The Foundations of Latin

The Foundations of Latin

by

Philip Baldi

Mouton de Gruyter
Berlin · New York 2002

Mouton de Gruyter (formerly Mouton, The Hague)
is a Division of Walter de Gruyter GmbH & Co. KG, Berlin.

An earlier version of this work was published in 1999 as volume 117 of the series
Trends in Linguistics – Studies and Monographs.

∞ Printed on acid-free paper which falls within the guidelines
of the ANSI to ensure permanence and durability.

Library of Congress Cataloging-in-Publication-Data

Baldi, Philip.
The foundations of Latin / by Philip Baldi.
p. cm. –
Includes bibliographical references and indexes.
ISBN 3-11-016294-6 (cloth : alk. paper)
ISBN 3-11-017208-9 (pbk. : alk. paper)
1. Latin language – Grammar, Historical. 2. Indo-European languages – Influence on Latin. 3. Inscriptions, Latin – Italy. 4. Italy – Languages. I. Title. II. Series.
PA2071.B35 1998
475–dc21 98-51324
CIP

Die Deutsche Bibliothek – Cataloging-in-Publication-Data

Baldi, Philip:
The foundations of Latin / by Philip Baldi. – Berlin ; New York : Mouton de Gruyter, 2002
ISBN 3-11-017208-9 br.
ISBN 3-11-016294-6 Geb.

Cover design: Sigurd Wendland, Berlin.
Printed in Germany.

Preface

Generations of students of Indo-European linguistics and of the Classics have been initiated to the history of the Latin language according to a familiar model. Most start with Palmer's *The Latin language*. The curious journey through Buck's *Comparative grammar of Greek and Latin* and Meillet's *Esquisse d'une histoire de la langue latine*. The committed know that the real answers lie somewhere in Sommer—Pfister's *Handbuch der lateinischen Laut- und Formenlehre*, and especially Leumann's *Lateinische Laut- und Formenlehre*, to which we might now add Sihler's massive revision and rewrite of Buck, the *New comparative grammar of Greek and Latin*. Specific interests and particular issues may take the serious to the works of Meillet—Vendryes, Devoto, Pisani, Niedermann, Prosdocimi and many others. But the road becomes rocky beyond Palmer, and without some specialized guidance or training, even the most ardent enthusiast of the history of Latin will find some of the issues impenetrable. Many give up on the topic, convinced that the arcane methodology and specialized techniques of Indo-European linguistics place the details beyond the grasp of the everyday scholar and ambitious student of Indo-European or classical studies.

In the United States, students of the Classics and the basic Indo-European foundations of Latin have relied primarily on Palmer and Buck. Each of these fine volumes has undeniable merits and strengths, and an equal number of evident flaws and weaknesses. The same is true of other synthetic accounts of the history of Latin. In some, there is no information on methodology and basic assumptions of historical-comparative research; in others there is a lack of discussion of the sources of information on which the historical linguistic analysis of Latin and the other Indo-European languages is based; in still others there is a lack of cultural and historical perspective, with the history of Latin presented as if the language had no speakers; most lack a coherent treatment of the inscriptional tradition, and of the non-Latin Italic languages; and in none of them is there a serious treatment of historical syntax, i.e. the evolution of Latin sentence structure from its Proto-Indo-European beginnings. In fact, not one of the available resources addresses the total picture of the history of Latin, methodologically, linguistically, materially and culturally, from its Indo-European origins up to the time when it starts to move in the direction of the Romance languages.

When I began this project over a decade ago, I imagined that *The foundations of Latin* would fill all the gaps in the previous literature, and add a few original twists of its own. It is my hope now that it will fill some of the gaps, and still add some original twists. This volume includes an account of historical linguistic methodology and the Indo-European family of languages which surpasses that found in comparable works. I have tried to justify linguistic details essential to the precise understanding of the history of Latin and the other Indo-European languages, and to

be explicit and current on matters relating to the inscriptional and philological record, and on the connections of Latin with other Indo-European languages. And I have attempted to provide up-to-date analysis and commentary which I hope makes the material contained in this volume considerably more than a reshaping of the known facts.

But gaps remain, the major one of which is a coherent treatment of historical Latin syntax. After writing a lengthy outline, I realized that the subject was so vast and fraught with theoretical and methodological peril that it would be pointless to undertake a synthesis in a volume of this sort. Phonology is manageable; once the theoretical decisions are made (such as whether to include laryngeals, glottalized consonants, voiceless aspirates, etc.), the details follow in a more or less straightforward manner. Morphology is more challenging, owing to the large number of morphological elements in a language of the Indo-European and Latin type, but it is nonetheless finite. Syntax, however, is vast and laden with conflicting data, and its analysis depends on crucial theoretical assumptions and subtle grammatical judgments. Historical syntax is particularly treacherous, for not only must a theoretical model be selected and justified, it must be appropriate for both the protolanguage and the target language, and the processes connecting the two must be stated with the same explicitness as those detailing the historical phonology and morphology. In effect, the syntax of Proto-Indo-European would have to be reconstructed before the historical syntax of Latin could be written. I realized that this cannot be accomplished in a synthetic account of the history of the language. It must be treated in a volume of its own.

This book is not intended to be a replacement for other volumes on the history of the Latin language. Leumann and now Sihler, for example, will continue to be primary sources on highly specific matters of phonology and morphology; Pisani will continue to provide the most comprehensive (though dated) account of the inscriptional record; and Niedermann will remain a capital resource of phonetic detail. At the same time, it is my expectation that specialists and non-specialists alike will find information and detail in these pages which they won't find outside the technical literature, and a synthesis of method, theory, data and analysis which is unavailable elsewhere. It is my hope that future generations of scholars will add *The foundations of Latin* to the list of standard resource works on the history of this fascinating language.

Contents

List of figures

List of maps

List of plates

List of tables

Acknowledgments

A project that takes as long as this one tends to involve a significant number of people. Since beginning *The foundations of Latin* in 1987, I have solicited the advice, opinions, and criticisms of many individuals. Because of the complicated nature of this book, and the different viewpoints which each colleague has brought to the task of commenting on its content, it has not been possible to accept every one of their suggestions. I hope each will understand why I have on occasion chosen to ignore, or to circumvent, their good counsel. It goes without saying that none of them is responsible for the errors or misjudgments which remain, especially because they all saw the work at different stages in its development.

Five colleagues stand out for special mention. Harm Pinkster (Amsterdam), Theo Vennemann (Munich), Frits Waanders (Amsterdam), Gary Miller (Florida), and William R. Schmalstieg (Penn State) read all or most of the manuscript in one of its early stages. Their advice has been invaluable, and their influence is everywhere. Archie Allen, Joseph Cotter, David Engel, Garrett Fagan, Aristoula Georgiadou, Paul Harvey, and Stephen Wheeler, my Classics colleagues at Penn State, have given me advice on technical matters in Greek and Latin. Alfred Bammesberger (Eichstätt) commented on an early draft of the chapters on phonology and morphology. James Mallory (Belfast) provided a critical reading of material on archaeology and the homeland question, and generously allowed me to make use of his maps of IE languages and proposals for the homeland. Bruce Campbell (Drake) commented on an early draft of the first five chapters. Christian Lehmann (Bielefeld) advised me on the letter of Claudius Terentianus. Snezana Milovanovic (Newfoundland) critiqued an early draft of the section on the PIE verb. I have profited from exchanges with Brent Vine (UCLA) and Rex Wallace (Massachusetts) on certain epigraphical matters, and Rama N. Sharma (Hawaii) on some Sanskrit details. Pierluigi Cuzzolin (Pavia) made me aware of the discovery of the Garigliano Bowl. Joe Eska (Virginia Polytechnic) advised me on some recent Celtic finds, and Konrad Koerner (Ottawa) provided information on aspects of the history of Indo-European scholarship. Werner Winter (Kiel), the editor of the series in which this book is appearing, gave a critical reading of the entire manuscript, and shared his vast knowledge of the field of Indo-European linguistics in his review. To him I am particularly grateful.

