Samoa) or by wealth (as in Vanuatu)? From the data that we have looked at, it is probably not possible to make a decision on this particular point.

Another question we could ask is, Who were the representatives outside the meeting house? Again, there are various possibilities. First, they may have been people who were not eligible to enter because they lacked the wealth (or perhaps because they were not "born into" the meetinghouse). Perhaps they were people who were not eligible to enter because they were dead (and whose presence was indicated by carved posts, instead). Again, on the evidence that is available we are not really in a position to come to a conclusion on this point.

Cultural reconstruction, difficult as it obviously is, is still a relatively simple matter in places like Polynesia. This is because the populations are located on small islands that are separated by large expanses of ocean, which means that day-to-day contacts were not possible between groups which might influence each other in unpredictable ways. (But there is plenty of archaeological evidence to indicate that different Polynesian peoples were in contact with each other, even over these huge distances.) Polynesian cultures also developed independently, because when Polynesian people settled on an island, there were never any other people living there. Cultural reconstruction in the Melanesian islands, however (and in the rest of the world, for that matter), is more complex because we have to remember that in many cases there were original populations (who may have left no distinct modern trace). There were also opportunities for continual day-to-day contact between people of different cultural backgrounds over many thousands of years. Under these conditions, it is possible for cultural innovations to have spread in a crisscross pattern over huge areas, thereby completely eliminating traces of the original cultures that we would need in order to apply any method of comparative cultural reconstruction successfully.

15.4 HISTORICAL LINGUISTICS

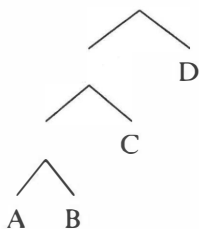
I have now discussed archaeology, oral history, and comparative culture as methods of reconstructing the cultural history of a society. These different methods all provide information that partly overlaps and is partly specific to each method. Archaeology and comparative culture can take us a long way back into the past. Only archaeology can give us reasonably accurate dates for cultural features, and only comparative culture can tell us anything about the nonmaterial culture of a society. Oral history can tell us something about the history of a society, but it cannot take us very far back in time when it comes to detailed information.

The final technique of cultural reconstruction that we have at our disposal is historical linguistics. Historical linguistics can allow us to go back quite a

few thousand years in time. This area of study can provide us with a number of different kinds of information about the history of a society, and this information can then be compared with the information that is provided by archaeology, oral history, and comparative culture as a double check. The kinds of things that historical linguists can tell us are described next.

15.4.1 Relative Sequence of Population Splits

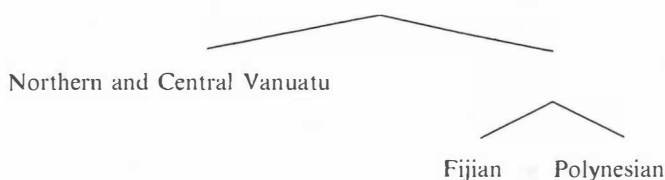
Take a situation such as the following, where there is a language family with four members, subgrouped as shown:



Here we have a number of languages that are all descended from a common ancestor. Languages A, B, and C all belong to a single subgroup, while language D belongs to a different subgroup of its own within the same family. The languages A, B, and C further subgroup such that A and B are more closely related to each other than either is to C. This subgrouping is, of course, arrived at by considering the shared linguistic innovations or changes that have taken place from the protolanguage.

A situation like this will tell us that, at one stage in the history of these languages, there was a single language (Proto-ABCD) which must have been spoken in a single community. This community then split—perhaps by migration, perhaps by a simple lack of contact between two areas without any migration taking place. The result of this split was that in one area, language D emerged, while in the remaining area, the ancestor to languages A, B, and C was spoken. Next, language C hived off from the protolanguage to modern A and B, and the final split was that which saw A and B become separate languages.

A subgrouping pattern of this kind would be compatible only with nonlinguistic evidence which suggests that speakers of language D split off relatively early from speakers of languages A, B, and C. Similarly, we would hope that nonlinguistic evidence would be compatible with the fact that C then split off from A and B and that the last split is also the most recent to have taken place. For instance, if we look at the languages of Polynesia and of Melanesia, we can draw a simple subgrouping diagram:



From this we can assume that all of the modern Polynesian languages go back to a common ancestor and this protolanguage split off from an earlier language that was ancestral to Proto-Polynesian and Fijian. Furthermore, we can assume that this split took place after the ancestor of Fijian and the Polynesian languages split off from the language that was also ancestral to the languages of northern Vanuatu.

If we are talking about language splits, we are presumably also talking about splits in the populations of the speakers of those languages. In the case of Polynesia and Melanesia, where the languages involved are spoken on small, isolated islands, we must also consider the relative age of migrations of entire peoples. I am speaking here of the relative age of population splits, not the absolute age. That is, on the basis of linguistic evidence, we can only say that the Fijian-Polynesian split took place later than the split from northern and central Vanuatu. We cannot say when these splits actually took place. The technique of glottochronology is one way in which some linguists attempt to provide actual dates for population splits, though there are few who would take this seriously now as an accurate indication.

Simply from an examination of a family tree, there is no way that we can tell which group moved away and which group remained in the original location (or, indeed, if both groups moved away in different directions). From the preceding family tree, for example, we cannot say for sure whether the Fijians migrated out of Polynesia or whether the Polynesians migrated out of Fiji (or whether both groups migrated out of northern or central Vanuatu). It is only by referring to nonlinguistic evidence that we can draw such conclusions. In the case of the Oceanic languages, archaeological evidence points to northern and central Vanuatu having been occupied considerably earlier than the islands of Polynesia, so it is probable that the Proto-Polynesians originated as a result of a migration out of Fiji.

15.4.2 The Nature of Cultural Contact

Often, when we are able to isolate copied words from directly inherited (i.e., indigenous) words, we can tell something about the nature of the cultural contact that took place at the time the lexical copying took place. Compare the following words in English, for example:

law justice
 freedom liberty
 kingship royalty

The words on the left are native English words, while those on the right are words copied from French after the invasion of England by the French-speaking Normans in 1066. While the pairs of words are similar in meaning, most people would probably agree that the issues described by the words to the right are more worth dying for than those described by words on the left. A banner reading *Justice and Liberty* is a more effective call to revolution than one that reads *Law and Freedom*. This fact suggests that when these words were copied from French into English, something was regarded as somehow "better" simply because it had a French name rather than an English name. That the French language had social prestige over English at the time is indicated by the following statement made by an Englishman in the English of the time:

Vor bote a man conne Frenss, me telth of him lute.

This translates into Modern English as follows:

Unless a man knows French, one thinks little of him.

Another example of this kind comes from the American Indian language called Navajo. The Navajo word for 'corn' is *nada:ʔ*. Nonlinguistic evidence tells us that the Navajo have only fairly recently acquired a knowledge of corn and that they learned about it from their neighbors, the Pueblo Indians, who speak a different language. Historically, we can reconstruct this Navajo word back to a compound, which literally meant 'enemy-food'. This suggests that when the Navajo and the Pueblo first came into contact with each other, the Navajo considered the Pueblo Indians to be enemies.

Finally, if we compare the vocabulary of modern Melanesian Pidgin with that of English, we would be able to reconstruct something of the nature of the social contacts that took place between Melanesians and Europeans when the language was in its formative years in the nineteenth century. A European is referred to as *masta* (from English 'master'), while Melanesians were referred to as *boi* (from English 'boy'). A Melanesian was a *boi* even if he happened to be married with five children of his own. A European, alternatively, was a *masta* even if he was not yet old enough to shave (though he would then be called a *pikinini masta* 'European boy'). Any Melanesian who managed to make it in the work situation and become an overseer could never be called a *masta* himself; the best he could hope for was to be called a *bosboi* (from 'boss boy').² The development of these terms clearly indicates that the

Europeans held power over the Melanesians when these words were originally incorporated into Melanesian Pidgin.

15.4.3 Sequences of Cultural Contact with Respect to Population Splits

It is sometimes possible to tell if certain cultural contacts took place before or after a population split took place. Let us look at the example of the introduction of the sweet potato into the Pacific.

We know from botanical evidence that the sweet potato was introduced into the whole Oceanic area relatively recently and certainly was not one of the crops that the Proto-Oceanic people brought with them when they settled the Pacific (such as bananas, breadfruit, and yams). Although sweet potato now seems to be well entrenched in the cultures of Papua New Guinea, it did not arrive there until around the sixteenth century, and it probably came from eastern Indonesia. The sweet potato is not indigenous to Indonesia, either; it was introduced to those islands by the Portuguese who first learned about it in South America in the fifteenth and sixteenth centuries. The sweet potato that seems so much at home in Polynesia today is also a recent arrival, though it was probably introduced directly from South America and spread from east to west.

It is also possible to argue on the basis of linguistic evidence that the sweet potato is a relative newcomer to Pacific diets. The word for 'sweet potato' throughout much of Polynesia is *kumala* or something similar. This word is possibly a direct copy of the word for 'sweet potato' in the Quechua language of Peru, where it is known as /kumar/. The same word is also found in many Melanesian languages, including Fijian and some of the languages of Vanuatu and the Solomon Islands (where it was almost certainly introduced in the nineteenth century). Normally, words in the Melanesian languages have undergone a large number of phonological changes that often make Proto-Oceanic words difficult to recognize. For instance, in the Paamese language of Vanuatu, original /*k/ is regularly lost, for example:

Paamese

| | | | |
|-----------|---|-------|---------------|
| *a kai | > | a:i | 'tree' |
| *a ika | > | ai | 'fish' |
| *kapika | > | ahi | 'Malay apple' |
| *masakit | > | mesai | 'sick' |
| *penako | > | hena | 'steal' |
| *a tansik | > | atas | 'sea' |

If there were a word of the shape **kumala* in Proto-Oceanic, by the regular changes in the history of Paamese, this should have ended up in Paamese as *umal*. In fact, Paamese has *kumala*, which preserves both the /k/ and the final vowel. This therefore suggests that Paamese acquired the word *kumala* (and presumably also the thing it referred to) after all of the other phonological changes had taken place in the language.

15.4.4 The Content of a Culture

Given the fact that a language bears a close relationship to the culture of the people who speak it, we can also tell something about the nature of the culture of a people simply by looking at the language they speak. This applies as much to a language that is in use today as to a reconstructed protolanguage.

A major aspect of the relationship that holds between a language and the culture of its speakers is the fact that there is always lexical richness in areas of cultural importance, and there is a corresponding lack of lexical development in areas that are of little importance culturally. Speakers of Polynesian languages typically have a number of different names for different kinds of bananas, sweet potato, and taro. Of course, we would expect that Inuit languages of the United States and Canada would not have any words at all for any of these things, since its speakers live in a place where it is more appropriate to develop lexical specialization for talking about snow.

When we apply this basic principle to a reconstructed protolanguage as a way of determining the content of a protoculture, we are using what is called the *Wörter und Sachen* technique of cultural reconstruction. *Wörter* is a German word meaning 'words', while *Sachen* means 'things'; the name of the technique itself translates as 'words and things'. Basically, the argument goes, if we can reconstruct a word for something in a protolanguage, then we can assume that the thing it refers to was probably either of cultural importance in the life of its speakers or environmentally salient.

A considerable amount of research has already been carried out on reconstructing the vocabulary of the Proto-Austronesian language that is the ancestor of all of the Austronesian languages spoken throughout the Pacific, as well as much of Southeast Asia. The reconstructed vocabulary for this language includes items expressing meanings such as the following:

| | | |
|-------------|---------------------------|------------------------------------|
| 'taro' | 'food garden' | 'derris poison (for killing fish)' |
| 'yam' | 'to weed' | 'high tide' |
| 'banana' | 'shoot, sucker of plant' | 'giant clam' |
| 'sugarcane' | 'wild pig' | 'seaweed' |
| 'sago' | '(of pig) root up ground' | 'conch shell' |

| | | |
|------------------|--------------------|--------------|
| 'breadfruit' | 'domestic pig' | 'fish scale' |
| 'orange' | 'canoe' | 'octopus' |
| 'pandanus' | 'sail' | 'clay pot' |
| 'betel nut' | 'sea travel' | 'shoot' |
| 'coconut' | 'paddle' | 'broom' |
| 'casuarina tree' | 'steer' | 'needle' |
| 'fallow land' | 'bail out (water)' | 'bow' |
| 'cultivate' | 'fish hook' | |

Applying this technique, the overall picture that emerges of the Proto-Austronesian-speaking society can be paraphrased as follows from the words of the Austronesian scholar Robert Blust (1995: 468 ff.):

They were settled people, occupied villages which contained some kind of public building and dwelling units, raised on posts (and thus entered by ladders), with thatched gabled roofs, internal fireplaces, and a number of storage shelves and wooden headrests. They possessed domesticated pigs, fowls and dogs. They hunted, wove, potted, used needle and thread, tattooed themselves, chewed betel nut and drank some kind of intoxicating drink... They had a well developed maritime technology, but also cultivated root crops, as well as rice and millet. They hunted heads, and used the bow and bamboo stakes in their hunting.

There is one further interesting point. For Proto-Austronesian, there are two reconstructed words for 'pig':

*babui 'wild pig'

*beyek 'domesticated pig'

Archaeological evidence indicates that there were originally no pigs in Melanesia and Polynesia. Also, the Oceanic languages only have a reconstructible word for 'tame pig', but none for 'wild pig'. This fits in nicely with the archaeological evidence, as we can conclude that it was Austronesian-speaking people who first introduced pigs into Melanesia and Polynesia. The only way to get to both of these areas from Southeast Asia is by sea, so it is logical that Proto-Oceanic would only have had a word for 'tame pig'. We would hardly expect people to have risked taking wild pigs with them in their ocean-going canoes, as wild pigs can be dangerous. Any wild pigs that we find in Melanesia and Polynesia today would therefore have to be the descendants of these original tame pigs that had escaped over the years and gone feral. Another interesting point is that, in many of the non-Austronesian languages of Papua New Guinea, the word for 'pig' seems to have been copied from forms derived from *beyek. This would be consistent with what I have just said, as there would have been no pigs at all in areas occupied by speakers of

non-Austronesian languages until they were introduced by the first speakers of Austronesian languages.

15.4.5 The Homeland of a People

In addition to giving us some ideas about the content of a protoculture, the *WÖRTER UND SACHEN* technique can also tell us something about the homeland of a language family. (Note that the original homeland of a language family is sometimes referred to in the literature by the German word *Urheimat*, from *Heimat* 'homeland', corresponding to the term *Ursprache* meaning 'protolanguage', from *Sprache* 'language'.)

From the Proto-Austronesian vocabulary that we have just examined, it is obvious that the ancestral people must have lived on an island or on the mainland very close to the sea. They clearly lived in a tropical rather than a temperate or cold environment. They lived in an area that had crocodiles, as there is a reconstructible word */*buqaja/* 'crocodile'. This fact alone rules out anywhere in Polynesia and many parts of island Melanesia as the Proto-Austronesian homeland, as these areas do not have native crocodiles. Using all of the linguistic data that we have, we can reconstruct for these people a homeland around Taiwan or southern China. We do know that around 10,000 years ago, the Chinese people pushed southward, presumably eventually pushing out the ancestors of the modern Austronesian speakers, who then spread to the Philippines and Indonesia and eventually to the Pacific area.

We can sometimes use the *Wörter und Sachen* technique to make some guesses about the actual routes followed by people in reaching their present locations. There is, for example, a word for 'owl' everywhere in Polynesia (except those areas that do not have owls). The word that we can reconstruct for this meaning in Proto-Polynesian is */*lulu/*. In Hawaii, there are owls, but the word that is used to refer to them is not a reflex of */*lulu/* but a quite different form altogether: */pueo/*. From this, scholars have argued that Hawaii might have been settled from an area where there are no native owls. One such area is the Marquesas Islands, near Tahiti. On arriving in their new home, the ancestors of the modern Hawaiians would have come across owls again, but these birds would have by then been new to them so they would have needed to find a new name.

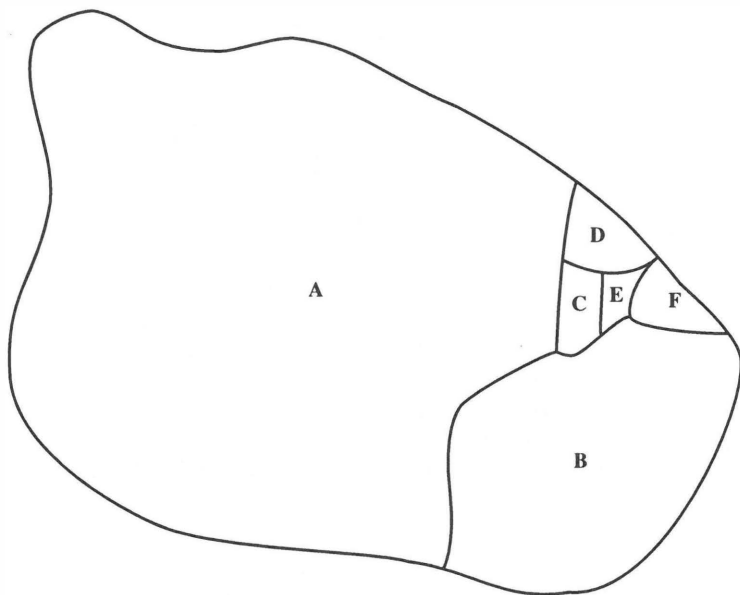
Biologists have argued that certain species of mosquitoes were spread to Polynesia by human settlement. In fact, in eastern Polynesia (Hawaii, Tahiti, and the Marquesas), the first Europeans in the area hardly noticed any mosquitoes at all. In these areas, the original word for 'mosquito', */*Jamuk/*, had taken on the new meaning of 'sandfly', which is a smaller insect but with an extremely itchy bite. In Hawaiian, the mosquito is now known by a different word altogether: */makika/* (which is possibly copied from English), and in Māori it is referred to as */waeroa/*, which literally means 'long legs'.