David Lembeck (State College) executed the graphics work, and Jeaninne Schonta of Penn State's Deasy Cartography Lab prepared the maps; both endured my nagging corrections and revisions. Melinda Blough (Penn State) helped with literary citations and philological corrections. The entire manuscript was formatted and set by Peter T. Daniels, whose expertise extends beyond typesetting and into technical matters of linguistics.

Thanks are also due to several organizations and institutions. The Dutch National Science Foundation (NWO) granted me a Research Fellowship for a semester in

Amsterdam in the Winter–Spring of 1987, where I began this undertaking. The Institute for the Arts and Humanistic Studies of Penn State University (Stanley Weintraub and George Mauner, Directors) awarded me a travel grant in the Summer of 1991 to begin inscriptional research, and to spend the summer at the University of Amsterdam. The same organization contributed funds toward the expense of preparing the maps for this volume. The Department of Classics at the University of Amsterdam provided me an occasional home away from home, in the most congenial and supportive surroundings imaginable for a visiting colleague. The National Endowment for the Humanities awarded me a Faculty Fellowship for 1994–1995, which allowed me to take off the entire academic year to work on this book. The Office of Research and Graduate Studies of Penn State's College of the Liberal Arts (Raymond Lombra, Dean) awarded me a research and travel grant which allowed me to spend three weeks in Italy in the Summer of 1994 to secure photographs of inscriptions, and to have maps and other graphics work prepared. During the Spring of 1996 I held the Fulbright Chair in Historical Linguistics at the University of Naples, where I completed a good deal of the rewrite of the manuscript.

Thanks also to my parents, John and Fran Baldi, who accompanied me to Italy in the Summer of 1994 and not only shared their passion for the country, but also helped to navigate the occasionally treacherous Italian museum bureaucracy. Final thanks go to my wife, Teri Lindner. She has been with this project from the beginning, through good times and bad. Fortunately, the good times have far outnumbered the bad. The good times include summers in Amsterdam and inscription hunts in Italy, where she took photographs and shared my joy at seeing and touching some ancient monument. On a daily basis, however, they include the infectious enthusiasm and inspiration which she imparts to those around her.

Special thanks, and a blanket exculpation, to Brent Vine (UCLA), who drew my attention to a number of errors of detail in the hardcover edition, most of which I have been able to correct in the paperback.

State College
September, 1998

Credits

Photographs and reproductions

Item	*Language/title*	*Source*
Plate 1	Faliscan	Soprintendenza Archeologica per l'Etruria Meridionale, Rome
Plate 2	Marrucinian	After Mommsen, *Die unteritalischen Dialekte*, Plate XIV
Plate 3	Marsian	Agence Photographique de la Réunion des Musées Nationaux, Paris. Musée du Louvre
Plate 4	Oscan	Soprintendenza Archeologica di Napoli e Caserta, Naples
Plate 5	Paelignian	Soprintendenza Archeologica di Napoli e Caserta, Naples
Plate 6	South Picene	Museo Archeologico, Ascoli Piceno
Plate 7	Umbrian	Foto Gavirati, Gubbio
Plate 8	Vestinian	Soprintendenza Archeologica di Napoli e Caserta, Naples
Plate 9	Volscian	Soprintendenza Archeologica di Napoli e Caserta, Naples
Plate 10	Gaulish/Latin	Musei Vaticani, Rome
Plate 11	Lepontic	Civiche Raccolte Archeologiche di Milano
Plate 12	Messapic	Biblioteca Comunale "Marco Gatti", Manduria
Plate 13	North Picene/Latin	Museo Archeologico Oliveriano, Pesaro
Plate 14	Raetic	Museo Nazionale, Trento
Plate 15	Raetic	Museo Civico, Padua
Plate 16	Sicel	After Conway—Whatmough—Johnson, *The Prae-Italic dialects of Italy*, vol. 2, p. 446
Plate 17	Venetic	Museo Civico, Padua
Plate 18	Latin	From Gordon, *Illustrated introduction to Latin epigraphy*, plate 1, with permission of the author and the University of California Press
Plate 19	Latin	From Gordon, *Illustrated introduction to Latin epigraphy*, plate 2, with permission of the author and the University of California Press
Plate 20	Latin	From Gordon, *Illustrated introduction to Latin epigraphy*, plate 3, with permission of the author and the University of California Press
Plate 21	Latin	Archivio Fotografico Vasari, Rome
Plate 22	Latin	From Gordon, *Illustrated introduction to Latin epigraphy*, plate 4, with permission of the author and the University of California Press
Plate 23	Latin	From Gordon, *Illustrated introduction to Latin epigraphy*, plate 5, with permission of the author and the University of California Press
Plate 24	Latin	From Gordon, *Illustrated introduction to Latin epigraphy*, plate 6, with permission of the author and the University of California Press
Plate 25	Latin	From Gordon, *Illustrated introduction to Latin epigraphy*, plate 49, with permission of the author and the University of California Press
Plate 26	Latin	École Française d'Archéologie, Athens

Alphabetic chart

Table 4.1 The development of the alphabet — After Almar, *Inscriptiones Latinae*, p. 10

Abbreviations and symbols

Languages

Alb.	Albanian
Arm.	Armenian
AS	Anglo-Saxon
Att.	Attic
Av.	Avestan
Bret.	Breton
CLat.	Classical Latin
Cret.	Cretan
Dor.	Doric
Eng.	English
Etr.	Etruscan
Falisc.	Faliscan
Fr.	French
Gaul.	Gaulish
Germ.	German
Gk.	Greek
Gmc.	Germanic
Goth.	Gothic
Heb.	Hebrew
Hier.	Hieroglyphic
Hitt.	Hittite
Hom.	Homeric
Icel.	Icelandic
IE	Indo-European
Ion.	Ionic
Ir.	Irish
It.	Italian
Lat.	Latin
Latv.	Latvian
Lemn.	Lemnian
Lesb.	Lesbian
Lith.	Lithuanian
Luw.	Luwian
MIr.	Middle Irish
Myc.	Mycenaean
Norw.	Norwegian
O–U	Oscan–Umbrian
OCS	Old Church Slavic
OE	Old English
OHG	Old High German
OIc.	Old Icelandic
OIr.	Old Irish
OLat.	Old Latin
OLatv.	Old Latvian
OLith.	Old Lithuanian
OPers.	Old Persian
OPruss.	Old Prussian
OSax.	Old Saxon
Osc.	Oscan
OWel.	Old Welsh
Paelign.	Paelignian
Phry.	Phrygian
PIE	Proto-Indo-European
Praen.	Praenestine
Rom.	Roman
Russ.	Russian
Sab.	Sabine
Skt.	Sanskrit
Slov.	Slovene
Span.	Spanish
Tibur.	Tiburian
Toch.	Tocharian
Umbr.	Umbrian
Ved.	Vedic
Ven.	Venetic
Volsc.	Volscian
Wel.	Welsh