These facts suggest that when the Polynesians first arrived in Aotearoa and Hawaii, there were no mosquitoes there at all (or that they had come from a place where there were no mosquitoes). The original name that people knew came to refer, instead, to another small insect that also had a bite which caused itching—the sandfly. When the mosquito finally made its way into these islands (perhaps only with the arrival of the first Europeans), the people had to find a new name to refer to it, either by copying the word from English or by creating a new compound from words that already existed in the language.

Numerous examples of the same kind indicate clearly that the Māori settled the much cooler island of Aotearoa from a more tropical location. We can probably reconstruct the Proto-Polynesians as being drinkers of kava, and the word in their language for the plant was *kava. The early Polynesians had probably developed a set of fairly elaborate ceremonies associated with the drinking of kava, in contrast to those Melanesian societies further to the west where kava has probably always been drunk in a more recreational and a much less ritualized way. We can be reasonably certain both that the Māori came from a tropical area and that kava ceremonies were part of Polynesian culture when they left more than 1,000 years ago because of the existence of reflexes of the original word *kava in Māori. The kava plant only grows in tropical climates and will not grow in New Zealand. When the Māori first arrived in Aotearoa, they certainly brought with them a knowledge of this tropical plant and the ritual with which it was associated. The name of the plant was retained, but it came to apply to another plant found in New Zealand that looked similar to the original kava plant; the name was reduplicated to indicate that the first settlers recognized that it was not exactly the same plant. So in Māori today, we have the *kawakawa* plant. That kava drinking was associated with ritual when the Māori arrived is indicated by the fact that the regular reflex of *kava in Māori is /kawa/, but this word has come to refer, instead, to the sprig of any tree that is used ceremonially, as well as ceremonial protocol in general.

Another way of reconstructing the homeland of a protolanguage involves the AGE-AREA HYPOTHESIS. This hypothesis says that the area that has the greatest diversity in terms of the number of first-order subgroups is likely to be the location of the original homeland. In saying this, we are assuming the lowest number of population movements to account for the geographical distribution of the subgroups (and remember that in historical linguistics we always choose the simplest and most reasonable solution to a problem rather than a more complex one, unless there are good reasons for preferring the more complex answer). Let us take an example. Imagine that we have a language family that is divided up into a number of subgroups which are located geographically as in map 15.2.

By the Age-Area Hypothesis, the original homeland is likely to have been the area in which the subgroups B, C, D, E, and F meet. This would require that we set up only one major population shift from the original area, that of



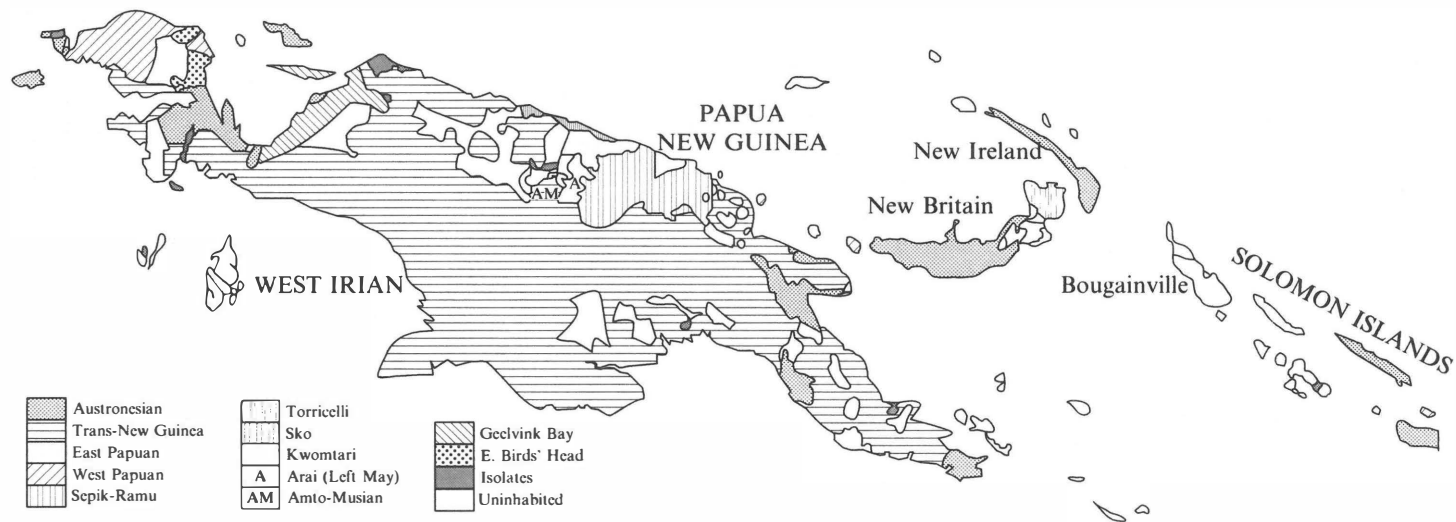
MAP 15.2 Using the Age-Area Hypothesis in an imaginary language family

the subgroup A which moved to the west. Alternatively, if we were to suggest that the area covered by A were to represent the original homeland, then we would need to argue for separate movements for the populations of B, C, D, E, and F to get to their present locations in the east.

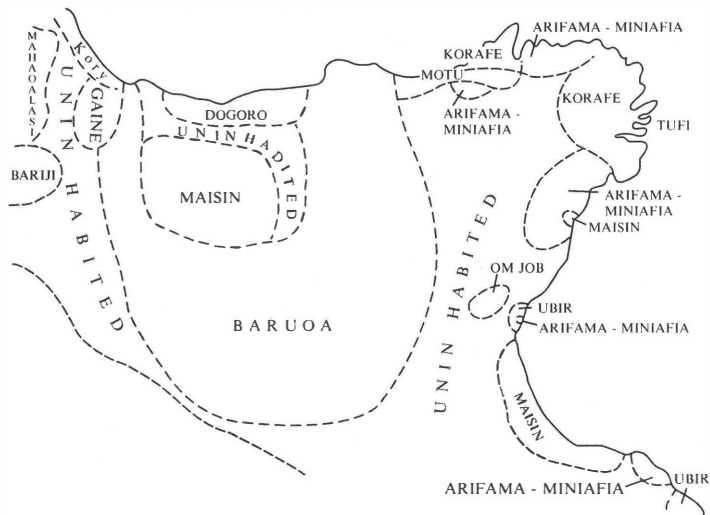
In Melanesia and the Pacific, the greatest area of subgrouping diversity in Austronesian languages is found in Melanesia rather than in Polynesia or Micronesia, and in Melanesia the greatest area of diversity is found in Papua New Guinea. This therefore suggests that the original homeland of the Oceanic languages lies somewhere in Papua New Guinea, and certainly not in the Solomon Islands, Vanuatu, New Caledonia, Fiji, Micronesia, or Polynesia.

Turning our attention now to the non-Austronesian languages of New Guinea, if we were ever able to demonstrate that these are all descended from a common ancestor (which nobody has so far been able to prove), then the most likely area of the original settlement would have been either the Sepik or the Bird's Head area of Irian Jaya, as map 15.3 indicates that these are the areas with the greatest numbers of distinct "phyla."

Sometimes we find that languages or language families are splintered, or discontinuous; that is, they are spoken in areas that do not join, and they are separated by other related languages or languages from other families. We can often consider this kind of evidence to support the idea that migrations have taken place, which result in originally contiguous groupings becoming separated. Ordinarily, we can assume that languages, or entire language



MAP 15.3 Languages of Papua New Guinea



MAP 15.4 Languages of Oro Province, Papua New Guinea

families, will occupy contiguous areas unless they are forced apart by some other factors.

The languages of the Tufi area of Oro Province in Papua New Guinea represent an interesting case of this kind of situation. In map 15.4, you can see that there are many discontinuous languages. The Maisin language is spoken in three separate areas, Notu in three areas, Korafe in two, Ubir in two, and Arifama-Miniafa in four. Apart from this, there are fairly large areas of unoccupied land in between languages. The inland Orokaiva people had a reputation traditionally of being a warlike people, and possibly what happened is that they pushed their earlier inland neighbors out of their original neighborhood into the safer uninhabited coastal areas, with the resulting very mixed-up looking linguistic map. The distribution of the languages in this area suggests that this was originally some kind of "refugee" area.

15.5 PALEOLINGUISTICS AND LANGUAGE ORIGINS

The term *PALEOLINGUISTICS* is not one that you will find in other textbooks of linguistics (as far as I know), because I made it up as I was writing this book. I created it because I felt that there was a need to talk about the reconstruction of the far distant past, beyond the time to which we have been able to reconstruct by means of the comparative method (but which nonlinguistic sciences such as archaeology can still tell us something about). The word derives, of course, from the prefix *paleo-*, which is attached to the

names of a number of scientific disciplines, and which means 'old' or 'ancient': for example, paleobiology, paleoecology, paleogeology, and paleozoology.³

Unfortunately, the comparative method of linguistic reconstruction does not allow us to go back as far in time as we would like. It is difficult to put dates to linguistic changes for which we do not have written records. It is probable that protolanguages such as Proto-Indo-European and Proto-Austronesian are not much more than 6,000 years old, certainly no older than 10,000 years old. The comparative method cannot take us further back in time for a simple reason: because languages gradually lose vocabulary over time, when they have been separated for a long period, they will have only a very small proportion of shared vocabulary. To set up systematic sound correspondences between languages, we need to have a reasonably large body of cognate items. When the corpus of shared items gets too small, we simply cannot recognize any systematic sound correspondences, at all, and without systematic sound correspondences, the comparative method becomes completely unworkable.

Archaeologists tell us that modern *Homo sapiens* (or modern humankind) is probably at least 100,000 years old. We do not know when human beings first acquired the capacity for language, but when humanity made its first major ocean crossing between Southeast Asia and what was then the continent of Sahul (which now consists of the islands of New Guinea, Australia, and Tasmania) at least 40,000 years ago (and possibly considerably more), the general assumption seems to be that those people were equipped with fully developed linguistic systems.

There are all sorts of interesting questions about language that we would like to have answers for. Did Proto-Human ever exist as a single language? If so, what was it like? Who spoke it? And where did it develop? Or did *Homo sapiens* independently develop the capacity for language in a number of different locations? If so, how many original languages were spoken at the dawn of humanity? Fascinating questions indeed, but so far not questions that we can satisfactorily answer. The limiting factor here is simply that after 100,000 years, any similarities that there might once have been between languages have been obliterated by such a long period of separation and constant linguistic evolution.

It is not just the fact that the languages themselves have been changing, either. In order to reconstruct a protolanguage, we need to have information on all of the daughter languages. If crucial features of the parent language were retained in a language for which we now have no records, then those features will be unreconstructible. Over the millennia, uncountable numbers of languages must have developed and then disappeared with no trace, for a variety of catastrophic reasons: warfare, famine, diseases, natural disasters, climate changes, and losses of territory with changing sea levels. So Proto-Human (if it ever existed) is unreconstructible not just because of limitations in the comparative method but also because we can never assemble the data that we need in order to be able to carry out such a reconstruction.

It has been recognized for a long time that the reconstruction of Proto-Human is a dead end, and as early as 1886, the prestigious Linguistic Society of Paris decreed that it would not host discussions concerning the origin of language, as it considered this pointless.

If we wanted to say anything at all about the nature of the first human language (or languages), the only possible course available to us would be to tackle the question as part of the quest for linguistic universals. A major thrust in linguistics in the past few decades has been the search for features of language that are common to all human languages. If we can establish that certain features are indeed found in all human languages, this raises the possibility that perhaps some aspects of language are "wired in" at birth as part of some kind of innate (as against learned) language capacity, and that we might even have inherited such genetic information all the way back from our ancestors who spoke Proto-Human. Obviously, this is just a theoretical possibility at the moment, and even if linguists were able to present us with some of the features of Proto-Human in this way, we would still be a long way from having reconstructed the language as such.

Even without attempting to go back as far as Proto-Human, we face severe problems when we try to link established language families further back than we have already been able to reconstruct. We know that Australia, New Guinea, and Tasmania were all settled ultimately from the same direction (i.e., from Southeast Asia), but we are unable to find any provable relationships between the languages of Australia and the languages of New Guinea and Tasmania, which were separated only when the rising oceans cut off Torres Strait (between New Guinea and Australia) about 8,000 years ago, and Bass Strait (between Tasmania and Australia) possibly 12,000 years ago. In fact, while the existence of Proto-Australian as the ancestor of all Australian languages has been widely assumed, it has never been satisfactorily proved by a rigid application of the comparative method. It has also long been known that the non-Austronesian languages of New Guinea are extremely diverse and fall into a significant number of completely unrelated language families. (Of course, the relationship of the Tasmanian languages will forever remain a mystery because the last fluent speaker of any of these languages died in 1876, and the information that was recorded on these languages before then was so poor that it is almost impossible to do anything useful with it in terms of linguistic reconstruction. We don't even know how many languages are represented in the data.)

Attempts to relate the Australian languages to languages further afield have been equally unsuccessful. For a while, scholars thought that the Dravidian languages might prove to be a good place to look, but it turned out that the similarities between the two groups of languages were too superficial to prove anything. The languages are typologically similar, but there is no evidence of systematic sound correspondence between the two families. Another scholar formulated what has come to be referred to as the **INDO-PACIFIC HYPOTHESIS**, which suggests that there is a large language family consisting of all of the non-Austronesian languages of Melanesia, Tasmania,

and the Andaman Islands (in the Bay of Bengal). This has remained nothing more than a hypothesis, and until someone can point to the existence of regular sound correspondences in any proposed sets of cognates, it is likely to continue to be regarded by mainstream historical linguists as being extremely suspect at best.

But if the Indo-Pacific languages, or just the Australian languages, or just the languages of New Guinea, do turn out to be descended from common ancestors, these ancestor languages are possibly going to be just as unreconstructible as Proto-Human, and for basically the same reasons that I have just given. It may be that we are talking about languages that go back more than 40,000 years in time. Between then and now, there have been great changes in sea level. Once, most of the ocean between Australia and New Guinea was probably occupied by people speaking an unknown number of languages, and when the sea levels rose over time, there must have been considerable realignment of occupation patterns and languages.

The archaeological evidence that I referred to in the first section of this chapter opens up a number of questions concerning the origin of the Australian languages. Are the modern Australian languages all descended from a single protolanguage that may have been spoken 40,000 (or even 60,000) years ago? Or is the spread of the modern languages much more recent? Or are some of the languages descended from an older, original language and others descended from a more recently introduced language? Until linguists are able to carry their reconstructions further back in time than we are able to do at the moment, these questions will have to remain unanswered.

The kinds of questions that I am addressing here are not restricted to a discussion of just the settlement of Melanesia, Australia, and Tasmania. Archaeologists tell us that the indigenous peoples of North and South America (now known by a variety of names, including Indians, American Indians, Native Americans, Amerindians, Eskimo, Inuit, Maya, Aztec, Inca, and so on) are all descended from people who migrated from Asia via a land bridge that once existed where the Bering Strait is now found. These migrations have not been dated with certainty, but most of the evidence so far suggests that they took place thousands of years after the migration of people into Australia and Melanesia. If the indigenous American languages were to be related in a single language family, then we would expect that the evidence for this relationship might be easier to find than with the languages of Melanesia and Australia. Unfortunately, linguists using the comparative method have been unable to come up with any reliable evidence that these languages are all descended from a common ancestor. These languages can be related into a number of separate large families, but there is, so far, no convincing evidence of any relationship further back in time between these large families.

While many attempts at palaeolinguistic comparisons fall far short of scientific respectability, the writings of Johanna Nichols since the mid-1980s have attracted considerable interest among some linguists, as well as archaeologists and others interested in establishing relationships at much greater time-depths than is possible using the comparative method. Nichols's

approach is more akin to population science in that she does not aim to study the evolution of individual languages, or even closely related groups of languages. Rather, she aims to study the history of "populations" of languages (1992). By this, she means that she considers large groupings of languages together, dealing not with particular features of individual languages but with broader general features of language groupings. Thus, she considers, for example, the languages of Australia or Africa as a whole. She pays attention not to whether structural features are present or absent, but to what are the statistical frequencies and distributions of features within these larger populations of languages.

Such linguistic markers are considered to be akin to biological markers in that they can be used to identify affinities between populations at considerable time-depths. She argues that if, in the languages of a continent (or some other large geographical area), a feature shows up with a high frequency, this distribution is not something that is due to recent diffusion. When several markers of this type are shared, this is taken as being indicative of historical affinity. Of course, such features must be known to be typologically unrelated. It would not be terribly meaningful, for example, to examine the distribution of SOV word order and postpositions as these two features tend to go hand in hand in historically unconnected languages.

She examined a sample of 174 languages, which she divided into three major geographic areas: the languages of Africa, the Middle East, northern Eurasia, and South and Southeast Asia; those of Australia, New Guinea, and Oceania; and those of the Americas. She included in her sample a significant range of the genetic variety found within each of these three areas. The kinds of linguistic features that she compares include things such as basic clause alignment (i.e., whether there is nominative-accusative or ergative-absolutive marking in the clause), the presence or absence of an inclusive/exclusive distinction in pronouns, the level of morphological complexity, whether or not inalienable and alienable possession are distinguished, and whether or not there are nominal classifiers.

The actual application and interpretation of Nichols's method is complex, and it is unlikely to become the standard model by which individual historical linguists will attempt to study linguistic relationships. However, she does draw some quite dramatic conclusions out of the data that she analyzes, and I now summarize some of her ideas.

Regarding Australia and New Guinea, she claims to have found evidence for the distribution of a number of features in some areas that are common to these languages and the western languages of her Old World grouping, such as case-marking systems, the lack of noun classes, ergativity, and lack of tonal systems. Another set of features—accusativity, the presence of noun classes, and the presence of tones—has a much narrower distribution in Australia and New Guinea. These features tend to recur in eastern Asia. Finally, she identifies features that are characteristic of the entire area of Australia and New Guinea, including relatively simple consonant inventories. From this,

she concludes that an early linguistic stratum occupied the entire continent of Sahul (which is how archaeologists refer to Australia and New Guinea together before they were separated in relatively recent times by rising sea levels). A second stratum resulted from a later linguistic colonization of the area with its greatest concentration of residual features in the northwest, which presumably represented the point of entry.

Applying the same kind of thinking at a world level, Nichols argues for a three-stage spread of human language since its origin in Africa over 100,000 years ago. Features that derive from this period, it turns out, are not discernible in modern languages, so we can only make assumptions about this period on the basis of nonlinguistic evidence. The second period of linguistic history involves a spread of languages from the Old World areas across Eurasia around the Pacific Rim and through to the Americas between about 60,000 and 30,000 years ago. This would have been the period in which the languages of Sahul arrived in what is now New Guinea and Australia. Finally, in the postglacial period, we see the development of complex and large-scale societies and the emergence of political and economic power, which has resulted in an overall reduction in linguistic diversity.

15.6 THE RELIABILITY OF CULTURAL RECONSTRUCTION

Having looked in detail at the kind of information that historical linguistics can provide about cultural history, we should ask ourselves an additional question: How *reliable* is this information, and how well does this information tie in with information provided by archaeology, oral history, and comparative culture? In general terms, what historical linguistics can tell us about cultural history depends on how we subgroup languages in a particular family and on what we reconstruct in the vocabulary of a protolanguage. Our conclusions about cultural history can therefore only be as accurate as our subgrouping and our lexical reconstruction.

You have already seen that subgrouping is not always certain. In some cases, there may be contradictory evidence when you are trying to set up subgroups, depending on what sorts of facts you choose to give more reliance to. For instance, some scholars have argued that the area of greatest subgrouping diversity within the Austronesian language family includes those Austronesian languages which are indigenous to Taiwan, off the coast of southern China. This fits in nicely with the proposition that I mention elsewhere in this chapter that the linguistic evidence suggests this part of the world as the Austronesian homeland. Linguists who make this particular subgrouping claim do so on the basis of shared grammatical and phonological innovations in the languages of Taiwan, but what we regard as a shared innovation or a shared retention depends on what we actually reconstruct in the protolanguage itself. If our grammatical or phonological reconstruction contains errors, then the

subgroupings based on those reconstructions will also be wrong. For instance, some linguists have claimed that it is in the Melanesian area that we have the area of greatest diversity in the Austronesian family (though most of these arguments have rested on lexicostatistical evidence, which you have already seen is not necessarily reliable). If this were true, then we would be speaking of a Melanesian homeland for Proto-Austronesian rather than a homeland in southern China.

Also, if our reconstruction of the content of the vocabulary of a protolanguage is inaccurate, then any statements that we make about the nature of the original culture and the original homeland may also be misleading. It is not difficult for our lexical reconstructions to be wrong, as the *Wörter und Sachen* technique that I described earlier is not completely infallible as a way of reconstructing the culture of a people in the past. While the comparative method produces fairly reliable reconstructions of the earlier forms of words, we cannot always guarantee that we have reconstructed the correct original meanings. We have already seen that semantic reconstruction cannot be carried out nearly as confidently as phonological reconstruction. For instance, the modern Algonquian languages of North America (mostly spoken in Canada) have words for 'whisky' that are compounds of 'fire' and 'water', as well as words for 'train' that are compounds of the words for 'iron' and 'horse'. By strictly applying the comparative method, it would be logically possible to reconstruct Proto-Algonquian words for both 'whisky' and 'train' that are based on these roots. Of course, we know from historical evidence that speakers of Algonquian languages came into contact with whisky and trains only with the arrival of the Europeans in the last few hundred years. These examples clearly involve parallel lexical developments, and such developments in related languages are often especially difficult to distinguish from shared innovations.

Where this technique produces cultural reconstructions that seem plausible within the bounds of what archaeologists already know, there is likely to be little significant dispute about their reliability. However, given that parallel semantic shifts can (and do) take place, it is logically possible for the *Wörter und Sachen* technique to produce archaeologically improbable protocultures. For instance, the form *tusi can be reliably reconstructed as a Proto-Polynesian word, and the reflex of this in most of the modern Polynesian languages means either 'write' or 'book' (or both). If we were to assume that this represents the original meaning of *tusi, then we would need to reconstruct Proto-Polynesian society as having been literate. At the time of European contact, however, none of these societies was literate, and there is no archaeological evidence of writing on any of the Lapita pottery (though people were decorating these pots with hand-drawn designs). We know from written records dating from the time of the early European missionaries that the modern reflexes of *tusi only came to refer to writing and books after contact with European missionaries. The original meanings of these words were more likely to have been 'make a mark', or something of that nature.

Earlier in this chapter I mentioned the work of Robert Blust in applying the *Wörter und Sachen* technique in the reconstruction of Proto-Austronesian culture. Blust concluded that iron was known by these early peoples, yet there is no archaeological support for this kind of reconstruction this early in history. Archaeologists are fairly confident that metallurgy appeared suddenly in Southeast Asia only about 2,200 years ago, which was well after the spread of the Austronesian languages had taken place. So, linguists probably have to be more careful in distinguishing between direct inheritance and parallel semantic shifts in ancient forms.

Reading Guide Questions

1. What is archaeology, and what kinds of historical information can archaeologists provide?
2. How reliable is the historical evidence provided by oral tradition? What factors influence the reliability of this kind of data?
3. What is meant by the term “comparative culture”? What kinds of historical information can it provide?
4. How can historical linguists tell us something about the relative order in which population splits take place?
5. What can we tell about the nature of cultural contact between two societies from linguistic evidence?
6. How can we tell something about the relative timing of a borrowed cultural feature from linguistic evidence?
7. What is the *Wörter und Sachen* technique of cultural reconstruction?
8. What are the problems involved in applying the *Wörter und Sachen* technique?
9. How can we make guesses about a people’s homeland and migration routes from the linguistic evidence?
10. What can historical linguistics tell us about the very ancient relationships between populations?
11. What is meant by “linguistic universals,” and what is their importance?
12. Can the existence of Proto-Human ever be demonstrated or disproven? Why?
13. What are the inherent weaknesses in cultural reconstruction?

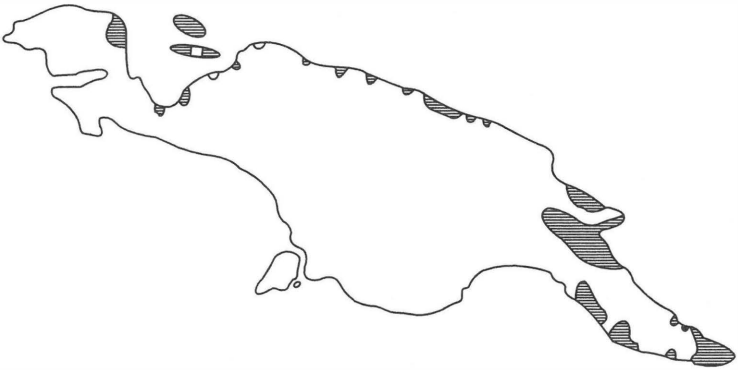
Exercises

1. In the languages of northern and central Vanuatu, the words for *kava* are derived from a form that can be reconstructed with the form *maloku. A word derived from the same original form appears in Fijian, with the meaning ‘quiet, subdued’ (which is how *kava* makes the drinker feel if it is sufficiently strong). The Polynesian languages have words derived from the reconstructed form /*kava/ to refer to the same thing. In some Polynesian languages, this word also means ‘bitter’ (which is what *kava* tastes like when

it is drunk). The languages of southern Vanuatu are not closely related to the Polynesian languages, and they have undergone many far-reaching phonological changes which generally make forms that are cognate with Polynesian words almost unrecognizable at first glance. The word for *kava* in the languages of southern Vanuatu are mostly something like /n?kava/, in which the initial syllable represents an earlier noun marker that has been reanalyzed as part of the root.

From all of this evidence, do you think that *kava* may have been discovered once, twice, or three times?

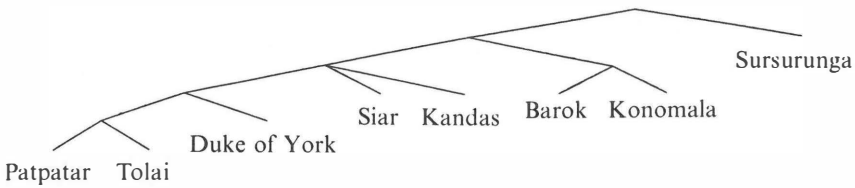
- Examine the following map showing the distribution of Austronesian languages on the mainland of the island of New Guinea. Assume that these languages originated outside of New Guinea, and say which direction you think they might have come from. Give your reasons.



- In southern New Ireland and northern New Britain in Papua New Guinea, the following languages are found:



These are all related within a single subgroup, the internal subgrouping of which (as suggested by the lexicostatistical evidence) is as follows:



Can you suggest a possible pattern of migration that is consistent with this subgrouping, and with the fact that languages in the nearest related higher-level subgroup are spoken to the immediate north of Barok? Note also that the languages spoken to the south of Tolai are completely unrelated non-Austronesian languages.

4. The following words have been reconstructed for Proto-Algonquian, the ancestor to the American Indian languages spoken in the areas shaded on the map below:

- *wɛfawɛ:ɪnɪfja 'American beech tree'
- *name:kwa 'lake trout fish'
- *a:ʧkikwa 'harbor seal'
- *atehkwa 'woodland caribou'



The following maps show the area where each of these four species is native. On the basis of this evidence, where might you suggest that the Algonquian languages originated from?



Harbour seal



Woodland caribou



American beech



Lake trout

5. In the past there have been theories that the Polynesians originated in South America and even from the islands off the coast of British Columbia. Most scholars regard such theories as belonging to the lunatic fringe. Why do you think they feel this so strongly?

Further Reading

- Lyle Campbell, *Historical Linguistics*, chapter 15, "Linguistic Prehistory," pp. 339–73.
 Theodora Bynon, *Historical Linguistics*, chapter 7, "Language and Prehistory," pp. 262–80.
 Raimo Anttila, *An Introduction to Historical and Comparative Linguistics*, chapter 21, "Change and Reconstruction in Culture and Linguistics," pp. 377–88.
 Morris Swadesh, "Linguistics as an Instrument of Prehistory."
 Donald Denoon and Roderic Lacey (eds.), *Oral Tradition in Melanesia*.
 Brian M. Fagan, *The Great Journey*.
 J. P. Mallory, *In Search of the Indo-Europeans*.
 Elizabeth J. W. Barber, *The Mummies of Ürümchi*.
 Johanna Nichols, *Linguistic Diversity in Space and Time*.
 John Mulvaney and John Kamminga, *The Prehistory of Australia*.
 Robert Blust, "The Prehistory of Austronesian-Speaking Peoples."
 Lyle Campbell, *American Indian Languages*, chapter 3, "The Origin of Native American Languages," pp. 90–106.

Datasets

The following sets of data are used in the exercises at the end of several chapters as an aid in acquiring different skills. Rather than repeat each set of data in each chapter, these Datasets are attached as an appendix, and students are referred to the Datasets by number in each particular question.

1 PALAUAN (MICRONESIA)

| | | |
|--------|-------|----------|
| *hataj | ʔað | 'liver' |
| *lajay | jarəs | 'sail' |
| *jalan | rajl | 'road' |
| *apuj | ɲaw | 'fire' |
| *mata | mað | 'eye' |
| *cinaɣ | sils | 'light' |
| *cucu | tut | 'breast' |
| *bulan | bujl | 'moon' |
| *batu | bað | 'stone' |
| *ikan | ɲikəl | 'fish' |
| *hujan | ʔull | 'rain' |
| *laɲit | jaɲəð | 'sky' |

| | | |
|--------|-------|----------|
| *buŋa | buŋ | 'flower' |
| *pəpu | wel | 'turtle' |
| *dəŋəy | reŋəs | 'hear' |

2 NGANYAYWANA (NEW SOUTH WALES, AUSTRALIA)

| | | |
|-------------|----------|-------------|
| *ŋa:nəŋ | anəŋa | 'who' |
| *wi:ŋan | igana | 'snow' |
| *ba:bəŋa | abəŋa | 'father' |
| *mi:ŋin | igina | 'star' |
| *mi:l | ila | 'eye' |
| *ga:bulga:n | abulgana | 'shark' |
| *bargan | argana | 'boomerang' |
| *winba | inba | 'fire' |
| *buruluŋ | ruluŋa | 'fly' |
| *wambupa | mbupa | 'kangaroo' |
| *bagar | gara | 'meat' |
| *ganəj | nəja | 'yam stick' |
| *dimin | mina | 'nits' |
| *guruman | rumana | 'boy' |
| *wigəj | gəjəjə | 'food' |
| *gugəŋa | gəwəŋa | 'child' |
| *gubila | bwila | 'possum' |
| *gijinma | ŋinma | 'scratch' |

3 MBABARAM (NORTH QUEENSLAND, AUSTRALIA)

| | | |
|--------|-----|------------|
| *jana | ne | 'stand' |
| *bamba | mba | 'belly' |
| *ɲaba | bo | 'bathe' |
| *wuna | no | 'lie down' |
| *jiba | be | 'liver' |
| *gumbi | mbi | 'penis' |
| *naga | ga | 'east' |
| *pulu | lu | 'he' |
| *gunda | ndo | 'cut up' |

4 YIMAS AND KARAWARI (EAST SEPIK, PAPUA NEW GUINEA)

| | Yimas | Karawari | |
|----------|-----------|----------|-----------------|
| *sikir | > tikit | sikir | 'chair' |
| *jakus | > jakut | jakus | 'string bag' |
| *samban | > tamban | samban | 'lover' |
| *panmari | > panmaɿ | panmari | 'male' |
| *sisin | > tirin | sisin | 'tooth' |
| *naniɲ | > naniɲ | janiɲ | 'fat' |
| *sambajm | > tambajm | sambajm | 'basket hanger' |
| *nawkwan | > nawkwan | jawkwan | 'chicken' |
| *nam | > nam | jam | 'house' |
| *sambin | > tambin | sambin | 'tail' |
| *simun | > timun | simun | 'cane' |
| *pariapa | > paɿapa | pariapa | 'verandah' |
| *manbaw | > manbaw | manbo | 'death adder' |
| *tumbaw | > tumbaw | tumbo | 'crocodile' |

5 LAKALAI (WEST NEW BRITAIN, PAPUA NEW GUINEA)

| | | |
|-------|-----|--------|
| *lani | ali | 'eat' |
| *ikan | ia | 'fish' |

| | | |
|---------|--------|---------|
| *lima | lima | 'hand' |
| *paʔa | vaha | 'leg' |
| *ʔate | hate | 'liver' |
| *kutu | utu | 'lice' |
| *ʔunsan | hura | 'rain' |
| *ʔanso | haro | 'sun' |
| *lipon | livo | 'tooth' |
| *danu | lalu | 'water' |
| *taŋi | tali | 'cry' |
| *tapine | tavile | 'woman' |

6 SUENA AND ZIA (MOROBE PROVINCE, PAPUA NEW GUINEA)

(These data have been slightly regularized.)

| Suena | Zia | |
|--------|--------|----------------|
| ni | ni | 'bird' |
| ŋo | jo | 'mercy' |
| wo | wo | 'meat, fish' |
| pu | pu | 'pig' |
| wa | wā | 'boat' |
| su | su | 'soup' |
| wi | wi | 'penis' |
| mu | mū | 'sap' |
| be | be | 'mouth' |
| pigi | pīgi | 'lime' |
| me | mē | 'shame' |
| ari | ari | 'vagina' |
| goroba | gorobo | 'cycad tree' |
| moka | moko | 'inside' |
| wena | weno | 'nose' |
| tuma | tumo | 'back of neck' |
| duba | dubo | 'throat' |

| | | |
|------|------|---------|
| papo | japo | 'name' |
| ema | emo | 'man' |
| me | me | 'urine' |

7 KORAFE, NOTU, AND BINANDERE (ORO PROVINCE, PAPUA NEW GUINEA)

| Korafe | Notu | Binandere | |
|--------|-------|-----------|-------------|
| poka | po | do | 'mercy' |
| poʔka | po | do | 'inside' |
| paʔka | pa | da | 'betel nut' |
| pawo | pawo | dao | 'name' |
| bipo | bi po | bido | 'banana' |
| susu | susu- | tutu | 'meaning' |
| toʔka | to | to | 'hole' |
| — | tewo | teo | 'bowl' |
| dubo | dubo | dubo | 'throat' |
| dika | di | — | 'tooth' |

8 PAAMESE (VANUATU)

| North | South | |
|-------|-------|-----------------|
| eim | aim | 'house' |
| amai | amal | 'reef' |
| ai | a:l | 'stinging tree' |
| oul | aul | 'maggot' |
| out | aut | 'place' |
| he | hel | 'step' |
| mea | mela | 'get up' |
| takul | takul | 'sago' |
| hae | hale | 'outside' |

| | | |
|--------|--------|----------|
| keil | kail | 'they' |
| teilaŋ | teilaŋ | 'sky' |
| tahē | tahel | 'wave' |
| moul | maul | 'alive' |
| mavul | mavul | 'broken' |
| houlu | haulū | 'many' |
| ateli | ateli | 'basket' |