Symbols

†	indicates an ungrammatical string
*	indicates a reconstructed form
**	indicates an impossible form
→	is replaced by
<	develops from
>	develops into
[ ]	indicates phonetic representation
/ /	indicates phonemic representation
⟨ ⟩	indicates orthographic representation

Other

1st	1st person
2nd	2nd person
3rd	3rd person
abl.	ablative
acc.	accusative
AcI	Accusativus cum infinitivo (Accusative with infinitive)
act.	active
adj.	adjective
adv.	adverb
aor.	aorist
athem.	athematic
Aux	Auxiliary
B.C.E.	Before the Common Era
C	consonant
C.E.	In the Common Era
CLE	Carmina Latina Epigraphica
cf.	confer (compare)
CIE	Corpus Inscriptionum Etruscarum
CIL	Corpus Inscriptionum Latinarum
con.	conjugation
conj.	conjunction
contemp.	contemporaneous
dat.	dative
dec.	declension
dep.	deponent
dial.	dialect
du.	dual
emph.	emphatic
fem.	feminine
fut.	future
gen.	genitive
imper.	imperative
imperf.	imperfect
impft.	imperfective
indic.	indicative
inf.	infinitive
INFL	inflection (as syntactic constituent)
instr.	instrumental
interj.	interjection
interr.	interrogative
lit.	literally
LL	Lachmann's Law
loc.	locative
masc.	masculine
med.	mediopassive
mid.	middle
n.	noun
neut.	neuter
nom.	nominative
NP	Noun Phrase
O	Object
opt.	optative
part.	participle
pass.	passive
perf.	perfect
pers.	person
pl.	plural
poster.	posterior
prep.	preposition
pres.	present
pret.	preterite
prim.	primary
pron.	pronoun
prt.	particle
rel.	relative
RV	Rig Veda
S	Subject
s.v.	sub voce (under the word)
SC de Bacch.	Senatus Consultum de Bacchanalibus
sc.	scilicet (that is to say)
sec.	secondary
sg.	singular
subj.	subjunctive
them.	thematic
V	vowel (as phonological segment)
V	verb (as syntactic constituent)
v.	verb (as part of speech)
vcd.	voiced
vclss.	voiceless
voc.	vocative
VP	Verb Phrase
X	3rd sg. masc./fem. subject

Transcription conventions

0.1. General

In general, phonetic and orthographic symbols used in this book are those found in basic treatments of phonetics (e.g. Ladefoged 1982), and in the standard handbooks for a particular language or field.

0.1.1. Consonant symbols

Stop consonants have their normal representation as ⟨p, t, k, b, d, g⟩. The voiceless labiodental, dental, and palatal fricatives are represented as ⟨f⟩, ⟨s⟩, and ⟨š⟩ respectively, while their voiced counterparts are represented as ⟨v⟩, ⟨z⟩, and ⟨ž⟩. Voiceless interdental fricatives are represented as either ⟨þ⟩ or ⟨θ⟩; their voiced counterparts are represented as ⟨ð⟩ or ⟨δ⟩. For voiceless velar fricatives the symbol ⟨χ⟩ is used; for the voiced velar fricative ⟨ɣ⟩ is used. The symbol ⟨h⟩ is used for the aspirate. The symbols ⟨l, r, y, w⟩ have their normal value; the symbol ⟨ɬ⟩ indicates a back (velar) *l*. The nasals ⟨m⟩ and ⟨n⟩ have their normal values, while ⟨ŋ⟩ indicates a velar articulation.

0.1.2. Vowel symbols

The vowels ⟨a, e, i, o, u⟩ have their "continental" values. A macron over a vowel (e.g. ⟨ā⟩) indicates that the vowel is long, while a breve (e.g. ⟨ă⟩) indicates that the vowel is short. A vowel with both a macron and a breve (e.g. ⟨ĕ⟩) is one which can be either long or short. The symbol ⟨ə⟩ marks a mid central vowel. A tilde over a vowel (e.g. ⟨ẽ⟩) usually indicates nasalization, except where noted below.

0.1.3. Accent symbols

The acute accent (e.g. ⟨é⟩) typically indicates the main accent, the circumflex (e.g. (⟨ê⟩ or ⟨ẽ⟩) the next most prominent accent, and the grave (e.g ⟨è⟩) the least prominent accent.

0.2. Symbols by language

0.2.1. Albanian

The symbol ⟨ë⟩ represents [ə]; ⟨y⟩ (also written ⟨ü⟩) represents a close high front rounded vowel.

0.2.2. Anatolian

The symbol ⟨ḫ⟩ in Anatolian most likely represents a fricative, possibly with palatovelar articulation. The symbol ⟨š⟩ represents [s], and ⟨z⟩ represents [ts].

0.2.3. Armenian

Consonants followed by a ' are aspirated, e.g. ⟨t'⟩. The symbol ⟨c⟩ represents [ts]. The symbol ⟨č⟩ represents [č], and its voiced counterpart [j] represents [ǰ].

0.2.4. Baltic

The letter ⟨ė⟩ in Lithuanian stands for a close *e*, i.e. [ę]. The circumflex written with a tilde (e.g. ⟨ã⟩) indicates smooth rising intonation. A hook under a vowel (e.g. ⟨ę⟩) represents prior nasalization. The letter ⟨c⟩ stands for [ts], while ⟨č⟩ represents [č].

0.2.5. Celtic

The letter ⟨c⟩ in Old Irish represents generally [k]. A vowel marked by an acute accent (e.g. ⟨á⟩) is long.

0.2.6. Etruscan

The letter ⟨s⟩ represents [s], and the letter ⟨ś⟩ represents [š].

0.2.7. Germanic

For the value of the Gothic diphthongs see chap. 2, table 2.9. The symbol ⟨hu⟩ in Gothic represents a labialized voiceless velar fricative [χ^{w}]. A ⟨q⟩ in Gothic repre-

sents a labialized velar [$k^{\underline{u}}$]. The letter [æ] in Old English is phonetically [æ], i.e. a low front vowel. The symbol ⟨ø⟩ in Old Icelandic represents a mid front rounded vowel.

0.2.8. Greek

Greek letters have their standard values. The archaic digamma (Ϝ) is pronounced as the labial glide [w]. The acute accent indicates high pitch, the grave indicates low pitch, and the circumflex combines acute and grave. In Mycenaean transcriptions, a ⟨q⟩ probably indicates a voiceless labiovelar stop.

0.2.9. Indo-Iranian

0.2.9.1. Sanskrit and Vedic

The acute accent (e.g. ⟨á⟩) indicates high pitch. Aspirated stops are indicated by a following ⟨h⟩ (e.g. ⟨ph⟩). A dot under a letter (e.g. ⟨ṭ⟩) indicates retroflex articulation. The letters ⟨c⟩ and ⟨j⟩ are phonetically [č] and [ǰ] respectively. The letter ⟨ś⟩ (formerly written ⟨ç⟩) represents [š]. Final ⟨ḥ⟩ is a voiceless velar aspirate. The letters ⟨ṛ⟩ and ⟨ḷ⟩ represent vocalic *r* and *l*. A dot over an m (⟨ṁ⟩) marks nasalization of the preceding vowel. A tilde over an n (⟨ñ⟩) indicates palatal articulation.

Unless specific material is being highlighted, the term "Sanskrit" (Skt.) is used throughout this volume to designate language forms from Old Indic, both Vedic and (classical) Sanskrit. In general, the forms cited are Vedic. Sanskrit examples are cited either as roots or as inflected forms. Roots are indicated by a hyphen at the end, as in the substantive *vṛ́ka-* 'wolf' or the verb *yuj-* 'join'. Inflected forms for substantives are typically the nominative singular, as in *vṛ́kaḥ* 'wolf', while for verbs they are either in the third person singular indicative, as in *bhárati* 'X carries', or the first person singular, as in *bhárāmi* 'I carry'. When Sanskrit forms are cited in a form ending in a final consonant, the consonant is changed according to the phonological rules of Sanskrit for finals. For example, the nominative singular of *ēkas* 'one' and the genitive singular *datás* 'of the tooth' are cited as *ēkaḥ* and *datáḥ* in accordance with the rules which convert final *s* to *ḥ*. In general, Sanskrit words are cited in their inflected form in this volume.

0.2.9.2. Iranian

In Avestan, the symbol ⟨č⟩ (also written ⟨c⟩) is phonetically [č]. The symbol ⟨å⟩ is roughly [ɔ]; ⟨ś⟩ represents [š] as in Sanskrit, and ⟨t̰⟩ is an unreleased *t*.

0.2.10. Italic

In Oscan transcriptions, the letter ⟨í⟩ represents a long open high front vowel [i:] and ⟨ú⟩ represents the open back vowel [o].

In Umbrian transcriptions, the letter ⟨ř⟩ has the phonetic value of [rs]; ⟨ç⟩ was probably pronounced [š].

In Faliscan and non-Roman Latin, the letter ⟨q⟩ is occasionally used to indicate a voiceless velar stop.

In Sicel transcriptions the letter ⟨í⟩ represents [i:].

The letter ⟨í⟩ in the transcription of the S. Picene inscriptions represents a long, close mid-palatal vowel, viz. [ę:]. The letter ⟨ú⟩ represents the equivalent in the velar region, i.e. a long, close mid-velar vowel [ǫ:], though there is some evidence that ⟨ú⟩ stands for a short [ǫ] before final *-m*. Also, ⟨í⟩ represents the final sound in *i*-diphthongs in most S. Picene areas.

In transcriptions of N. Picene the symbol ⟨ś⟩ represents a dental sibilant [s], while the symbol ⟨š⟩ most likely represents a palatal sibilant like [š], though the possibility exists that it is a variant of ⟨ś⟩. The letter ⟨ú⟩ possibly represents the lower high rounded back vowel [ʊ].

0.2.11. Slavic

The symbols ⟨ĭ⟩ and ⟨ŭ⟩ represent the vowels ь short *i* (as in *pit*) and ъ short *u* (as in *put*) respectively. A hook under a vowel (e.g. ⟨ę⟩) represents nasalization. The symbol ⟨ě⟩ stands for the lower mid front vowel [ɛ]. The symbol ⟨c⟩ represents [ts], while the symbol ⟨č⟩ represents [č].

0.2.12. Tocharian

The symbol ⟨ṣ⟩ represents palatalized *k*; the symbol ⟨c⟩ represents palatalized *t*; and the symbol ⟨ś⟩ represents palatalized *s*. A tilde over an *n* (⟨ñ⟩) indicates palatalized *n*. A dot under an m (⟨ṃ⟩) shows that the preceding vowel is nasalized. The symbol ⟨ä⟩ represents a vowel of uncertain quality, possibly something like the high central [ɨ].

0.2.13. Proto-Indo-European

Raised ⟨h⟩ (e.g. ⟨b^h⟩) marks aspiration on the preceding stop. Voiceless palatals are indicated by ⟨k̂⟩; voiced palatals are marked by ⟨ĝ⟩ and ⟨$\hat{g}^h$⟩. Voiceless labiovelars are indicated by ⟨k^{μ}⟩; voiced labiovelars are indicated by ⟨g^{μ}⟩ and ⟨$g^{\mu h}$⟩. Glottalized

stops (ejectives) are indicated by a ' after a consonant, e.g. ⟨t'⟩. When marked, semivowels are indicated by ⟨u̯⟩ and ⟨i̯⟩. When marked, syllabic resonants are indicated by an open circle under a letter, viz. ⟨m̥, n̥, l̥, r̥⟩. When discussed, long resonants are indicated by a macron (e.g.⟨m̥̄⟩). The PIE laryngeals are assigned the symbols ⟨h_1, h_2, h_3⟩.

Chapter 1
Indo-European and the Indo-European languages

1.1. Introduction

By their very nature, languages have histories which reach far back in time. It is the task of historical-comparative linguistics to recover the details of that history. From the evidence of the ages, including the present, a composite picture of the linguistic past is assembled, piece by piece, according to established principles of investigation. The recovery effort requires a body of data, a theory of language change, and a set of methods which are sophisticated enough to lead to well-motivated historical inferences.

If the language under inquiry has a long and ample literary tradition, providing monuments of itself as it has moved through time, the task is somewhat simpler than it is if the language has no recorded history. The rich and deep written remains of English, for example, provide excellent material records which reach back more than 1,000 years. For this language, it is possible to trace the evolution of a single sound, form, word, or construction through written texts, revealing the patterns of change and development which have been at work in the language throughout its recorded history. Such a wealthy literary tradition puts the historian of English in an enviable position, with fixed points from which to proceed in the explanatory process, such as Old English texts, Middle English literature, and so on, right up to the present day. This wealth of data allows sophisticated accounts of the evolution of the structures of the language over a considerable span of time. But the English language did not begin at some arbitrary point which has been labeled "Old English" by historians of the language, say 700 C.E. Indeed, the history of English stretches well into the remote past, into a period before there was an identifiable language called English, and well before there were any written monuments to record its linguistic shape. It is the deeper task of the historian of English to go beyond the identifiable material documents and to stretch the methods of historical inquiry into periods before the advent of writing. Only when this is done can the history of the language be in any sense complete.

And what of Latin, a language whose first identifiable monuments stem from the 6th century B.C.E.? Can the linguistic history of such an "ancient" language be written? Unlike Modern English, Latin is the starting point, not the end point, of a literary tradition. Is it possible to go back in time to recover the history of a language with no direct antecedents providing evidence of some earlier documented stage? The history of English is revealed not only by Old, Middle, and Modern English materials, but also by the copious information provided on early dialects which has been transmitted in the texts. But there is no Latin before Latin as it is attested; how can the historical investigation of this language be carried out?