9 MOTU (CENTRAL PROVINCE, PAPUA NEW GUINEA)

| | | |
|---------------------|--------------------|-------------------|
| *tama | tama | 'father' |
| *taŋi | tai | 'cry' |
| *tari | tadi | 'younger brother' |
| *yita | ita | 'see' |
| *yate | ase | 'liver' |
| *tina | sina | 'mother' |
| *tiavu | siahu | 'sweat' |
| *mate | mase | 'die' |
| *yutu | utu | 'louse' |
| *pune | pune | 'bird' |
| *ɔaŋi | lai | 'wind' |
| *leŋi | rei | 'long grass' |
| *bara | bada | 'big' |
| *diba | diba | 'right' |
| *geru | gedu | 'nape of neck' |
| *garo | gado | 'language' |
| *g ^w ada | g ^w ada | 'spear' |
| *lata | rata | 'milk' |
| *labia | rabia | 'sago' |
| *maɔa | mala | 'tongue' |
| *wabu | vabu | 'widow' |
| *walo | varo | 'vine' |
| *vui | hui | 'hair' |

| | | |
|---------|--------|---------|
| *vavine | habine | 'woman' |
| *api | lahi | 'fire' |
| *au | lau | 'I' |

10 SEPA, MANAM, KAIRIRU, AND SERA (COASTAL SEPIK, PAPUA NEW GUINEA)

| Sepa | Manam | Kairiru | Sera | |
|-----------|---------|---------|----------|-------------|
| tamota | tomoata | ramat | reisiouk | 'man' |
| waine | aine | mwoin | tamein | 'woman' |
| mata | mata | mata | tapuŋ | 'eye' |
| giŋa | gaŋa | kwokala | suvətəŋ | 'nose' |
| talŋo | kuŋi | tələŋa | tenepiŋ | 'ear' |
| lima | debu | kawi | ləyəŋ | 'arm, hand' |
| lulu | ruru | sus | tuit | 'breast' |
| dala | dara | sinai | tenei | 'blood' |
| ŋamali | amari | warəŋ | rau | 'sun' |
| kalewa | kalea | kaleo | bul | 'moon' |
| wabubu | rodo | abwuŋ | puiŋ | 'night' |
| ndanu | daŋ | rian | rain | 'water' |
| makasi | makasi | nau | na | 'sea' |
| partu | patu | buŋ | ak | 'stone' |
| ewa | ewa | luf | teiŋ | 'fire' |
| kai | kai | kai | ai | 'tree' |
| undu | udi | wur | bur | 'banana' |
| keu | keu | wonau | biŋ | 'dog' |
| manu | maŋ | mian | main | 'bird' |
| mota | moata | vaniu | meni | 'snake' |
| ika | ika | siasi | mwoiŋ | 'fish' |
| ŋalambuti | laŋo | ləmwok | laŋ | 'fly' |
| namu | naŋ | niam | nənei | 'mosquito' |
| pela | pera | pial | nou | 'house' |
| wawaraki | wauwau | bunbun | wuipul | 'white' |
| mbotambo | zimzimi | silsir | neknək | 'black' |

| | | | | |
|---------|--------|--------|-------------|---------|
| ndisuau | tumura | marir | marir | 'cold' |
| kani | kaŋ | an | əain | 'eat' |
| sopu | mai | miai | ma | 'come' |
| lako | lako | liak | pi | 'go' |
| teke | teke | tai | pontenen | 'one' |
| lua | rua | wulu | eltiŋ | 'two' |
| toli | toli | tuol | eltiŋ pal | 'three' |
| wati | wati | viat | eltiŋ eltiŋ | 'four' |
| lima | lima | vələri | pingari? | 'five' |

11 BURDUNA (WESTERN AUSTRALIA)

| | | |
|-------------|-----------|----------------|
| *pampura | papura | 'blind' |
| *tulupku | tułucku | 'crane' |
| *ŋaʔu | ŋaja | 'I' |
| *kawuŋka | kawuka | 'egg' |
| *kaŋʔara | kaʔara | 'root' |
| *papu | pawu | 'father' |
| *ŋampu | ŋapu | 'tree' |
| *waŋkan | waʔkan | 'chest' |
| *kuʔara | kujara | 'two' |
| *tuŋtu | tuʔu | 'narrow' |
| *mu[an]kara | mu[ack]ua | 'parrot type' |
| *cipa | ciwa | 'dive' |
| *kumpu | kupu | 'urine' |
| *puka | puwa | 'bad' |
| *kuŋʔal | kuʔal | 'daughter' |
| *ŋaŋka | ŋaʔka | 'beard' |
| *tuʔuŋkaʔi | tuʔukaʔi | 'honey' |
| *pacapuʔu | pajawuʔu | 'dangerous' |
| *mukul | mu:l | 'aunt' |
| *jimijca | jimica | 'scratch' |
| *kaŋpar | katpar | 'spider's web' |
| *puŋkuʔi | pukuʔi | 'kangaroo' |

| | | |
|-------------|----------|-----------------|
| *paʔari | pajari | 'fight' |
| *paca | paja | 'drink' |
| *ɣuŋʔa | ɣuʔa | 'lie' |
| *cuʔwa | cuwara | 'hiding' |
| *ɣuŋkun | ɣukun | 'rotten' |
| *ʔa:paca | ʔa:waja | 'wild plum' |
| *kəkul | kawul | 'testicles' |
| *parumpa | parupa | 'wattle tree' |
| *piŋʔa | piʔa | 'mud' |
| *waŋka | waka | 'speak' |
| *miniŋca | minica | 'centipede' |
| *piŋkaci | piŋkaji | 'dish' |
| *tiŋti | tiŋti | 'clitoris' |
| *jukari | juwari | 'stand' |
| *kankala | katkala | 'wild potato' |
| *jakan | ja:n | 'spouse' |
| *kucuru | kujuru | 'word' |
| *cinticinti | citijiti | 'willy wagtail' |
| *maŋʔa | maʔa | 'arm' |
| *mintulu | mitulu | 'fingernail' |
| *mika | miwa | 'back' |
| *pukura | pu:ra | 'devil' |
| *waŋʔa | waʔa | 'give' |
| *cukuʔu | cu:ɖu | 'smoke' |
| *macun | majun | 'turtle' |
| *kukulara | ku:lara | 'dove' |

12 QUÉBEC FRENCH (CANADA)

(In these examples, the symbol ɥ is used to represent a high, front unrounded glide.)

| Standard French | Rural Québec French | |
|-----------------|---------------------|------------|
| kanadjē | kanadzjē | 'Canadian' |
| | pətsi | 'small' |
| | batsy | 'beaten' |

| | | |
|-----------|-----------|----------------|
| tʃe | tsʃe | 'kill' |
| tyb | tsʏb | 'tube' |
| tip | tsɪp | 'guy' |
| tigɤ | tsɪg | 'tiger' |
| dir | dzir | 'say' |
| kʁokodil | kʁokɔdzɪl | 'crocodile' |
| dyɤ | dzyr | 'strong' |
| ēdjē | ēdzjē | 'Indian' |
| kɔ̄dʒɪɤ | kɔ̄dzʒɪɤ | 'drive' |
| avœgl | avœg | 'blind' |
| pœpl | pœp | 'people' |
| pɤɔpɤ | pɤɔp | 'clean' |
| vinegɤ | vineg | 'vinegar' |
| tabl | tab | 'table' |
| filtɤ | fɪlt | 'filter' |
| kɔ̄vĕkɤ | kɔ̄vĕk | 'convince' |
| pakt | pak | 'pact' |
| āsābl | āsām | 'together' |
| septābɤ | septām | 'September' |
| ɔ̄bɤ | ɔ̄m | 'shade' |
| ʒœgl | ʒœŋ | 'jungle' |
| lāg | lāŋ | 'tongue' |
| lādmē | lānmē | 'the next day' |
| pāɤskə | pāskə | 'because' |
| mɛɤkɤɔ̄di | mɛkɤɔ̄dzi | 'Wednesday' |
| pāɤl | pāl | 'speak' |
| tɤwa | twa | 'three' |

13 TIENE (DEMOCRATIC REPUBLIC OF CONGO)

(High tone is marked by ' and low tone by ` over the vowel.)

Common Bantu Tiene

| | | |
|--------|--------|------------|
| *-bádà | -bála | 'marriage' |
| *-bádj | -baale | 'tomorrow' |

| | | |
|---------|--------|-------------------|
| *-bàmbà | -baama | 'poisonous snake' |
| *-béédè | -bééle | 'milk' |
| *-bèng | -bɛɛ | 'become red' |
| *-cá | -sa | 'do' |
| *-cété | -sété | 'nail' |
| *-cǐ | -sí | 'inhabitant' |
| *-céndé | -síéné | 'thorn' |
| *-cíndí | -síéné | 'squirrel' |
| *-di | -le | 'be' |
| *-dí | -le | 'eat' |
| *-dǎmbò | -dieme | 'sign' |
| *-dǎngá | -dia | 'smoke' |
| *-gèd- | -kel- | 'egg' |
| *-gòndò | -gɔɔɔ | 'moon' |
| *-gǔ̀ | -kolo | 'hill' |
| *-jàdà | -zala | 'hunger' |
| *-jǐb- | -yíéb- | 'know' |
| *-kádá | -kálá | 'crab' |
| *-kín- | -kén- | 'dance' |
| *-pùtò | -fuute | 'payment' |
| *-pùd- | -fúl- | 'blow' |

14 CYPRIOT ARABIC

Data are from Borg (1985), and I have retranscribed some of his symbols using IPA. A dot under a consonant in Old Arabic denotes emphatic phonation type. (While these data are correct as far as they go, the real picture is much more complex and looking only at these data will most likely give a somewhat misleading picture about the relationship between Old Arabic and Cypriot Arabic.)

Old Arabic Cypriot Arabic

| | | |
|--------|-------|--------------|
| bayḏah | peḏe | 'egg' |
| badal | pitel | 'he changed' |
| sabt | sift | 'Saturday' |
| ṭa:b | tap | 'blows' |

| | | |
|---------|---------|---------------------------------|
| ʒipt | ʒift | 'I brought' |
| fatal | fitel | 'he twirled' |
| ʃu:f | suf | 'wool' |
| θawb | θawp | 'shift' |
| ðahr | ðaxr | 'black' |
| ða:q | tak | 'he tasted' |
| faħað | fizet | 'he begged' |
| nabi:ð | mpit | 'wine' |
| ha:ða: | aða | 'this (<i>masc.</i>)' |
| qaṭṭab | kattep | 'he startled' (OA 'he frowned') |
| sawda:ʔ | sawta | 'black (<i>fem.</i>)' |
| xafab | xafep | 'wood' |
| ʕadas | ʕates | 'lentils' |
| ħaṭab | xatap | 'firewood' |
| qamar | lamar | 'moon' |
| ħalq | xank | 'mouth' |
| maʃl | masl | 'whey' |
| ħabl | xapl | 'rope' |
| ism | ism | 'name' |
| baṭn | patn | 'belly' |
| ðakar | takar | 'male' |
| ʃuʕlu:k | saʕalúk | 'poor (<i>masc.</i>)' |
| kaʕkah | káʕake | 'round cake' |
| kalb | kilp | 'dog' |
| qalb | kalp | 'heart' |
| ra:h | rax | 'he went' |
| ħalaf | xilef | 'he swore' |
| sa:ʕah | saʕa | 'hour' |
| ʃabay | sipʕe | 'he painted' |
| ʕari:b | ʕaríp | 'foreign' |
| ʃuyl | ʃoʕol | 'work' |
| ʔahl | exl | 'parents' |
| ʔakl | ikl | 'food' |
| biʔr | pir | 'well' |

15 NYULNYULAN (WESTERN AUSTRALIA)

(Stress is on the initial syllable of all words in all languages: *rr* is a tril; *r* is a glide; *y* is used for the palatal glide. Otherwise the transcriptions are in IPA. There is no voicing contrast; *b*, *d*, and *k* are used as per the conventions of the orthographies of these languages.)

| Bardi | Nyulnyul | Nimanburru | Yawuru | Nyikina | |
|-----------|----------|------------|-----------|-----------|--------------------------------------|
| a:mba | wamb | wa:mba | wamba | wamba | 'man' |
| wuraŋ | wuriŋ | warəŋ | ʒaŋɖu | ʒaŋɖu | 'woman' |
| karrbina | — | karrbi:nə | karrbina | karrbina | 'heavy shield' |
| juŋka.miŋ | juŋk | juŋku | juŋku | juŋku | 'fire' ('firefly' in Bardi) |
| irrkili | irrkil | yirrkili | yirrakulu | yirrakulu | 'boomerang, pindan wattle' |
| ba:wa | bab | ba:bə | baba | baba | 'baby' |
| baŋjuɖu | baŋjuɖ | baŋjuɖə | baŋjuɖa | baŋjuɖa | 'fish poison' |
| bi:ni | bin | bi:nə | bina | bina | 'rotten' |
| buru | buru | buru | buru | buru | 'place, ground' |
| da:ŋku | daŋk | da:ŋku | daŋku | daŋku | 'jaw' |
| jawal | jabal | jabal | jabal | jabal | 'story' |
| ji:wa | jib | ji:bə | jiba | jiba | 'boomerang' |
| jinal | jinal | jinal | jinal | jinal | 'spear' |
| ju:rru | ju:rr | ju:rru | ju:rru | ju:rru | 'snake' ('biting insect' in Nyikina) |
| ka:ŋji | kaŋj | ka:ŋji | kaŋji | kaŋji | 'bone' |
| ku il | ku ibil | ku ibil | ku ibil | kulibil | 'turtle' |
| morr | makarr | makurr | makurr | makurr | 'road' |
| miyala | mija | mija ə | mija a | mija a | 'sit' ('be awake' in Bardi) |
| nola | nawul | nawulu | nawulu | nawulu | 'club' |
| ŋa:nka | ŋank | ŋa:nka | ŋanka | ŋanka | 'language' |
| wa: i | wa | wa: i | wa i | wa i | 'meat' |
| aŋkurr | waŋkurr | waŋkurr | waŋkurr | waŋkurr | 'tears' |
| ara | war | wara | warəŋ | warəŋ | 'other' |
| i:ndu | winduk | wi:nduw | — | winduku | 'curlew' |
| aŋki | yaŋk | aŋki | yaŋki | yaŋki | 'what' |
| i:wala | yibal | yi:ba ə | yibala | yibala | 'old man' |

16 AROMA, HULA, AND SINAUGORO
(CENTRAL PROVINCE, PAPUA NEW
GUINEA)

| Aroma | Hula | Sinaugoro | |
|--------|--------|-----------|---------------|
| pune | — | pune | 'pigeon' |
| opi | kopi | kopi | 'skin' |
| vau | vau | vatu- | 'stone' |
| — | pai | bati | 'chop' |
| ama | ama | tama | 'father' |
| ina | ina | tina | 'mother' |
| ayi- | ayi | tayi | 'cry' |
| uli | uli | tuli | 'sew' |
| inaye | inaye | tinaye | 'bowels' |
| ui | ui | tui | 'knee' |
| upu | upu | tubu | 'grandparent' |
| ia | yia | yita | 'see' |
| uu | yuu | yutu | 'louse' |
| yae | ae | yate | 'liver' |
| ulia | yulia | yulita | 'octopus' |
| laa | laa | lata | 'milk' |
| mae | — | mate | 'die' |
| nau | nau | natu | 'child' |
| yaoi | aoi | yatoi | 'egg' |
| upa | kupa | — | 'short' |
| — | kavu | kayu | 'ashes' |
| auli | kauli | kauli | 'left hand' |
| — | kopa | koba | 'chest' |
| one | — | kone | 'sand' |
| wau | kwau | — | 'tie' |
| — | kwari | kwari | 'hit' |
| wareya | kwarea | — | 'die' |
| — | kwamo | kwamo | 'cough' |
| pipiya | pipiya | bibiya | 'lip' |
| poyi | poyi | boyi | 'night' |

| | | | |
|--------|--------|--------|--------------|
| — | poka | boga | 'belly' |
| — | para | bara | 'big' |
| | kupa | guba | 'sky' |
| ripa | ripa | diba | 'right hand' |
| repa | repa | deba | 'head' |
| lapia | lapia | labia | 'sago' |
| riri | — | didi | 'finger' |
| roye | — | doye | 'back' |
| karo | karo | garo | 'voice' |
| kovu | — | goyu | 'smoke' |
| ima | yima | yima | 'hand' |
| mauli | mayuli | mayuli | 'alive' |
| manu | manu | manu | 'bird' |
| mona | mona | mona | 'fat' |
| mina | mina | mina | 'brain' |
| maa | — | mata | 'eye' |
| maḍa | maa | maa | 'tongue' |
| — | melo | melo | 'boy' |
| numa | numa | numa | 'house' |
| nivi | nivi | nivi | 'dream' |
| niu | niu | niu | 'coconut' |
| nemo | nemo | nemo | 'mosquito' |
| leyi | leyi | leyi | 'long grass' |
| ḍayi | ayi | ayi | 'wind' |
| wayi | wayi | wayi | 'wallaby' |
| meyi | meyi | meyi | 'urinate' |
| arawa | yarawa | yarawa | 'wife' |
| vane | vane | vane | 'wing' |
| vui | vui | yui | 'hair' |
| vira | vira | vira | 'how many?' |
| vue | vue | yue | 'moon' |
| vavine | vavine | vavine | 'woman' |
| vua | vua | yua- | 'fruit' |
| vonu | vonu | yonu | 'full' |
| valivu | — | valiyu | 'new' |
| lovo | lovo | loyo | 'fly' |

| | | | |
|----------|--------|----------|-------------------|
| varo | — | varo | 'plant' |
| vaivai | vaivai | — | 'flour' |
| ōara | ara | ara | 'name' |
| ōavala | avala | avala | 'wet season wind' |
| unu | yunu | yunu | 'breadfruit' |
| ulo | yulo | yulo | 'pot' |
| uria | yuria | yuria | 'betel nut' |
| yaniyani | aniani | yaniyani | 'eat' |
| — | oro | yoro | 'mountain' |
| mari | mari | mari | 'sing' |
| milo | milo | milo | 'dirty' |
| rawa | rawa | rawa | 'sea' |
| lala | laa | laja | 'sail' |
| walo | walo | walo | 'vine' |
| wai | wai | wai | 'water' |
| wapu | — | wabu | 'widow' |