Though the task is somewhat more difficult than that confronting the historian of English (and somewhat easier than that confronting the historian of an unwritten language), the histories of Latin and of other ancient languages without recorded ancestors can not only be written, they can be written in detail. Historical investigation of such systems can proceed as long as there are other related languages available which furnish some material clues. That is, if it can be established that there is a genetic affinity between the language of inquiry and some other language or languages, research can proceed even in the absence of anterior written records and extensive texts. The analysis of the history of English, for example, is enhanced by materials from related languages such as Gothic, Old Icelandic, and Old High German, which allow the linguistic historian to proceed beyond Old English and to investigate the course of the language before the advent of written records, before it can even be identified as English. To write the history of an ancient language like Latin, it is necessary to make similar use of information provided by genetically related languages, analyzed with a consistent and rigorous methodology. Like English, Latin is a language which has, to pursue the customary genealogical analogy, sisters, cousins, aunts, and daughters, all of which carry clues about its history and that of its parent, Proto-Indo-European (PIE). The survey which follows provides a brief review of the methodology by which ancestral stages of languages are reconstructed, the evidence used to establish the existence of the Indo-European (IE) family of languages, and a summary of the languages which make up this, the most important language family in the world.

1.2. Methodological preliminaries

During the approximately two centuries in which the interrelationships among the IE languages have been systematically studied, techniques which demonstrate genetic affiliations among languages have been developed with considerable success. Chief among these techniques is the comparative method, which takes shared linguistic features as its data and provides procedures for establishing ancestral proto-forms. Since the focus of this chapter is Indo-European and not methods of reconstruction, the review of the comparative method will be based only on data from IE languages. This method, however, is not restricted in its applicability to IE languages. It is the methodological centerpiece of historical linguistic inquiry in language families of widely different size and structure. For more than a century now, specialists have been refining this and other methods which guide their efforts to retrace the linguistic past.

1.3. The comparative method and the regularity principle

There are many issues which must be considered to capture the central notions involved in the comparative method. Among these issues are the universality, replicability, and limits of the method. Universality is a central concern. Procedures which are effective only with certain languages, or types of languages, are of limited general value. Replicability is also a crucial issue. A method which yields different results to different investigators can lay no claim to a scientific basis. Finally, the limitations of the method are also of pivotal concern, as the time-depth and the structural consistency which can be attained by reconstruction techniques bear significant implications for the status of protolanguages.

Most research on linguistic change, past and present, has been guided by principles, methods, and assumptions which were formulated in the nineteenth century. It was during this period that the Indo-European, Semitic, and Finno-Ugric languages provided the main empirical foundations for modern theories of language change and contributed to the development of methods for reconstruction. Semitic was scientifically established as a linguistic entity by the late eighteenth century (the term "Semitic" was first used by August Schlözel in 1781), and the unity of Finno-Ugric was demonstrated by Sámuel Gyármathi (1799). This work was influential in the nineteenth century when such scholars as Rasmus Rask, who proposed an essay on Germanic phonology for a prize by the Danish Academy of Sciences in 1811 (published 1818), and Franz Bopp, who began the comparative linguistic analysis of Sanskrit (1816), were undertaking systematic investigations of the IE language family.

Work done on the IE languages during the nineteenth century continues to provide a methodological focal point for the conduct of studies of language relationships. This focus has resulted from the long and well-established traditions of the IE languages and of IE scholarship. There is a wealth of written data from this language family, with salient materials from nearly every branch. It has a wide geographical and cultural spread, extending from Europe to India at the time of its discovery and now spanning the globe. It has a rich time depth, with recorded data stretching deep into the second millennium B.C.E. And, the IE family has a relative homogeneity among its eleven well-attested branches which facilitates comparison and reconstruction.

The models of description and the associated theoretical paradigms which have emerged from the intensive comparative study of the IE and other languages contain at least the following central notions relating to the establishment of genetic units and the reconstruction of ancestral (proto)languages. These notions have been applied to non-IE languages as well, and their apparent universal utility is a basic premise of historical-comparative linguistic analysis.

1) A significant percentage of cognates (words with a common origin) in core vocabulary areas must be demonstrated in order to establish genetic affinity

between languages. This is a crucial step in the comparison of lexical items across languages, which allows the formation of reconstruction hypotheses (see table 1.1 for a list of such cognate forms). The demonstration of cognacy further involves the notion that there are reliable ways to recognize borrowings, such as Eng. *tsunami* from Japanese, or Eng. *paternal* from Latin; onomatopoeia, such as Eng. *bubble* and *buzz*; and accidents of linguistic history which result in chance lexical similarities, such as Eng. *man* and Karitiana *man* 'husband' (Karitiana is a Tupian language spoken in the Amazon region of Brazil). None of these can play a part in either the postulation of a genetic unit or in the reconstruction equation itself.

2) Phonological change is regular; that is, in any given language, all tokens of a particular phonological type will change in the same way across time, under like conditions. Because of the principle of regularity, in which exceptions and irregularities are explained by further subregularities, one expects to find consistent and systematic sound correspondences (matchings) between cognate forms in cognate languages. This is the classic nineteenth century position which is often labeled as the "Neogrammarian Principle" (after its nineteenth century originators), or the "regularity principle": regular sound change has no exceptions; when exceptions occur, they are governed by other recoverable rules. A claim of genetic relatedness for two or more languages must be substantiated by sets of recurring correspondences in phonological systems, and replicable reconstruction can only proceed if some form of the regularity hypothesis is adopted. The IE languages show such phonological regularity to a great degree. The well-known sound changes which define a set of changes from the reconstructed protolanguage into Germanic known as "Grimm's Law" provide an excellent example of such regularity. The exceptions to Grimm's Law, which are covered by Verner's Law[1] and other associated regular sound changes, provide the original cornerstone for what has proven to be a well-validated principle.

3) The regularity of systematic correspondences between related languages makes possible the application of the comparative method, a procedure for postulating reconstructed proto-forms on the basis of the attested evidence of the descendant languages. The comparative method is based on the principle that sets of recurring sound correspondences between two related languages continue blocks of positionally determined sounds from the mother language (Hoenigswald 1960: 132). The comparative method requires data from at least two languages, though many more can be placed into the reconstruction equation. The method is supplemented by the ancillary procedure of internal reconstruction, which generates hypotheses concerning the earliest recoverable evidence from within a single language by investigating the internal alternation of linguistic elements.

Reconstructions based on the application of the comparative method can be verified by reference to independently established structural principles (typology).

1 Grimm's and Verner's Laws are discussed in chap. 2. A useful summary of these and other "laws" in IE linguistics is Collinge (1985).

These principles, which in part define linguistic universals and establish parameters for structural configurations, provide an external check on the empirical validity of a reconstruction.

4) Historical processes of various sorts—phonological, morphological, lexical/semantic—can compromise sound changes, concealing shared features between languages. Auxiliary explanatory devices such as analogy (grammatical change), borrowing, or areal contact strengthen the notion of regular sound change by providing mechanisms to explain some kinds of irregularity, thereby accounting for the absence of firm correspondences. For example, the recognition that analogy can confound the operation of a regular sound change by either causing it not to apply, or by obscuring its results, has played an important role in the formulation of the comparative method. The rich array of analogical processes which work sometimes in conjunction, and sometimes at odds, with phonological processes has resulted in a strengthening of the notion of regular sound change. And while the unpredictable and non-formalizable nature of analogy, as well as its sometimes unrestricted explanatory power, have been points of interest and debate among historical linguists, its dominant role in language change, at least in the IE family, is undeniable.

Of course, it is reasonable to ask why it is that phonological systems remain full of irregularity when normal sound change is regular (excluding so-called sporadic changes, which can apply to some words and not others). The answer is a paradox. On the one hand, sound change is regular but often creates irregularity, especially in paradigms where one form may not fit the environment of the sound change at work. On the other hand, analogy, which is irregular and unpredictable in its operation, establishes regularity by eliminating allomorphy within paradigms. Take the example of Latin rhotacism, whereby (roughly) *s* becomes *r* between vowels, using the word *honor* 'honor' for illustration.[2] The Old Latin form of the nominative of this word is *honōs*. By the normal application of the rhotacism rule, the paradigm of *honōs* looks as follows (singular only).

nom. *honōs*
gen. *honōris*
dat. *honōrī*
acc. *honōrem*
abl. *honōre*

This paradigm is irregular in that there is an allomorphic variation between the nominative form *honōs* and the oblique stem *honōr-*. The irregularity is resolved by the analogical restructuring of the nominative form to *honor* (with regular shortening of the *ō* before final *-r*), thereby re-establishing a monomorphemic stem *honōr-* and restoring paradigmatic regularity. Despite the regularity that it creates, the ana-

2 This example is deliberately oversimplified. A more detailed discussion of rhotacism is provided in chap. 6.

logical restructuring of the stem to *honōr-* throughout the paradigm appears to be unpredictable. Other candidates for such a restructuring process like *corpus*, *corporis* 'body' never developed a nominative form **corpor*,[3] and there is really no way to know in this case why *honōs* was affected while *corpus* was not.

5) The relatability of irregular forms, especially irregular morphological forms, is an index of genetic affinity among languages and an important tool in the recovery of proto-stages. Implicit in this notion is the assumption that language change in general is not only regular in its operation, but that it leads to greater regularity in the resulting linguistic system (though it is certainly true that regular sound change can create irregular outputs). Such regularity is consistent with the notion that speakers must be able to preserve communication across generations, and that the system which results from a change or series of changes must be internally consistent. Irregular forms (i.e. those not derivable by general rule) are for the most part historical residue and are relatively limited in a language system; they are not generally introduced in the process of language change. Furthermore, when such irregularities, especially morphological irregularities, are found to correspond across languages, they provide a clear index of genetic relatedness. For example, the widespread irregularity of the verb 'to be' in IE languages indicates quite clearly that the various languages in which this verb is represented must be related to each other. Not only is 'to be' irregular, but it is consistent in its irregularity across languages. See table 2.18 on p. 76 for a few sample paradigms of this verb.

6) In language families with traditions of literacy, sound changes can often be recovered from written data. Because of the lengthy literary histories of many of the IE languages, there is considerable orthographic evidence of changes in individual languages as they are recorded through time. The interpretation and understanding of these orthographic representations requires a thorough philological understanding of the texts themselves and a sensitivity to the writing habits and traditions of diverse groups, from illiterate stonecutters inscribing monuments, to everyday citizens scribbling graffiti, to semi-educated scribes recording documents, to skilled language users. Such expertise enables investigators to plot changes in the physical linguistic record, to interpret them, and to integrate them into the theory of sound change.

The six criteria just discussed, as well as the many additional subcriteria and associated methodologies, are based on a conception of language change which is for the most part phonologically based. That is, the techniques of linguistic reconstruc-

---

3 An * before a segment or morph indicates that the form is reconstructed. Most often in this volume the * will indicate that the form is reconstructed PIE, e.g. **h₁ekuo-* 'horse'. Frequently, however, the * is used to represent an unattested form of pre-Latin, pre-Greek, etc. (e.g. *generis* 'of the race' < **geneses*, Osc. *factud* 'do' < **fakitōd*), or a possible but unattested form (e.g. Lat. **corpor* 'body'). A ** indicates that a word-form is not only unattested, but impossible. Ungrammatical syntactic constructions are indicated by a †, e.g. Lat. †*cum mē* for *mēcum* 'with me'.

tion which have proven especially useful in the recovery of ancestral stages in the IE languages are most appropriately characterized in terms of phonological processes (see Kiparsky 1995). The "ripple effect" of phonological change as it reshapes morphs, and ultimately can result in a reorientation of grammatical categories and processes, is a characteristic feature of change in the IE family. This is not to say, of course, that there are no independent morphological or syntactic changes, or that there is no independently established comparative morphology or comparative syntax. In general, however, the regularity principle is a principle of phonology.

The foregoing notions concerning linguistic change and reconstruction have led to fruitful results in the understanding of the IE family and have set the methodological standard for research on language change in studying non-IE languages as well.

The claim that two or more languages are genetically related carries with it the associated claim that the two share common ancestry. Methods of historical investigation should provide a means of recovering the ancestral system, attested or not. The initial demonstration of relatedness is in many ways the easy part; establishing well-motivated intermediate and ancestral forms is quite another matter. Among the difficulties are the following: How many shared similarities are enough to prove relatedness conclusively, and what sort of similarities must they be? Which features in which of the languages being compared are older? Which are innovations? Which are borrowed? How are similarities and differences evaluated for significance? What assumptions are made about the relative importance of lexical, morphological, syntactic, and phonological characteristics, and about directions of language change?

Though all of these questions come into play in any reconstruction effort, there is an initial assumption: if two or more languages share a feature which is unlikely to have arisen by accident, borrowing, or as the result of some typological tendency or language universal, then it is assumed to have arisen only once and to have been transmitted to the two or more languages from a common source. The greater the number of shared features that are discovered and securely identified, the sounder the relationship. One "rule of thumb" in IE reconstruction is that for a morph to be projected back to PIE, it should be represented in at least three descendant language groups, and if only three, they should not all be geographically contiguous. Another stipulates that at least one member of the European and one member of the Asian sides of the family be represented before the form can be securely postulated for the parent language.

In determining genetic relationship and reconstructing proto-forms using the comparative method, it is customary to start with vocabulary. A list of possible cognates which is likely to produce a maximum number of common inheritance items, known as the basic vocabulary list, provides many of the words to be investigated, such as first-order kinship terms, pronouns, body parts, lower numerals, and others.