17 PROTO-GAZELLE PENINSULA AND DAUGHTER LANGUAGES (NEW BRITAIN, PAPUA NEW GUINEA)

| Proto-Gazelle | Pila-Pila | Nodup | Vatom | Lunga-Lunga | |
|----------------------|-----------|--------------------|--------|---------------------|-----------------|
| *rat ^u | rat | rat ^u | rat | rat ^u | 'basket' |
| *vup ^u | vup | vuvu | vup | vuv ^u | 'fishtrap' |
| *ram ^u | ram | ram ^u | ram | ram ^u | 'club' |
| *vasian ⁱ | vaian | vaian ⁱ | vaian | vasian ⁱ | 'sling' |
| *saman ⁱ | aman | aman ⁱ | aman | saman ⁱ | 'outrigger' |
| *pal ⁱ | pal | pal ⁱ | pal | pali | 'house' |
| *liplip ⁱ | liplip | livilivu | liplip | — | 'fence' |
| *pem ^u | pemu | pem ^u | pem | pem ^u | 'axe' |
| *pisa | pia | pia | pia | pisa | 'ground' |
| *tirip ^u | tirip | tirivu | tirip | tiriv ^u | 'green coconut' |
| *kabaŋ ⁱ | kabaŋ | kabaŋ ⁱ | kabaŋ | kabaŋ ⁱ | 'lime' |
| *up ^u | up | uvu | — | uv ^u | 'yam' |

| | | | | | |
|-----------------------|--------|---------------------|--------|----------------------|----------|
| *talisa | talia | talia | talia | talisa | 'nut' |
| *papi | pap | pavu | pap | — | 'dog' |
| *taŋis ⁱ | taŋi | taŋi | taŋi | taŋis ⁱ | 'cry' |
| *iap ⁱ | iap | iavu | iap | iav ⁱ | 'fire' |
| *mulisi | muli | muli | muli | mulisi | 'orange' |
| *beso | beo | beo | beo | beso | 'bird' |
| *lisi | li | lia | li | lisi | 'nits' |
| *sikilik ⁱ | ikilik | ikilik ⁱ | ikilik | sikilik ⁱ | 'small' |
| *tas ⁱ | ta | tai | ta | tas ⁱ | 'sea' |

Notes

Chapter 1

1. [Campbell and Poser (2008: ch 3) have a lot of information about Sir William Jones and his role in establishing historical linguistics. They point out that while he is often credited with the ideas of modern historical linguistics, in fact he is building on a tradition that was already aware of much that is attributed to Jones himself. I have modified Crowley's text here a little, but I leave the quotation from Jones because of the role it has played in teaching modern historical linguistics.—C.B.]

2. There is one sense in which we can say that Latin is a "dead language." In medieval and Renaissance Europe, the language of international scholarship and education was Latin, which was based on the written classical varieties of the language that was spoken during the heyday of the Roman empire, about 2,000 years ago. After the 1600s, written Latin became more and more rare as the local vernaculars (English, French, German, Dutch, Italian, etc.) replaced Latin to the point where Latin is now used only as an official language of the Roman Catholic Church for certain religious functions (and there is a continuing trend away from Latin in the church as well). While spoken Latin did not die, we could argue that the situation with regard to the written language is somewhat different.

3. Another area of possible substrate influence in France is the counting system. French has vestiges of a sexigesimal (base-sixty) system, unlike most

of the other languages it is related to. This may reflect influence from an earlier population whose counting system was adopted by speakers of Late Latin.

4. Here the C stands for any consonant, the V for any vowel.

Chapter 2

1. In the study of the history of languages, the symbol * is used to mark a form that has never actually been heard or written, but which is inferred or RECONSTRUCTED in a protolanguage on the basis of evidence that is available. In chapter 5, we look at how we arrive at such reconstructions.

2. For example, this change has also happened in the history of Japanese.

3. There are many versions of the sonority hierarchy, but they differ in minor ways from one another. The major article for the identification of sonority as an important tool in linguistics is by Clements (1988, 1990).

4. In this example, the reconstructions are Proto-Oceanic. R here is not an IPA symbol; it's a cover term for a sound whose exact pronunciation was not known, but it was possibly a uvular fricative [ʀ].

5. The reconstructions here are to pre-Ambrym, not to Proto-Oceanic.

6. We've reconstructed the earlier forms using data from other languages. There's more information about how this is done in chapter 5.

7. You might be familiar with a similar process in synchronic phonology called the OCP, or Obligatory Contour Principle.

8. Although anaptyxis and epenthesis are given here as synonymous, in the literature of historical linguistics there is some variation in the way that these terms are used. Some writers use the term epenthesis as a cover term for excrescence, anaptyxis, and prothesis together, while others prefer epenthesis to anaptyxis when referring specifically to the insertion of a vowel between two consonants occurring in a consonant cluster.

9. This example comes from Hock (1991: 123).

10. The idea of features in phonology goes back to Noam Chomsky and M. Halle (1968). Since then there has been a lot of work on phonology which recognizes the limitations of describing sounds in terms of binary features (noting that many of the features given here as binary are, in fact, gradient: for example, voicing is highly variable in various languages). However, features are still a useful way of conceptualizing changes like fusion.

11. The reconstructions are "pre-Greek"—that is, a level intermediate between Proto-Greek and Proto-Indo-European.

12. The reconstructions are taken from Fortson (2004: 299).

13. There is more information on the dialects of American English in Labov et al. (2006).

14. There are a number of other changes like this. You will see some in chapter 4.

15. The symbol $\underset{\cdot}$ under a letter tells you that the nasal is acting as a nucleus to a syllable. In the Indo-European literature, syllabic consonants are marked by $\underset{\cdot}$ but here we're using the IPA symbol.

16. There were some other intermediate changes here, such as the raising of [o] to [u] (kentom > kentum) and the weakening of the final syllable.

17. Cheyenne and Arapaho, other Algonquian languages, also show sound changes that are rare from the point of view of other families.

18. For more discussion of unusual sound changes in Oceania languages, see Blust (2005).

Chapter 3

1. This used to be more common, but since the rise of Optimality Theory and non-rule-based approaches to synchronic phonology, students tend not to be familiar with this area of linguistics.

2. There is a more elegant way to write this rule. We could write it $*[\text{voiced}] > [\text{nasal}] / [\text{nasal}] (\text{C}) _$. However, this rule is much more general than that given in the text, as it applies to *any* voiced segment, not just vowels and voiced stops. We have no data here for voiced consonants other than stops. Further data would be necessary to evaluate which rule is correct.

3. The reconstructions are proto-Polynesian.

4. Data are from Lincoln (1976). In earlier editions, the Banoni data included an alveolar affricate [ts]; however, this is a spelling convention in Lincoln (1976) for [tʃ].

5. Another way to write this rule for vowel harmony would be the following: $*\emptyset > V_\alpha / V_\alpha C _ \#$. That is, in the relevant environment, a vowel of the same type is added.

Chapter 4

1. If you haven't taken a phonology class, you might not be familiar with some of the terms used in this chapter, such as PHONEME and ALLOPHONE. In that case, you should look through Katamba (1993) or Clark and Colin Yallop (1995) to remind yourself about these terms (or to get more information about what they mean).

2. Although the Motu spelling system distinguishes *s* from *t*, this is only because the European missionaries who devised the spelling system were not familiar with the concept of the phoneme, and simply assumed that because [s] and [t] need to be distinguished in English, they should also be distinguished in Motu.

3. The ancestor language of many of the modern languages of Europe and South Asia.
4. You have already seen examples of changes of this type, for example, in sec. 2.8.
5. This is in dialects of English which have lost /r/ in postvocalic position.
6. In some other dialects, such as stereotypical “Long Island” English, the [g] has been reintroduced.

Chapter 5

1. These languages are not the only members of the Polynesian subgroup of Austronesian, so unless we consider data from more languages, we won't be sure that we are reconstructing to Proto-Polynesian, and not to some other intermediate ancestor.

2. This goes back to a principle in science known as Ockham's Razor. In the case of historical reconstruction, this means that you should only posit changes for which you have evidence (albeit indirect evidence sometimes), and you should not posit needless steps. For example, a hypothesis of *s > h is strictly better than one of *s > *z > ʒ > fi > h, if there is no evidence for the intermediate stages. In this case, the more complex solution might even be falsifiable, since you might have evidence that *z didn't change at all!

3. There is, of course, a trap in making this assumption. It might be that a change has occurred in all the languages, and therefore reconstructing the same sound projects a change too far back. Sometimes there will be evidence that a change has occurred in all languages. For example, the resulting phoneme system might be lopsided (which would give you evidence that something is “missing”; one phoneme might have a much more frequent distribution than others of its class (implying that it is the result of a merger), or you might have evidence from the surrounding sounds, for which see below.

4. The principle of MAJORITY RULES, as this is called, only works in some circumstances. However, we need to talk about subgrouping before we can explain about the circumstances where reconstructing the form with the widest distribution will lead you astray. (Chapter 6 is all about subgrouping.) For now, just be aware that there are situations where “majority rules” will be misleading and that it should be only one of a set of factors that you consider when deciding on your reconstructions.

5. There are good reasons for this in language processing. It is more economical.

6. A capital letter here indicates that we are unsure of the realization of the sound. Capital letters are also used in reconstruction in historical linguistics for situations where a sound of indeterminate value is reconstructed—for example, if we know that the reconstructed phoneme was probably a type of

lateral sound, but we aren't sure whether it was palatal or alveolar. This is a little different from the use of capitals here, where the capital indicates that we aren't sure which correspondence set the word belongs to.

7. Recall that the capital C stands for any consonant, and the capital V for any vowel.

8. [Tauberschmidt (2005) gives the Sinaugoro word for 'mother' as *sina*. I have left Crowley's data since Sinaugoro seems to be dialectally complex.—CB].

9. These procedures are adapted from a lecture handout originally written by Harold Koch.

Chapter 6

1. The transcription here is IPA, not the orthographic spelling (or transliteration) of the numbers.

2. Lyle Campbell has done considerable work in making these arguments; some of his papers are listed in the further reading section. I also have some discussion of language, culture, and the earliest human language in sec.15.5.

Chapter 7

1. There are other things going on in some of these words, but ignore that. There are other patterns with \bar{e} and \bar{o} , too, but for now we will consider a subset of the data. For the first three verbs, the form in the first column is the first-person singular present, the second form is the first-person singular perfect, and the third form is the first-person singular aorist. For the other words, the forms are different derivational forms or different case forms.

Chapter 8

1. Various quantitative methods are used in work on language variation and change. Such work involves not only quantifying the features that may vary between speakers (such as the length of voice onset time) but also codifying the degree of difference between the varieties of the language being compared. Dialectometry, for example, is a method for quantifying the overall difference between two speech varieties (it is usually used between geographical dialects of language). We will not be discussing these methods any further here, however.

2. The exact rate of change does not really matter for lexicostatistics, only that the rate is approximately constant.

3. There is some discussion of the length of the word list and whether it matters in McMahon and McMahon (2006).

4. [Earlier editions of this book provided a guide on how to calculate time depth using this formula. I have condensed the section and left out these instructions because glottochronology has been comprehensively discredited.—C.B.]

5. There has been some work on using NeighborNets and Swadesh (basic vocabulary lists), but thus far such work has produced the same results as lexicostatistical classifications and preliminary classifications. While such work is important for illustrating the uses of such methods, it has not, to be blunt, told us anything about linguistic relationships that we didn't already know. McMahon and McMahon (2003) provide some discussion of this topic.

6. In this section I give an overview of the methods and concepts used rather than detailed discussion of the mathematics behind the methods or explicit instructions about how to make the calculations. The further reading at the end of this chapter contains suggestions for further information beyond the general introduction I am giving here. Bryant et al. (2005) and Johnson (2008: 182–215) are the clearest introductions, and further reading can be found in the references in those papers.

Chapter 9

1. Campbell and Poser (2008) have extensive discussion of the work of this time and its contribution to later historical linguistics, including the work of Gesner, Paris, and others.

2. The sounds that are represented by the digraphs *bh*, *dh*, and *gh* in Sanskrit and by *ph*, *th*, and *kh* in Greek are voiced and voiceless aspirated stops, respectively.

3. 'Thorp' is an old word in English for 'village', but now it occurs only in place names, such as Marblethorpe, Scunthorpe, and so on.

4. Another problem for the strict application of the comparative method is analogy, for which see secs. 12.4.2 and 10.2.

5. Some grammatical conditioning can turn out to be phonological conditioning. For example, a clitic might have created an environment for a sound change which applied (or did not apply); if that clitic were later lost, we would have a change which appears to be conditioned by word class.

Chapter 10

1. if you have not taken a class in morphology, I recommend having a look at Haspelmath (2002) or Bauer (2003) so you can become more familiar with the key concepts that we will be talking about in this chapter.

2. A great deal of this section is based on Koch (1996), but I have condensed his discussion and amalgamated some of his categories.
3. The data here are from Comrie (1989: 89ff).

Chapter 11

1. Incidentally, the Japanese word for ‘thank you’—*arigato*—was borrowed, too, from Portuguese *obrigado*.
2. The technical term for this is “necronym taboo.”
3. Some of these words exist in other varieties of English, of course, but this process is much more productive in Australia and New Zealand than it is in other dialects.

Chapter 12

1. For more information and examples, see Hamel (1994).
2. The following argument is taken from Bowern (2008).
3. The mechanisms in this section are originally from Harris and Campbell (1995).
4. This example comes from Harris and Campbell (1995: 100–101).

Chapter 13

1. [Linguists differ in how they treat such exceptions; most linguists would recognize exceptions like these without necessarily positing a parallel phonemic system.—CB]
2. The figures presented here are from the so-called fourth floor study (Labov 1972), and the token that participants were asked to say was “fourth floor,” not *car*.
3. Somewhat confusingly, the term “(lexical) diffusion” is also used for the spread of changes across language and dialect boundaries (as you saw in sec. 9.4), as well as for the spread of a sound change through the lexicon of an individual language, variety, or speaker.
4. There is considerable debate on the extent to which diffusion is a mechanism in language change. See Phillips (2006) and Labov (2007) for some discussion.

Chapter 14

1. The notion of an “official” language of a country is a very recent concept in the history of languages, and while these days we often think

of boundaries between languages like boundaries between countries, many multilingual societies are either contained within nation borders or crosscut them.

2. Another version of this joke has “American” as the punch line.

3. Exactly the same change has happened in the Yolŋu Matha languages of northern Australia, where loans from the Makassar language do not take agreement morphology. For more information, see, for example, Beulah Lowe (1960).

4. Examples of the former arguments include Dion and Poplack (2007) and King (2000), while examples based on the latter include Aikhenvald (2004) and the essays in Aikhenvald and Dixon (2006).

5. In early work on pidgin and creole languages, it was thought that pidgins developed into creoles when the children of pidgin speakers started learning the language as a first language. It is now clear, however, that there is much more to creole formation than this. Creoles can form without a prior pidgin stage, although pidgins can also form part of the input to creoles.

6. This language was also called Police Motu in the earlier literature.

7. There are some problems with this idea, for a couple of reasons. For example, since creoles (like pidgins) have the majority of their vocabulary from one language, it follows that the creators of the creole have access to that vocabulary. The relationship between pidgins and creoles is considerably more complex than is presented here.

8. Most of the discussion here is based on Thomason (1997, 2001); the Michif data are from Bakker and Papen (1997).

9. [In the third edition of this book, Terry Crowley set up esoterogeny and exoterogeny as a challenge to the comparative method and family-tree model, and suggested that the family-tree model should be abandoned because it could not model such languages. I have revised this section because I think it is important to see the representational questions about language relationship as separate from the study of change. Furthermore, I do not reject the family tree as one model among several for representing language diversification; it is not the only way that languages diversify, but it is a prominent one in the history of the world. Rather than putting Tok Pisin and other languages in the “too hard” basket, as Crowley suggests is done by people who don’t wish to adopt eso/exoterogeny, I’d suggest we might need more than one type of model.—C.B.]

10. Many Māori see the possible loss of their language as a threat to their cultural identity and are taking steps to ensure that the language does not disappear. Older speakers of Māori are now being involved in special childcare centers and preschools known as “kōhanga reo” (literally: ‘language nest’) in which only Māori is used. Thousands of children are now growing up as fluent speakers of Māori. A monolingual Māori dictionary was recently published.

Chapter 15

1. Some have suggested that the European tradition of the unicorn derives from early accounts of the rhinoceros, when few Europeans had actually traveled to Africa.

2. Speakers of modern varieties of Melanesian Pidgin have over the past few decades become increasingly aware of the English sources of these words, and these kinds of distinctions are becoming rare. Many people now use purely descriptive labels to refer to people, which do not assign particular status to either race. So, *masta* becomes *waitman*, *boi* becomes *blakman*, and the word *masta* has become a neutral term meaning 'boss', whether European or Melanesian.

3. [The term "paleolinguistics" is now used often, although not always in the sense that Terry Crowley coined it. It is also used to mean 'linguistic prehistory' in general; that is, using language to make inferences about the past beyond the written record.—C.B.]