Table 1.1. Some basic Indo-European terms

A. Numerals from 1–10

	one	two	three	four
Skt.	ēkaḥ	dvá(u)	tráyaḥ	catváraḥ
Gk.	οἶνη 'ace'[a]	δύ(Ϝ)ο δύω	τρεῖς	(Att.) τέτταρες, (Hom.) τέσσαρες
Lat.	ūnus	duo	trēs	quattuor
Hitt.		dā-	teri-	
Toch. A		wu	tre	śtwar
B		wi	trai	śtwer
OIr.	óen	dáu, dó	tri	ceth(a)ir
Goth.	ains	twai	þreis	fidwōr
OCS	(jed-)inŭ	dŭva	trĭje	četyre
Lith.	víenas	dù	trỹs	keturì
Arm.	mi[b]	erku	erek'	č'ork'
Alb.	një, nji	dy	tre, tri	katër

	five	six	seven	eight
Skt.	páñca	ṣáṭ-	saptá	aṣṭá(u)
Gk.	πέντε, πέμπε	ἕξ	ἑπτά	ὀκτώ
Lat.	quīnque	sex	septem	octō
Hitt.			šipta-	
Toch. A	päñ	ṣäk	ṣpät	okät
B	piś	ṣkas	ṣukt	okt
OIr.	cóic	sé	secht n-	ocht n-
Goth.	fimf	saíhs	sibun	ahtau
OCS	pętĭ	šestĭ	sedmĭ	osmĭ
Lith.	penkì	šešì	septynì	aštuonì
Arm.	hing	vec'	ewt'n	ut'
Alb.	pesë	gjashtë	shtatë	tetë

	nine	ten
Skt.	náva	dáśa
Gk.	ἐννέ(Ϝ)α	δέκα
Lat.	novem	decem
Hitt.		
Toch. A	ñu	śäk
B	ñu	śak
OIr.	nói n-	deich n-
Goth.	niun	taíhun
OCS	devętĭ	desętĭ
Lith.	devynì	dẽšimt
Arm.	inn	tasn
Alb.	nëndë	dhjetë

Table 1.1. Some basic Indo-European terms *(continued)*

B. Animal names

	mouse	wolf	cow	sheep
Skt.	mū́ṣ-	vṛ́kaḥ	gáuḥ	áviḥ
Gk.	μῦς	λύκος	βοῦς, Dor. βῶς	ὄ(Ϝ)ις
Lat.	mūs	lupus	bōs	ovis
Hitt.			wa-wa-(i)- (Hier. Luw.)	ḫawi- (Hier. Luw.)
Toch. A			ko	
B	maścītsi (pl.)	walkwe	keu	(pl.) āwi(?)
OIr.		olc 'evil'	bó	ói
Goth.	mūs (OHG)	wulfs	kýr (OIc.)	ouwi (OHG)
OCS	myšĭ	vlĭkŭ	gumĭno 'threshing floor'	ovĭca
Lith.		vil̃kas	gùovs (Latv.)	avìs
Arm.	mukn	gayl	kov	hoviw 'shepherd'
Alb.	mi	ulk	ka	

	pig	dog	horse
Skt.	sūkaráḥ	śvā́	áśvaḥ
Gk.	ὗς	κύων	ἵππος, Myc. i-qo
Lat.	sūs	canis	equus
Hitt.			azu(wa) (Hier. Luw.)
Toch. A		ku	yuk
B	suwo	ku	yakwe
OIr.		cú	ech
Goth.	swein	hunds	eoh (OE)
OCS	svinija	súka (Russ.) 'bitch'	
Latv.	su/ivēns 'young pig'	šuõ (Lith.)	ešvà, ašvà (OLith.) 'mare'
Arm.		šun	ēš 'donkey'
Alb.	thi		

C. Body parts

	foot	heart	eye	tongue
Skt.	pā́t		ákṣi	jihvā́
Gk.	Att. πούς (gen. ποδός)	καρδίᾱ	ὄμμα, ὄψομαι 'I will see'	
Lat.	pēs (gen. pedis)	cor (gen. cordis)	oculus	lingua (OLat. dingua)
Hitt.	pata- (Hier. Luw.)	kard-		
Toch. A	pe	kri	ak	käntu
B	paiyye	kāryāñ (pl.)	ek	kantwo
OIr.	ís 'below'	cride	enech	teng
Goth.	fōtus	haírtō	augō	tuggō
OCS	pěšĭ 'on foot'	srŭdĭce	oko	języky̆
Lith.	pādas 'sole'	širdìs	akìs	liežùvis
Arm.	otn	sirt	akn	lezu
Alb.	(për)posh 'under'		sy	

Table 1.1. Some basic Indo-European terms *(continued)*

D. Kinship terms				
	mother	father	sister	brother
Skt.	mātā́	pitā́	svásā	bhrā́tā
Gk.	(non-Att.) μᾱ́τηρ	πατήρ	ἔορ (voc.) 'daughter'	φρᾱ́τηρ 'member of a phratry'
Lat.	māter	pater	soror	frāter
Hitt.				
Toch. A	mācar	pācar	ṣar	pracar
B	mācer	pācer	ṣer	procer
OIr.	máthir	athir	siur	bráth(a)ir
Goth.	mōðir (OIc.)	fadar	swistar	brōþar
OCS	mati		sestra	bratrŭ, bratŭ
Lith.	mótė 'woman'		sesuõ	brólis
Arm.	mayr	hayr	k'oyr	elbayr
Alb.	motrë 'sister'			

E. General terms				
	full	family, race	month	dead, death
Skt.	pūrṇáḥ	jánaḥ 'person'	mā́saḥ	mṛtáḥ
Gk.	πλήρης	γένος	μήν	ἄμβροτος 'immortal'
Lat.	plēnus	genus	mēnsis	mortuus
Hitt.				mer-, mir- 'to die, disappear'
Toch. A			mañ	
B	pällew (full, of moon)		meñe	
OIr.	lán	gein 'birth'	mí	marb
Goth.	fulls	kuni	mēna, mēnōþs	maúrþr 'murder'
OCS	plŭnŭ		měsęcĭ	mĭrǫ, mrěti
Lith.	pìlnas		mė̃nuo	miřti
Arm.	li	cin 'birth'	amis	meṙanim 'die'
Alb.	plot		muaj	

a Gk. οἶνη 'ace', as on dice, is the appropriate cognate for comparison in this chart. However, it does not exist as part of the counting system. Greek makes use of the other PIE form designating 'one', viz. **sem-*, seen in εἷς, μία, ἕν.

b Arm. *mi* is related to Gk. εἷς, μία, ἕν 'one' (< **sem-*).

From these and other data, sets of equations (correspondences) are established. Correspondence statements stipulate that in a given environment, phoneme **X** of one language will correspond to phoneme **Y** of another language consistently and systematically if the two languages are descended from a common ancestor. See table 1.1 for some examples.

The comparative method can be illustrated through the examination of a few lexi-

Table 1.2. Selected lexical items illustrating the comparative method

	'mouse'	'mother'	'nine'
Skt.	mū́ṣ-	mātā́	návaḥ
Gk.	μῦς	μάτηρ	ἐννέ(Ϝ)α
Lat.	mūs	māter	novem
OHG	mūs	mōðir (OIc.)	niun (Goth.)
	'dead'	'dog'	'family, race'
Skt.	mṛtáḥ	śvā́	jánaḥ
Gk.	ἄμβροτος	κύων	γένος
Lat.	mortuus	canis	genus
Goth.	maúrþr	hunds	kuni
	'I am'	'I vomit'	'old'
Skt.	ásmi	vámiti 'X vomits'	sánaḥ
Gk.	εἰμί	ἐμέω	ἔνος 'last year's'
Lat.	sum	vomō	senex
Goth.	im	vāma (OIc.) 'sickness'	sineigs

Table 1.3. Nasals aligned for comparison

	'mouse'	'mother'	'nine'	'dead'	'dog'	'family, race'	'I am'	'vomit'	'old'
Skt.	m-	m-	-n-	m-	-n	-n-	-m-	-m-	-n-
Gk.	m-	m-	-nn-	-m(b)-	-n	-n-	-m-	-m-	-n-
Lat.	m-	m-	-n-	m-	-n-	-n-	-m	-m-	-n-
Gmc.	m-	m-	-n-	m-	-n-	-n-	-m	-m-	-n-

cal items from IE sister languages,[4] restricting the data to fairly clear cases (table 1.2).[5] The discussion will be restricted to the nasals *m* and *n*, arranged for comparison in table 1.3.