Language References

Following is a list of languages used as problems or to illustrate major points in the text. As stated at the beginning of the book, I have avoided quoting the sources of my information in the text to avoid creating a less readable and overly academic style. The list of languages below indicates the main sources of the information used. Sources without dates indicate personal communication [to Terry Crowley] or untitled and unpublished notes, while sources listed without names indicate Crowley's own fieldnotes or general knowledge. [CLB indicates Bower's fieldnotes.]

| | |
|------------------|------------------------|
| Abau | Bailey (1975) |
| Afrikaans | Burgers (1968) |
| Alamblak | Bruce (1979) |
| Algonquian | Arlotto (1972) |
| Angkamuthi | Crowley (1983) |
| Arifama-Miniafa | Lynch (1977b) |
| Aroma | Ross (1988) |
| Attic Greek | Cowan (1971) |
| Bahasa Indonesia | |
| Bandjalang | Crowley (1978) |
| Banoni | Lincoln (1976) |
| Bardi | Bower (2004), CLB |
| Binandere | Farr and Larsen (1979) |

| | |
|-----------------------|---------------------------------------|
| Bislama | Crowley (1995) |
| Burduna | Austin (1981) |
| Cypriot Arabic | Borg (1985) |
| Djambarrpuyŋu | CLB |
| Dusun | D. J. Prentice |
| Dutch | |
| Dyirbal | Dixon (1972) |
| Enggano | |
| Fijian | Capell (1973), Schütz (1985) |
| Canadian French | Walker (1984) |
| Futuna East | |
| Futuna West | |
| Gamilaraay | Austin et al. (1980) |
| Georgian Alice Harris | |
| German | |
| Gothic | Bloomfield (1967) |
| Græek | Bloomfield (1967) |
| Gumbaynggir | Eades (1979) |
| Hawaiian | William A. Foley |
| Hiri Motu | |
| Hula | Ross (1988) |
| Huli | Brian Cheetham |
| Icelandic | Cowan (1971) |
| Idam | Bailey (1975) |
| Ilokano | Bloomfield (1967) |
| Italian | Cowan (1971), Arlotto (1972) |
| Jajgir | Crowley (1979) |
| Kabana | Thurston (1987) |
| Kairiru | Laycock (1976) |
| Kara | Baumont (1981) |
| Karawari | William A. Foley |
| Karnic languages | CLB |
| Kaytetye | Koch (1996) |
| Koiari | Tom Dutton |
| Koita | Tom Dutton |
| Korafe | Lynch (1977b), Farr and Larsen (1979) |
| Kuman | |
| Kwaio | Keesing (1975) |
| Lakalai | Johnston (1978) |
| Lardil | Dixon (1980) |
| Latin | Bloomfield (1967), Arlotto (1972) |
| Lenakel | Lynch (1977a) |
| Lezgian | Schulze (forthcoming) |
| Māori | Williams (1985), William A. Foley |
| Maisin | Lynch (1977b) |

| | |
|------------------|--|
| Manam | Laycock (1976) |
| Manga | Hooley (1971) |
| Mapos | Hooley (1971) |
| Marshallese | John Lynch |
| Mbabaram | Dixon (1980) |
| Mekeo | Ross (1988) |
| Moriori | King (1989) |
| Motu | |
| Mountain Koiari | Tom Dutton |
| Mpakwithi | Crowley (1981) |
| Murut | D. J. Prentice |
| Ndao | Walker (1980) |
| Nganyaywana | Crowley (1976) |
| Notu | Lynch (1977b), Farr and Larsen (1979) |
| North Sarawak | Blust (2002) |
| Nyikina | Stokes (1982) |
| Nyulnyulan | CLB |
| Old English | Arlotto (1972) |
| Orokaiva | Lynch (1977b) |
| Paamese | Crowley (1982) |
| Palauan | William A. Foley |
| Patep | Hooley (1971) |
| Rarotongan | William A. Foley |
| Romanian | Cowan (1971) |
| Rotuman | |
| Samoan | Marsack (1973), William A. Foley |
| Sanskrit | Bloomfield (1967) |
| Sawu | Walker (1980) |
| Sepa | Laycock (1976) |
| Sera | Laycock (1976) |
| Sinaugoro | Ross (1988), Tauberschmidt (2005) |
| Sissano | Laycock (1973) |
| Southeast Ambrym | Parker (1970) |
| Spanish | Bloomfield (1967) |
| Suau | |
| Suena | Farr and Larsen (1979) |
| Tagalog | Dempwolff (1934–1938) |
| Tahitian | Clark (1979) |
| Tiene | Ellington (1977) |
| Toba Batak | Dempwolff (1934–1938) |
| Tok Pisin | |
| Tolai | |
| Tongan | William A. Foley |
| Trukese | |
| Turkic languages | Johanson and Csató (1998), Öztopçu et al. (1996) |

| | |
|--------------|---|
| Turkish | Lewis (2000), CLB |
| Turkmen | Johanson and Csató (1998) |
| Ubir | Lynch (1977b) |
| Udi | Harris (2000; 2002), Alice Harris (pc to CLB) |
| Uradhi | Crowley (1983) |
| Wagau | Hooley (1971) |
| Wallisian | |
| Wampar | Holtzknecht (1989) |
| Wiradjuri | Austin et al. (1980) |
| Yakut | Krueger (1962), Straughn (2006) |
| Yandruwandha | Breen (2004) |
| Yawuru | CLB |
| Yimas | William A. Foley |
| Yuwaaliyaay | Austin et al. (1980) |
| Zia | Farr and Larsen (1979) |

References

- Aikhenvald, Aleksandra (2004). *Language Contact in Amazonia*. Oxford: Oxford University Press.
- Aikhenvald, Aleksandra Y. and R. M. W. Dixon (eds.) (2006). *Grammars in Contact: A Cross-Linguistic Typology*. Oxford: Oxford University Press.
- Aitchison, Jean (2001). *Language Change: Progress or Decay?* 3rd ed. Cambridge: Cambridge University Press.
- Allan, Scott (1989). Review of *Introduction to Historical Linguistics*. *Te Reo* 32: 95–99.
- Anderson, Henning (1980). Morphological change: toward a typology. In J. Fisiak (ed.), *Historical Morphology*, pp. 1–50. New York: Mouton.
- Anderson, Stephen (1988). Morphological change. In F. Newmayer (ed.), *Linguistics: The Cambridge Survey*, pp. 111–35. Cambridge: Cambridge University Press.
- Anttila, Raimo (1972). *An Introduction to Historical and Comparative Linguistics*. New York: Macmillan.
- Appel, René, and Pieter Muysken (1987). *Language Contact and Bilingualism*. London: Edward Arnold.
- Arlotto, Anthony (1972). *Introduction to Historical Linguistics*. Boston: Houghton Mifflin.
- Austin, Peter (1981). Proto-Kanyara and Proto-Mantharta historical phonology. *Lingua* 54: 295–333.
- Austin, Peter (1990). Classification of Lake Eyre languages. *La Trobe University Working Papers in Linguistics* 3: 171–201.
- Austin, Peter, Corinne Williams, and Stephen Wurm (1980). The linguistic situation in north central New South Wales. In B. Rigsby and P. Sutton (eds.), *Contributions to Australian Linguistics* (Series A, No. 59), pp. 167–80. Canberra: Pacific Linguistics.

- Bailey, D. A. (1975). The phonology of the Abau language. In *Abau Language, Phonology and Grammar*, Ukarumpa: Summer Institute of Linguistics, pp. 5–58.
- Bakker, Peter, and Richard Papi (1997). Michif: a mixed language based on Cree and French. In Sarah Thomason (ed.), *Contact Languages: A Wider Perspective*, pp. 295–364. Amsterdam: John Benjamins.
- Barber, Elizabeth J. W. (1999). *The Mummies of Ürümchi*. London: Macmillan.
- Bauer, Laurie (2003). *Introducing Linguistic Morphology*. Washington, D.C.: Georgetown University Press.
- Beaumont, C. H. (1981). *The Tigak Language of New Ireland*. (Series B, No. 58). Canberra: Pacific Linguistics.
- Blevins, Juliette (2003). The phonology of Yurok glottalized sonorants: Segmental fission under syllabification 1. *International Journal of American Linguistics* 69(4): 371–96.
- Blevins, Juliette, and Andrew Garrett (1998). The origins of consonant-vowel metathesis. *Language* 74(3): 508–56.
- Bloomfield, Leonard [1933](1967). *Language*. London: Allen and Unwin.
- Blust, Robert (1995). The prehistory of the Austronesian-speaking peoples: A view from language. *Journal of World Prehistory* 9(4): 453–510.
- Blust, Robert (1996). Review of *Introduction to Historical Linguistics*. *Oceanic Linguistics* 35/2:328–35
- Blust, Robert (2002). Kiput historical phonology. *Oceanic Linguistics* 41(2): 384–439.
- Blust, Robert (2005). Must sound change be linguistically motivated? *Diachronica* 22(2): 219–69.
- Borg, Alexander (1985). *Cypriot Arabic*. Monburg: Deutsche Morgenländische Gesellschaft.
- Bowern, Claire (2004). Bardi verb morphology in historical perspective. Ph.D. dissertation, Harvard University, Cambridge, Massachusetts.
- Bowern, Claire (2008). The diachrony of complex predication. *Diachronica* 25(2): 161–85.
- Bowern, Claire (forthcoming). Historical linguistics. In *Cambridge Encyclopedia of the Language Sciences*.
- Bowern, Claire, and Harold Koch (2004). Introduction: subgrouping methodology in historical linguistics. In Claire Bowern and Harold Koch (eds.), *Australian Languages: Classification and the Comparative Method*, Current Issues in Linguistic Theory, vol. 249, chap. 1, pp. 1–16. Amsterdam: John Benjamins.
- Breen, Gavan (2004). Evolution of the verb conjugations in the Ngarna languages. In Claire Bowern and Harold Koch (eds.), *Australian Languages: Classification and the Comparative Method*, Current Issues in Linguistic Theory, vol. 249, chap. 10, pp. 245–94. Amsterdam: John Benjamins.
- Britain, David (2001). Space and spatial diffusion. *Handbook of Language Variation and Change*, pp. 603–637. Oxford: Blackwell.
- Bruce, L. (1979). A grammar of Alambak (Papua New Guinea)'. Ph.D. dissertation, Australian National University, Canberra.
- Bryant, David, F. Filimon, and Russell Gray (2005). Untangling our past: languages, trees, splits and networks. In R. Mace, C. Holden, and S. Shennan (eds.), *The Evolution of Cultural Diversity: Phylogenetic Approaches*, pp. 69–85. London: UCL Press.

- Burgers, M. P. O. (1968). *Teach Yourself Afrikaans*. New York: David McKay.
- Bynon, Theodora (1979). *Historical Linguistics*. Cambridge: Cambridge University Press.
- Campbell, Lyle (1997). *American Indian Languages: The Historical Linguistics of Native America*. New York: Oxford University Press.
- Campbell, Lyle (1999). *Historical Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- Campbell, Lyle (2003). Beyond the comparative method? In Barry Blake and Kate Burridge (eds.), *Selected Papers from the Fifteenth International Conference on Historical Linguistics*, Amsterdam: John Benjamins, pp. 33–58.
- Campbell, Lyle, and William Poser (2008). *Language Classification: History and Method*. Cambridge: Cambridge University Press.
- Capell, A. (1973). *A New Fijian Dictionary*. Suva: Government Printer.
- Chambers, J. K., Peter Trudgill, and Natalie Schilling-Estes (eds.) (2001). *Handbook of Language Variation and Change*. Oxford: Blackwell.
- Chomsky, Noam (1965). *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.
- Chomsky, Noam, and M. Halle (1968). *The Sound Pattern of English*. New York: Harper and Row.
- Clark, J. E., and Colin Yallop (1995). *An Introduction to Phonetics and Phonology*. Oxford: Blackwell.
- Clark, Ross (1979). Language. In Jesse D. Jennings (ed.), *The Prehistory of Polynesia*, pp. 249–70. Cambridge: Harvard University Press.
- Clements, G. Nick (1988). The sonority cycle and syllable organization. In Wolfgang U. Dressler, Hans Luschützky, Oskar Pfeiffer, and John Rennison (eds.), *Phonologica 1988: Proceedings of the 6th International Phonology Meeting*, pp. 63–76. Cambridge: Cambridge University Press.
- Clements, G. Nick (1990). The role of the sonority cycle in core syllabification. In J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology 1: Between the Grammar and Physics of Speech*, chap. 17, pp. 283–333. New York: Cambridge University Press.
- Comrie, Bernard (1989). *Language Universals and Linguistic Typology*. 2nd ed. Oxford: Blackwell.
- Cowan, William (1971). *Workbook in Comparative Reconstruction*. New York: Holt, Reinhart and Winston.
- Crowley, Terry (1976). Phonological change in New England. In R. M. W. Dixon (ed.), *Grammatical Categories in Australian Languages*. Canberra: AIAS.
- Crowley, Terry (1978). *The Middle Clarence Dialects of Bandjalang*. Canberra: Australian Institute of Aboriginal Studies.
- Crowley, Terry (1979). Yaygir. In R. M. W. Dixon and Barry J. Blake (eds.), *Handbook of Australian Languages*, Vol. 2, pp. 146–94. Canberra: Australian National University Press.
- Crowley, Terry (1981). The Mpakwithi dialect of Anguthimri. In R. M. W. Dixon and Barry J. Blake (eds.), *Handbook of Australian Languages*, Vol. 1, pp. 363–84. Canberra: Australian National University Press.
- Crowley, Terry (1982). *The Paamese Language of Vanuatu*. Canberra: Pacific Linguistics.

- Crowley, Terry (1983). Uradhi. In R. M. W. Dixon and Barry J. Blake (eds.), *Handbook of Australian Languages*, Vol.3, pp. 306–428. Canberra: Australian National University Press.
- Crowley, Terry (1995). *A New Bislama Dictionary*. Suva: Institute of Pacific Studies and Pacific Languages Unit, University of South Pacific.
- Crystal, David (2000). *Language Death*. Cambridge: Cambridge University Press.
- Dempwolff, Otto (1934–1938). *Vergleichende Lautlehre des Austronesischen Wortschatzes*. Berlin: Reimer.
- Denoon, Donald, and Roderic Lacey (eds.) (1981). *Oral Tradition in Melanesia*. Port Moresby: University of Papua New Guinea and Institute of Papua New Guinea Studies.
- Dion, Nathalie, and Shana Poplack (2007). *Confronting Synchrony with Diachrony in the Study of Linguistic Change: The 18th International Conference on Historical Linguistics (ICHL)*. Université du Québec à Montréal.
- Dixon, R. M. W. (1972). *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.
- Dixon, R. M. W. (1980). *The Languages of Australia*. Cambridge: Cambridge University Press.
- Dixon, R. M. W. (1994). *Ergativity*. Cambridge: Cambridge University Press.
- Dixon, R. M. W., and Barry Blake (1979). *Handbook of Australian Languages*, vol. 1. Canberra: Australian National University Press.
- Durie, M., and M. Ross (eds.) (1996). *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*. New York: Oxford University Press.
- Eades, Diana (1979). *Gumbaynggir*. Canberra: Australian National University Press, pp. 244–361. In Dixon, R. M. W., and Blake, B. (eds.), *Handbook of Australian Languages*, Vol. I. Canberra: Australian National University Press.
- Eckert, Penelope, and John Rickford (2001). *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press.
- Ellington, J. (1977). Aspects of the Tiene language. Ph.D. dissertation, University of Wisconsin–Madison.
- Fagan, Brian M. (1987). *The Great Journey: The Peopling of Ancient America*. London: Thames and Hudson.
- Farr, James, and Robert Larsen (1979). A selective word list in ten different Binandere languages. Mimeo.
- Fortson, Benjamin (2004). *Indo-European Language and Culture: An Introduction*. Oxford: Blackwells.
- Fox, Anthony (1995). *Linguistic Reconstruction*. Oxford: Oxford University Press.
- Garrett, Andrew (1990). The origin of NP split ergativity. *Language* 66(2): 261–96.
- Georg, Stefan, and Alexandr Vovin (2003). From mass comparison to mess comparison. *Diachronica* 20(2): 331–62.
- Gildea, Spike (ed.) (1999). *Reconstructing Grammar: Comparative Linguistics and Grammaticalization*. Amsterdam: John Benjamins.
- Givón, Talmy (1999). Internal reconstruction: as method, as theory. In Spike (ed.), *Reconstructing Grammar: Comparative Linguistics and Grammaticalization*, pp. 107–60. Amsterdam: John Benjamins.

- Goddard, Ives (1982). The historical phonology of Munsee. *International Journal of American Linguistics* 48(1): 16–48.
- Gray, R. D., S. J. Greenhill, and R. M. Ross (2007). The pleasures and perils of Darwinizing culture (with phylogenies). *Biological Theory* 2(4): 360–75.
- Greenberg, Joseph (1963). *Languages of Africa*. Bloomington: Indiana University Press.
- Greenberg, Joseph (1987). *Language in the Americas*. Stanford: Stanford University Press.
- Grimm, Jakob (1880). *Geschichte der deutschen Sprache*. Leipzig.
- Gudschinsky, Sarah (1964). The ABCs of Lexico statistics (Glottochronology). In Dell Hymes (ed.), *Language in Culture and Society*, pp. 612–23. London: Harper and Row.
- Haas, Mary R. (1969). *The Prehistory of Languages*. The Hague: Mouton.
- Hale, Mark (1998). Syntactic change. *Syntax* 1(1): 1–17.
- Hale, Mark (2007). *Historical Linguistics: Theory and Method*. Oxford: Blackwell.
- Hamel, Patricia (1994). *Grammar and Lexicon of Loniu*. Vol. C-103. Canberra: Pacific Linguistics.
- Harris, Alice C (2000). Where in the word is the Udi clitic? *Language* 76(3): 593–616.
- Harris, Alice (2002). *Endoclitics and the Origin of Udi Morphosyntax*. Oxford: Oxford University Press.
- Harris, Alice (2008). Reconstruction in syntax: reconstruction of patterns. In G. Ferraresi and M. Goldbach (eds.), *Principles of Syntactic Reconstruction*, pp. 73–95. Amsterdam: John Benjamins.
- Harris, Alice, and Lyle Campbell (1995). *Historical Syntax in Cross-linguistic Perspective*. Cambridge: Cambridge University Press.
- Harrison, K. David (2007). *When Languages Die: The Extinction of the World's Languages and the Erosion of Human Knowledge*. Oxford: Oxford University Press.
- Haspelmath, Martin (2002). *Understanding morphology*. London: Arnold.
- Heath, Jeffrey (1998). Hermit crabs: formal renewal of morphology by Phonologically mediated affix substitution. *Language* 74:728–59.
- Heggarty, P. (2008). Linguistics for archaeologists: a case-study in the Andes. *Cambridge Archaeological Journal* 18(01): 35–56.
- Hock, Hans Henrich (1991). *Principles of Historical Linguistics*. 2nd rev. ed. Berlin: Mouton de Gruyter.
- Hock, Hans Heinrich, and Brian D. Joseph (1996). *Language History, Language Change, and Language Relationship: An Introduction to Historical and Comparative Linguistics*. Berlin: Mouton de Gruyter.
- Holtzknicht, Susanne (1989). *The Markham Languages of Papua New Guinea*. Vol. C-115. Canberra: Pacific Linguistics.
- Hombert, Jean-Marie, John Ohala, and William Ewan (1979). Phonetic explanations for the development of tones. *Language* 55(1): 37–58.
- Hooley, Bruce A. (1971). Austronesian languages of the morobe district, Papua New Guinea. *Oceanic Linguistics*: 79–151.
- Jeffers, Robert J., and Ilse Lehiste (1980). *Principles and Methods for Historical Linguistics*. Cambridge: MIT Press.
- Jespersen, Otto. 2008/1922. *Language; its nature, development and origin*. 6th ed. N. P.: Ghose Press.

- Johanson, Lars, and Éva Ágnes Csató (eds.) (1998). *The Turkic Languages*. London: Routledge.
- Jones, Sir William. 1798. "Third Anniversary Discourse: on the Hindus" *Asiatick Researches* 1: 415–431 (Delivered 2 February 1786).
- Johnson, Keith (2008). *Quantitative Methods in Linguistics*. Oxford: Blackwell.
- Johnston, Ray (1978). Steps towards the phonology and grammar of Proto-Kimbe' Mimeo. Ukarumpa: Summer Institute of Linguistics.
- Joseph, Brian (1990). Review of *Introduction to Historical Linguistics*. *Language* 66(3): 633–34.
- Joseph, Brian D., and Richard D. Janda (2003). *The Handbook of historical linguistics*. London: Blackwell.
- Katamba, Francis (1993). *Morphology*. London: Palgrave Macmillan.
- Keesing, R. M. (1975). *Kwaio Dictionary*. Canberra: Australian National University, Department of Linguistics, Research School of Pacific Studies.
- King, Michael (1989). *Mori Mori: A People Rediscovered*. New York: Viking.
- King, Ruth (2000). *The Lexical Basis of Grammatical Borrowing*. Amsterdam: John Benjamins.
- Koch, Harold (1996). Reconstruction in morphology. In M. Durie and M. Ross (eds.), *The Comparative method Reviewed: Regularity and Irregularity in Language Change*, chap. 8, pp. 218–63. New York: Oxford University Press.
- Kroch, Anthony (2002). Syntactic change. In *Handbook of Contemporary Syntactic Theory*, pp. 699–729. London: Blackwell.
- Krueger, John (1962). *Yakut Manual*. Bloomington: Indiana University Press.
- Labov, William (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William (2001). *Principles of Linguistic Change: Social Factors*. 2 vols. Oxford: Blackwell.
- Labov, William (2007). Transmission and diffusion. *Language* 83(2): 344–387.
- Labov, William, S. Ash, and C. Boberg (2006). *The Atlas of North American English: Phonetics, Phonology, and Sound Change: A Multimedia Reference Tool*. Berlin: Walter de Gruyter.
- Lahiri, Aditi (2003). *Analogy, Levelling, Markedness: Principles of Change in Phonology and Morphology*. Manuscript.
- Langacker, Ronald W. (1968). *Language and Its Structure*. New York: Harcourt, Brace.
- Langlois, Annie (2004). *Alive and Kicking: Areyonga Teenage Pitjantjatjara*. Canberra: Pacific Linguistics.
- Laycock, D. C. (1973). Sissano, Warapu and Melanesian pidginisation. *Oceanic Linguistics* 12: 245–78.
- Laycock, D. C. (1976). *Austronesian Languages: Sepik Provinces*. Vol. C-39, pp. 399–418. Canberra: Pacific Linguistics.
- Lee, Jennifer (1987). *Tiwi Today: Study of Language Change in a Contact Situation*. Canberra: Australian National University, Research School of Pacific and Asian Studies.
- Lehmann, Winfred P. (1962). *Historical Linguistics: An Introduction*. London: Holt, Rinehart and Winston.
- Lewis, Geoffrey (2000). *Turkish Grammar*. 2nd ed. Oxford: Oxford University Press.
- Lichtenberk, Frantisek (1991). Semantic change and heterosemy in grammaticalisation. *Language* 67(3): 474–509.

- Lightfoot, David (2006). *How New Languages Emerge*. Cambridge: Cambridge University Press.
- Lincoln, Peter Craig (1976). Describing Banoni. Ph.D. dissertation, University of Hawaii, Honolulu.
- Lowe, Beulah (1960). Grammar lessons in Gupapuygu. Mimeo. Milingimbi.
- Lynch, John (1977a). *Lenakel Dictionary*. Canberra: Pacific Linguistics.
- Lynch, John (1977b). Notes on Maisin: an Austronesian language of the northern province of Papua New Guinea? Mimeo.
- Maiden, Martin (2004). Into the past: morphological change in the dying years of Dalmatian. *Diachronica* 21: 85–111.
- Mallory, J. P. (1989). *In Search of the Indo-Europeans: Language, Archaeology, and Myth*. London: Thames and Hudson.
- Marsack, C. C. (1973). *Teach Yourself Samoan*. London: English University Press.
- Matisoff, James A. (1990). On megalocomparison. *Language* 66(1): 106–20.
- McGregor, William (ed.) (2006). *Nekes and Worms' Australian Languages*. The Hague: Mouton.
- McMahon, April, and Robert McMahon (2003). Finding families: quantitative methods in language classification. *Transactions of the Philological Society* 101(1): 7–55.
- McMahon, April, and Robert McMahon (2006). *Language Classification by Numbers*. Oxford: Oxford University Press.
- Mufwene, Salikoko (2001). *The Ecology of Language Evolution*. Cambridge: Cambridge University Press.
- Muller, Max (1864). *Lectures on the Science of Language*. London: Longman.
- Mulvaney, John and John Kamminga (1999). *The Prehistory of Australia*. Sydney: Allen and Unwin.
- Nakhleh, Luay, D. Ringe, and T. Warnow (2005). Perfect phylogenetic networks: a new methodology for reconstructing the evolutionary history of natural languages. *Language* 81: 382–420.
- Nichols, Johanna (1992). *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.
- Nida, Eugene (1962). Analogical Change. In W. Anderson and N. C. Stageberg (eds.), *Introductory Readings in Language*, pp. 86–92. New York: Holt, Rinehart, and Winston.
- Ostler, Nicholas (2005). *Empires of the Word: A Language History of the World*. New York: HarperCollins.
- Öztopçu, Kurtuluş, Zhoumagaly Abuov, Nasir Kambarov, and Youssef Azemoun (1996). *Dictionary of the Turkic Languages*. London: Routledge.
- Parker, G. J. (1970). *Southeast Ambrym Dictionary*. Canberra: Pacific Linguistics.
- Phillips, Betty S. (2006). *Word Frequency and Lexical Diffusion*. London: Palgrave Macmillan.
- Rankin, R. L. (2003). The comparative method. In Brian D. Joseph and Richard D. Janda (eds.), *The Handbook of Historical Linguistics*, pp. 183–212. Oxford: Blackwell.
- Rask, R. (1818). *Undersøgelse om det gamle nordiske eller islandske sprogs oprindelse*. Copenhagen.

- Ringe, Don (2003). Internal reconstruction: In *The Handbook of Historical Linguistics*, Brian D. Joseph and Richard D. Janda, (eds.), 244–61. Maiden, Mass.: Blackwell.
- Romaine, Suzanne (1994). *Language in Society: An Introduction to Sociolinguistics*. New York: Oxford University Press.
- Romaine, Suzanne (1988). *Pidgin and Create Languages*. London: Longman.
- Ross, Malcolm D. (1988). *Proto-Oceanic and the Austronesian Languages of Western Melanesia*. Vol. C-98. Canberra: Pacific Linguistics.
- Ross, Malcolm (2007). Calquing and metatypy. *Journal of Language Contact*. Thema 1: 116–143.
- Schmidt, Annette (1985). *Young People's Dyirbal: An Example of Language Death from Australia*. New York: Cambridge University Press.
- De Saussure, Ferdinand (1986). *Course in General Linguistics*. Chicago: Open Court Press.
- Schulze, Wolfgang (forthcoming). *Person, Klasse, Kongruenz, 2*. Munich: Lincom Europa.
- Schütz, Albert J. (1985). *The Fijian Language*. Honolulu: University of Hawaii Press.
- Shapiro, Marianne (1990). *De Vulgari Eloquentia, Dante's Book of Exile*. Lincoln: University of Nebraska Press.
- Stokes, Bronwyn (1982). A description of the Nyigina language of the Kimberley region of Western Australia. Ph.D. dissertation, Australian National University, Canberra.
- Straughn, Christopher (2006). Sakha-English dictionary. Unpublished ms.
- Swadesh, Morris (1964). Linguistics as an instrument of prehistory. In Dell Hymes (ed.), *Language in Culture and Society: A Reader in Linguistics and Anthropology*, pp. 575–84. London: Harper and Row.
- Taubenschmidt, Gerhard (2005). Sinaugoro dictionary. Unpublished ms.
- Thomason, Sarah (1997). *Contact Languages: A Wider Perspective*. Amsterdam: John Benjamins.
- Thomason, Sarah (2001). *Language Contact*. Edinburgh: Edinburgh University Press.
- Thomason, Sarah, and Terrence Kaufman (1988). *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press.
- Thurgood, Graham (2002). Vietnamese and Tonogenesis. *Diachronica* 19(2): 333–63.
- Thurston, William R. (1987). *Processes of Change in the Languages of North-western New Britain*. Vol. B-99. Canberra: Pacific Linguistics.
- Trask, R. L. (1994). *Language change*. London: Routledge.
- Traugott, Elizabeth C., and Richard B. Dasher (2005). *Regularity in Semantic Change*. Cambridge: Cambridge University Press.
- Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society* 1(2): 179–95.
- van Kemenade, Ans, and Nigel Vincent (eds.) (1997). *Parameters of Morphosyntactic Change*. Cambridge: Cambridge University Press.
- Van der Tuuk, H. N. (1971). A grammar of Toba Batak [a translation by Jeune Scott-Kemball of Tabasche Spraakunst (1864/1867)]. Koninklijk Instituut voor Taal-, Land- en Volkenkunde, Translation Series 13. The Hague: Martinus Nijhoff.
- Verner, Karl (1875). Eine Ausnahme der ersten Lautverschiebung, *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen* 23(2): 97–130.

- Walker, Alan Trevor (1980). *Sawu: A language of Eastern Indonesia*. Ph.D. dissertation, Australian National University, Canberra.
- Walker, Douglas C. (1984). *The Pronunciation of Canadian French*. Ottawa: University of Ottawa Press.
- Wilkins, David P. (1996). Natural tendencies of semantic change and the search for cognates. In M. Durie and M. Ross (eds.), *The Comparative Method Reviewed*, pp. 264–304. New York: Oxford University Press.
- Williams, Herbert W. (1991). *A Dictionary of the Maori Language*. Wellington: GP Books.
- Yip, Moira (2002). *Tone*. Cambridge: Cambridge University Press.

Index

- Aboriginal Australia, 300
absolute, 229
accents, 13
accusative, 192–193, 227, 229, 292
acronyms, 211
Admiralty Islands, 237
adpositions, 219
affixes, 109, 140, 192, 219, 234–236, 286
Afrikaans, 44, 57, 285, 288
Age-Area Hypothesis, 313, 314
agglutinating languages, 220, 225, 292
Alamblak, 232, 233
Albanian, 269, 270
Algonquian, 47, 115, 286, 322, 347
allomorphs, 123, 124, 131, 186, 227
allophones, 66, 69, 73, 99, 171, 189, 190, 252, 266
alternation, 43, 46, 124–131, 167, 192, 259
alveolar, 18, 37, 38, 40, 42, 73, 74, 269, 347
Amara, 271
ambiguity, 94
Ambrym, 57
Amerind, 116
analogy, 189–193, 203, 239, 240, 350
Anatolian, 70, 127
Angkamuthi, 27, 68, 72
Angles, 29
antipassive, 229
Aotearoa. *See* New Zealand
apheresis, 27, 68
apocope, 27–28, 30, 36, 43, 74
Arabic, 6, 204, 207, 267, 337
Arapaho, 347
archaeology, 122, 299–300, 305–306, 316, 321
areal features, 270
Arifama-Miniafa, 316
Aroma, 340
aspiration, 44, 167
assimilation, 16, 37–44, 130
assimilation at a distance, 42
asterisks, 126
attitudes, 18
Australian English, 11, 26
Australian languages, 27, 130, 140, 227, 233, 289, 318, 319

- auxiliaries, 233, 234
 Aymara, 153
- Bahasa Indonesia, 5, 9, 11, 122–123,
 212, 267
 Bahasa Malaysia, 204–205
 Balkans, 269–271
 Bandjalang, 181, 206, 226–228
 Banoni, 57, 60, 61, 69, 347
 Bantu, 271, 336
 Bardi, 130, 193, 194, 339
 basic vocabulary, 138, 140, 148–150,
 250
 Bayesian phylogenetics, 159–161
 Belgium, 264
 Bering Strait, 319
 Bible, 199
 bifurcation, 199, 201
 bilabial, 18, 35, 39
 Binandere, 331
 binary, 155, 346
 Bislama, 35, 170, 203, 211–212, 274,
 304
 blends, 212
 Blevins, Juliette, 31
 Bloomfield, Leonard, 248
 Blust, Robert, 347
 borrowing, 5, 28, 45, 152, 158, 170, 187,
 205–208, 267, 272
 Bougainville, 60, 315
 bound morphemes, 274
 boundaries, 222
 boundary shift, 186
 branches, 110, 155, 158
 breaking, 36, 209, 232, 236
 broadening, 199, 200, 201, 202
 Bulgarian, 269
- calquing, 267
 Campbell, Lyle, 218
 Canada, 264
 Caucasus, 13, 240
 causes, 13–15, 36, 46, 214, 226, 289
 Celtic, 6, 14
 Central America, 229
 Central Australia, 13, 192
 Chaldaic, 163
 chance, 5, 108–109, 112, 113, 164, 283
 chance resemblances, 109, 116
 change, 7, 8, 11–19, 23
 arbitrariness, 178
 conditioning, 94
 directionality, 235
 frequency, 88
 gradualness, 273
 naturalness, 87
 observation of, 246
 regularity, 191
 value judgments, 246
 change in progress, 249
 character sets, 155–159
 Chinese, 14, 271, 312
 Chomsky, Noam, 253, 346
 Chukchi, 188
 civilization, 12
 clade, 153
 class, 204
 class markers, 225
 classification, 218
 clipping, 210
 clitics, 234, 235
 cluster reduction, 29
 clusters, 13, 28–31, 36, 39, 130, 251
 cognate sets, 81, 84, 93–94, 99, 116,
 142–152, 169, 193, 210–214, 319
 colonization, 321
 common development, 111
 comparative culture, 303, 304, 305,
 306, 321
 comparative method, 79, 97, 101–102,
 110–111, 116, 121, 162–164, 173,
 178–180, 193, 213, 272, 284, 304,
 316–322, 350, 352
 as algorithm, 162
 in syntax, 217
 compensatory lengthening, 126
 complementary distribution, 67, 73, 74,
 96, 98–101, 166, 167, 169, 189, 213,
 251, 267
 complete loss, 67
 compression, 210, 211
 computational methods, 151, 152
 conditioned sound change, 56, 67, 72,
 94–95, 98, 99, 128, 166, 171
 constituent order, 229, 230, 231, 232
 constructions, 241
 contour tone, 46
 convergence, 269, 270, 271, 272

- convergent development, 170
 copying. *See* borrowing
 core vocabulary, 138, 144, 148–150, 206
 replacement, 138
 correspondence sets, 83, 91, 94–96, 98,
 100–103, 152, 155, 159, 166, 217
 distribution of forms, 86
 Cree, 285, 286
 creole, 273, 274, 277–286, 352
 cultural reconstruction, 321
 culture contact, 305
 Cypriot Arabic, 31, 337
- Danish, 269
 Dante, 7
 degrammaticalization, 237
 derivation, 248
 devoicing, 42, 114, 124, 125, 179
 diachrony, 3, 247
 dialect chains, 180
 dialects, 11, 26, 32, 34, 36, 40–41, 66, 74,
 121, 173–180, 252, 273, 274, 346,
 348–351
 diffusion, 179, 239, 241, 269, 272, 320,
 351
 diminutives, 211
 dingo, 300
 diphthongs, 28, 36
 dissimilation, 43, 44, 167, 186
 distance matrix, 157
 distance-based methods, 137
 divergence, 5
 Diyari, 39, 158, 192
 Djambarrpuyŋu, 185
 DNA clock, 148
 doublet, 28, 208
 drift, 112, 113, 258
 Dutch, 110, 111, 137, 180, 269, 277,
 285, 345
 Dyrbal, 291–293
- education, 261
 Efate, 289
 empirical linguistics, 247
 Enga, 301
 Enggano, 43, 57, 113
 English, 4–6, 14, 26, 29, 31, 40, 66, 109,
 163, 188, 200, 246, 264, 290, 307,
 313, 345, 350, 353
- environments, 96
 epenthesis, 31, 346
 ergativity, 227, 228, 229, 239, 292, 320
 Eskimo, 319
 esoterogeny, 287, 288, 352
 etymology, 163, 168, 203–205
 euphemism, 202
 Eurasia, 320
 Eurocentrism, 181
 evolution, 317
 excrescence, 30, 31, 346
 exoterogeny, 287–288, 352
 extension, 240
- family of languages, 7, 11, 112
 family tree, 8, 9, 110, 111, 137, 144, 148,
 153, 285, 301, 307
 family tree model, 180
 feature, 13, 34–35, 37–46, 87, 112, 138,
 155, 171, 218, 269, 320
 Fijian, 5, 9, 11, 26, 40, 67, 113, 169, 277
 307, 309
 final devoicing, 42, 124
 Finno-Ugric, 163
 fission, 36
 flap, 68
 Flemish, 163, 264
 folk etymology, 203
 fortition, 25, 42
 Fortson, Benjamin, 70, 346
 French, 4, 6, 14, 27, 34, 68, 110, 128, 187,
 203, 212, 264, 272, 308, 335, 345
 frication, 44
 front vowels, 57, 66, 97
 fusion, 16, 33, 34, 35, 47, 222, 223, 225,
 346
 fuzziness, 248
- gaps, 17, 18, 84, 87–90, 210
 Garrett, Andrew, 31
 geminate, 38
 gender, 256
 genealogies, 301
 genetically related, 78
 Georgian, 241
 German, 12, 14, 25, 42, 44, 57, 110, 111,
 124, 163, 164, 176, 180, 192, 200,
 201, 210, 231, 258, 264, 269, 270,
 274, 275, 288, 310, 312, 345

- Gilliéron, Jules, 178
 glottal stop, 46
 glottochronology, 136, 148–149, 151,
 307, 350
 Gothic, 6, 12, 164–167
 grammar, 6, 15, 109, 184, 219, 238, 272,
 279–285, 292–293, 303, 304
 grammatical change, 219
 grammatical conditioning, 173, 350
 grammatical functions, 227
 grammaticality judgments, 249
 grammaticalization, 121, 218, 233–239
 Grassmann, Hermann, 166
 Greek, 7, 34, 35, 44, 70, 71, 125, 127,
 138, 163, 164, 165, 166, 167, 192,
 240, 267, 268, 269, 270, 350
 Greenberg, Joseph, 115
 Greenland, 163
 Grimm, Jakob, 12, 164
 Guichard, Etienne, 163
 Gurindji, 285
- haplogy, 29
 harmony, 42, 43
 Harris, Alice, 217
 Hawaiian, 56–60, 79–95, 99, 101,
 112–115, 168–169, 260, 275, 289,
 312
 Hawaiki, 303
 Hebrew, 7, 163, 164, 168, 268
 heuristics, 78, 162
 Hindi, 229
 Hiri Motu. *See* Motu
 Hittite, 70, 127, 228, 239
 homeland, 8, 312–314, 321, 322
 homonyms, 70, 203, 287
 homophones, 70
 Hula, 340
 Huli, 42
 hyperbole, 202, 203
 hypercorrection, 204–205, 256
- Icelandic, 39, 40, 164
 iconicity, 4
 Ilokano, 33
 indeterminacy, 248, 250
 Indo-European, 70, 115–116, 125, 127,
 150, 163, 269, 271, 281, 347
 inflecting languages, 221
 inflection, 220
 inflectional cycle, 221
 innovation, 112–114, 181, 208, 210
 innovation-based method, 136
 interference, 202, 203, 266, 267, 270, 283,
 290
 internal reconstruction, 121, 125–130,
 193, 218, 238
 Inuit, 13, 310, 319
 Irian Jaya, 314
 irregular change, 113–114, 125
 isoglosses, 175–180
 isolating, 223
 Italian, 6, 7, 9, 110, 111, 163, 224, 264,
 268, 269, 345
- Jespersen, Otto, 12
 Jones, William, 6
- Kabana, 209, 271
 Kairiru, 36, 333
 Kara, 25
 Karawari, 329
 Karnic, 39, 158, 192
 Kartvelian, 240
 kava, 172, 206, 313
 Kaytetye, 185–186
 Kiput, 28
 Kiribati, 208, 304
 Koch, Harold, 351
 Koiari, 140–144, 153–157
 Koita, 140–144, 153–157, 230
 Korafe, 316, 331
 Kove, 271
 Kuman, 16
 Kuwae, 289
- Labov, William, 254–255, 346, 351
 Ladin, 10
 language change, 6, 11–18, 173, 190, 194,
 204, 219, 248, 254, 258, 265, 273,
 284, 285, 292, 299, 351
 gradualness, 247
 speed, 173
 language contact, 208, 230, 264, 265,
 288, 289
 language death, 10, 291, 293
 language families, 108
 language relationship, 7, 136, 164, 352

- language split, 173
 laryngeals, 125, 127
 laterals, 24, 25
 Latin, 6–11, 14, 25, 31, 70, 71, 127, 130,
 138, 163–167, 188, 204, 220, 224,
 238, 260, 268, 275, 284, 291,
 345–346
 Laz, 240–241
 Lenakel, 257–259
 lengthening, 35
 lenition, 24–26, 29, 41, 47, 57, 87, 90, 185
 lexical change, 109, 137, 138, 151, 208,
 210, 211, 265
 lexical copying, 207
 lexical diffusion, 258–259
 lexical replacement, 150
 lexicon, 101, 138, 148, 164, 179, 184,
 205, 208, 259–261, 265, 272, 277,
 284, 287, 288, 351
 lexicostatistics, 136, 139–140, 148–151,
 155–156, 349
 lexifier languages, 280, 282
 light verbs, 239
 lingua franca, 10, 268, 288
 linguistic areas, 270
 liquids, 25, 32, 89, 99, 101
 loan translation, 267
 loans, 68
 loss, 25–29, 32, 35, 36, 47, 56, 67–69,
 73, 75, 89, 112–113, 124, 229,
 293, 352
 lower class, 256, 261
 Luwian, 70, 127

 Madang, 301
 Maisin, 214, 316
 majority rules, 86, 348
 Makassar, 352
 Manam, 236, 333
 Māori, 5, 9, 30, 171, 290, 302–303,
 312–313, 352
 manner of articulation, 38–41, 44
 Marquesas Islands, 312
 Mbabaram, 109
 mechanisms, 239, 258
 Media Lengua, 285
 Mekeo, 47
 Melanesia, 257, 307
 Melanesian Pidgin, 284, 308
 merger, 69–72, 348
 metaphor, 202
 metathesis, 32–33
 Michif, 285–286
 middle class, 254–257
 Middle East, 6, 163, 320
 Middle English, 186
 migration, 300
 Milne Bay Province, 241
 minimal pairs, 69–70, 75, 101, 257
 mixed languages, 272, 281, 285–286
 monolingualism, 264
 morpheme ordering, 188
 morphologization, 234, 235
 morphology, 109, 137, 171, 184, 188,
 189, 193, 194, 218, 225, 238, 239,
 240, 267, 269, 285, 286, 350, 352
 Motu, 4, 18, 32, 41, 55–58, 70, 75, 97–99,
 207, 214, 218–219, 230, 250–252,
 267, 277–279, 347, 352
 Mountain Koiari, 144, 157
 Mpakwithi, 68
 multilingualism, 265
 multistate, 155
 mutual intelligibility, 180

 narrowing, 199, 200
 nasal harmony, 57
 nasalization, 34, 129
 nasalized vowels, 286
 Navajo, 308
 Ndebele, 187, 207
 NeighborNet, 157–158
 neogrammarians, 162, 170
 networks, 152–153, 258, 300
 New High German, 192
 New York City, 254
 New Zealand, 302
 New Zealand English, 66
 Nichols, Johanna, 319, 320
 nonce borrowings, 260
 Northern Paamese, 175
 Notu, 316, 331
 Nyikina, 193, 194, 339

 occlusivization, 31
 Oceania, 181, 320, 347
 Oceanic languages, 113, 114, 233, 307,
 311, 314

- OCP, 346
 Old English, 186
 Old Irish, 35, 70
 order of changes, 58, 60, 188, 283
- Paamese, 4, 5, 9, 11, 171–175, 177–179, 210, 222, 232, 236, 266–269, 309–310
- Pacific Rim, 321
- Palauan, 23, 25
- paleolinguistics, 316, 353
- Pama-Nyungan, 185
- Papua New Guinea English, 202
- parallel development, 112, 114
- parsimony methods, 158
- partial assimilation, 38, 40
- partial loss, 67
- partial similarity, 190
- passives, 228, 253
- pejoration, 200
- Perak, 205
- perfect phylogenies, 159
- phoneme, 18, 25, 35, 66–74, 85, 88–92, 96, 98–99, 101–102, 128, 131, 166, 169, 189, 213, 250, 252, 266, 267
- phonemic loss, 67
- phonemic merger, 70–72
- phonemic shift, 69
- phonemic split, 72–73, 267
- phonetic change, 25, 65, 66, 67, 69, 73
- phonetics, 45
- phonological change, 113–114, 131, 171, 185, 199, 259, 309
- phonological reduction, 238
- phonotactics, 16, 250
- pidgin, 273–274, 277–284, 291–293, 352
 versus creoles, 279
- pitch, 46
- place names, 350
- plausibility, 87
- Police Motu. *See* Motu
- Polynesia, 79–81, 115, 130, 163, 209, 305, 306–312, 314
- polysynthetic, 225, 226, 286
- population shift, 313
- possession, 229
- pre-Greek, 34
- prelanguage, 121, 130, 131
- prepositions, 221, 237
- primitive languages, 12
- productivity, 211
- progressive assimilation, 38, 39, 226
- pronouns, 109
- prothesis, 32, 68, 346
- Proto-Austronesian, 113–115, 310–312, 322–323
- Proto-Human, 116, 317
- Proto-Karnic, 193
- protolanguage, 6–10, 61, 65, 67, 69, 70, 79, 85, 88, 91, 97, 101–103, 112, 121, 130, 136, 149, 153, 165, 194, 205, 213, 219, 272, 303, 306, 310, 312–313, 317, 322, 346
 reality of, 102
- Proto-Nyulnyulan, 193, 213
- Proto-Oceanic, 89, 112, 167, 169, 311, 312, 317, 318
- Proto-Polynesian, 79, 81, 85, 88, 90–94, 112, 115, 130, 213, 284, 307, 312, 322, 348
- proto-world. *See* Proto-Human
- Proto-Yolŋu, 185
- Provençal, 186
- Pueblo Indians, 308
- quantitative methods, 136
- Québec French, 272
- Quechua, 153
- race, 256
- Rarotongan, 79–95, 101, 112–115, 168–169
- Rask, Rasmus, 164
- reanalysis, 45, 66, 81, 187, 194, 228, 229, 239, 240, 258
- reconstruction, 79, 238
 procedures, 91
- reflexes, 79, 85, 89, 91, 93, 101, 169–170, 313, 322
- register, 253
- regressive assimilation, 38–39, 47
- regularity, 275
- reinforcement, 187
- rhotacism, 25, 26
- rhotics, 24, 25, 89
- Romanian, 9, 10, 11, 269, 270
- Romansh, 10, 11, 264
- rooted tree, 153

- Rotuman, 208, 213
 Ruhlen, Merrit, 115
 rule change, 61
 rules, 55–61, 94, 111, 124, 173, 184, 217,
 240–241, 249–252, 266, 312
 Russian, 110, 191

 Sahul, 317, 321
 Samoan, 14, 79–96, 101, 112–115,
 122–124, 130, 168, 275, 284, 303
 Sango, 285
 Sanskrit, 6, 7, 11, 44, 70, 71, 163–167,
 207, 210, 350
 Saussure, Ferdinand de, 3, 4, 19, 127,
 190, 247, 248, 250
 Scots English, 66
 semantic change, 113, 199, 201, 203, 210,
 212
 Semantic interference, 266
 semantics, 171, 184, 266, 303, 304
 Sepa, 333
 Sera, 333
 serial verbs, 232–233, 237
 Shakespeare, 11, 199
 shared features, 219
 shared innovation, 111, 112, 114, 115,
 322
 shared retention, 111
 shift, 12, 17, 25, 59, 60, 67, 69, 109, 114,
 199, 201–203, 212, 256, 290, 292,
 338
 similarity
 systematicity, 168
 simplification, 16, 17
 Sinaugoro, 97–99, 214, 340, 349
 Slavic, 31
 social class, 254, 257
 social systems, 181
 sociolinguistics, 254
 Solomon Islands, 272, 314
 sonority, 24, 25, 346
 sound change, 12, 16, 17, 24–26, 29, 33,
 35–37, 44–47, 55, 59–61, 65–67,
 72–73, 79, 85, 87, 95, 97, 102,
 109–113, 122–125, 128–130, 152,
 170–171, 178–179, 184–186, 190,
 213, 238, 258, 347, 351
 conditioned by word class, 173
 environments, 217
 naturalness, 86
 regularity, 168
 sound correspondences, 82–85, 92–94,
 97–100, 103, 128, 149, 164–170, 213,
 214, 247, 284, 317, 319
 South Africa, 44, 288
 South America, 285, 309, 319
 Southeast Ambrym, 27, 28
 Southern Paamese, 175
 Spanish, 7, 9, 31, 163, 200, 277, 279,
 289
 speech communities, 109
 Sprachbund, 270
 state, 12, 18, 122, 124, 155, 159, 171,
 250
 subgroup, 28, 108, 110–115, 127, 139,
 145, 152–153, 158, 306, 314, 348
 substratum, 13–14
 Suena, 330
 suppletion, 188
 SVO word order, 229
 Swadesh, Morris, 140–141
 Swahili, 289
 sweet potato, 309
 Sye, 15, 220, 234
 synchrony, 65, 130, 247
 syncope, 16, 28, 29
 syntactic change, 217, 238
 systematic similarities, 5

 taboo, 150, 202, 203, 209, 351
 Tagalog, 33
 Tai-Kadai, 46
 Tasmania, 10, 293, 300, 318
 taxon, 153
 Tiene, 336
 tilde, 34
 time depth, 116, 148, 150–151, 350
 Time of Darkness, 301
 Tocharian, 70
 Tok Pisin, 4, 10, 42, 246, 259, 260, 261,
 271, 274, 275, 276, 277, 280, 281,
 282, 283, 284, 289, 352
 Tolai, 5, 9, 11, 218–219, 230, 280–281
 tone, 45–46, 271, 336
 Tongan, 79–101, 112, 114–115, 168–170,
 206, 257
 transitivity, 227
 Trans-New Guinea, 316

- Trukese, 47
 Turkish, 6, 42
 typology, 131, 218, 231, 232
- Ubir, 316
 Ukrainian, 31, 32
 umlaut, 43, 74
 unconditioned loss, 55
 univerbation, 226
 universals, 109, 318
 unpacking, 35, 47
 unrooted tree, 153
 unusual changes, 113
 upper class, 255, 256
 Uradhi, 23, 25, 57, 71, 72
 uvular, 13, 269, 270, 346
- Vanuatu, 4, 15, 27, 35, 113, 163, 174,
 202, 232, 274, 304, 314
 variability, 204, 248–249, 252–257, 293
 variation, 174
 velar, 16, 18, 34, 40, 67, 70, 88, 214,
 267
 Verner, Karl, 167
- voiceless, 18, 24, 35, 39–40, 46, 57, 69,
 124, 165, 350
 Voltaire, 163
 vowel breaking, 36
 vowel harmony, 42, 43, 347
 vowels, 16, 17, 24, 25, 27–30, 34–36, 46
 reconstruction, 83
- Wampar, 209
 Wangkumara, 158, 192
 Warluwarra, 41
 wave model, 173, 179
 word boundary, 171
 word classes, 171
 word order, 220
 Wörter und Sachen, 310, 312, 322–323
- Yandruwandha, 39, 158, 192, 240
 Yanyuwa, 41
 Yawarrawarrka, 39, 158
 Yawuru, 339
 Yimas, 225, 232, 329
 Yolŋu, 207, 352
 Young People's Dyrbal, 291

"I wish this book had been available ten years ago, when I was a student of historical linguistics. Its clear style and lack of Indo-European language-based data would make it an ideal textbook for students of historical linguistics in the UK."

—Scott Allan, *Te Reo*

"Crowley has written out the mainlines of historical linguistics in an exceptionally lucid and engaging style."

—Robert Blust, *Linguistics*

"A most refreshing treatment of historical linguistics at a beginning level . . . This book is an important one for those of us steeped in Western tradition . . . Western linguists will welcome the wealth of novel and highly usable examples."

—Brian D. Joseph, *Language*

All languages change, just as other aspects of human society are constantly changing. This book is an introduction to the concepts and terminology of historical linguistics, the study of language change over time. It covers the major areas of historical linguistics, presenting concepts in a clear and accessible way. Examples of language change are given throughout, with special emphasis on the English language. While the needs of undergraduate students of linguistics have been kept firmly in mind, the book will also be of interest to the general reader seeking to understand language change.

For this fourth edition, a number of new sections have been added, covering new problems and several new datasets. Existing materials have been updated with new sections on grammaticalization, tonogenesis, morphological change, and new methods in language classification.

— was Senior Lecturer in Linguistics at the University of Waikato in New Zealand until 2005.

— is Assistant Professor of Linguistics at Yale University.

Cover design by Black Hat Design