Reconstruction can only be undertaken with appropriately compared segments. It is clear that nasal segments match in 'mouse', 'mother', 'dog', 'family, race', 'I am', 'vomit', and 'old', but this is less clear in 'nine' and 'dead'. What of the double νν in Gk. ἐννέ(Ϝ)α? One possibility is that εν- is a prefix, placing the first ν outside the equation.[6] Similarly with ἄμβροτος 'immortal': the α- is a prefix meaning

4 The two most complete etymological accounts of the PIE lexicon are Pokorny (1951–1959) and Watkins (1985). A semantically organized dictionary of the IE languages is Buck (1951). Unless otherwise specified, words in this and the following tables can be assumed to be cognates.

5 In the interest of avoiding gender-bias in glosses, third person verb forms will be glossed with X instead of 'he', 'she', 'it', or some combination of these forms.

6 It is more likely that the difference in the initial segments of Greek and the other languag-

'not' (= Lat. *in-*, Goth. *un-*, etc.), and the β results from a tendency in Greek to break up sonorant clusters like -μρ- (also -μλ-) through the insertion of -β-, resulting in -μβρ- (cf. the parallel process in Lat. *camera* > Fr. *chambre* 'room'). So the *m*'s do indeed align, leaving a consistent set of *m* and *n* correspondences (figure 1.1).

m : m : m : m n : n : n : n

Figure 1.1. Comparative alignment of nasal segments

These alignments represent the horizontal or comparative dimension. Now the segments are "triangulated", adding the vertical, or historical, dimension (figure 1.2).

Figure 1.2. Triangulation of nasal segments

After all the relevant data have been checked and their distributional patterns investigated, a hypothesis can be made concerning the proto-sound. In these two cases the most immediately evident hypothesis is that the proto-sounds were **m* and **n* (figure 1.3).

Figure 1.3. Reconstruction of **m* and **n*

This analysis leads to the statement that **m* > ('develops into') *m* and **n* > *n* in the various daughter languages.

Invariable correspondences such as these are more the exception than the rule in historical-comparative linguistics. It is far more common to find sets in which only a few of the members have identical segments. But the method of comparative reconstruction, when supplemented with sufficient information about the internal structure of the languages in question, can still yield strong results, as illustrated in table 1.4.

At issue are the correspondences which include *s*, *h*, *ḫ*, and *r*. In 'stand' the correspondence set is *s* : *s* : *s* : *s*, with the initial *h* of Gk. ἵστᾱμι the result of reduplication of the root initial *s* and its replacement by *h* (i.e. < **sistāmi*). This replacement

---

es is due to the fact that the word for 'nine' began with a "laryngeal" consonant (**h*$_1$) in the protolanguage which is recoverable as ε(ν)- in Greek, but was lost in the other languages. Laryngeals are discussed in chaps. 2 and 6.

Table 1.4. Selected Indo-European cognates

	'stand'	'old'	'tooth' (gen.sg.)	'be' (various forms)
Skt.	sthā- 'stand'	sánaḥ	datáḥ	ástu 'let X be!'
Gk. (Dor.)	ἵστᾱμι 'I stand'	ἕνος	ὀδόντος	ἔω (ᾖ) 'if, whenever I am'
Lat.	stāre 'to stand'	senex	dentis	erō 'I will be'
Goth.	standan 'to stand'	sineigs	(OCS slovese 'word's')	ist 'X is'

reproduces the pattern found in the initial sound of the set 'old', namely *s* : *h* : *s* : *s*. In final position there is the correspondence *ḥ* : *s* : *s* : *s* in 'old' and 'tooth', and in medial position *s* : ∅ : *r* : *s* in 'be' (figure 1.4). What is or are the proto-sound(s)?

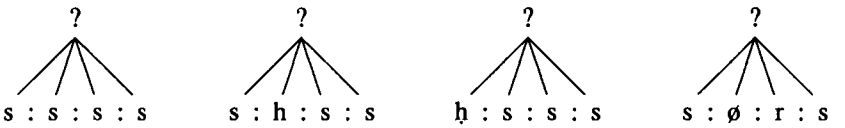

Figure 1.4. Reconstructing **s*

An analysis of these data leads straight to **s* for all four correspondences. For the first set 'stand' there is complete agreement of correspondences, and **s* seems a reasonable reconstruction in the absence of evidence to the contrary. For the second set, a pattern can be discerned by postulating that **s* > *h* in Greek initially (ἵστᾱμι, ἕνος), a lenition (weakening) process found in a number of languages. Such weakening is further evidenced in the forms γένεος and ἔω, where medial **s* > ∅, in which the process has completely obliterated the original **s*. The final *ḥ* of Sanskrit *sánaḥ* and *datáḥ* results from a specific rule of Sanskrit by which final *s* after vowels becomes *ḥ*. And the medial Latin *r* in *erō* is the result of rhotacism, whereby Latin converts intervocalic *s* to *r* (cf. *flōs* 'flower', (gen.) *flōris*). These hypotheses can be verified with countless similar examples.

From these few, admittedly simplified, illustrations it can be seen that the comparative method, when supplemented by adequate information about the internal structure and phonetic processes of the languages in question, and by a consideration of other relevant data, can produce consistent reconstructions of ancestral forms. It is with such methods that PIE is being reconstructed. For a review of comparative reconstruction methodology, see Fox (1995).

Research on a wide variety of the world's languages has shown that there are many strict phonological correspondences of the type found in IE languages. For example, some Native American languages allow for the orthodox application of the comparative method of reconstruction with consistent results. Similar results in both grouping and reconstruction have been achieved in the Austronesian, Dravidian, Finno-Ugric, and many other families. The end product has been a continuous

vindication of the comparative method and the repeated demonstration that its applicability extends beyond its original (Indo-European) domain.

But doubt about the method persists in some quarters. First, it has been asked whether the method is both circular and too powerful, in effect guaranteeing its own success. Critics claim that in the actual application of the comparative method, linguists do something like the following:

a. Make an informed guess about the relatedness of the languages to be compared;
b. Decide what forms are cognate on the basis of gross structural similarity and loose semantic relatability;
c. Reconstruct proto-forms by applying internal and comparative reconstruction methods;
d. Explain away exceptions and irregularities by appeal to analogy and other auxiliary notions;
e. Build "semantic bridges" which adjust results in order to arrive at a more or less unique form with a consistent sound–meaning correspondence.

Critics of the traditional comparative method consistently point to the relatively small number of "good cognates" (i.e. those in which appeal to explanatory devices outside the comparative method is unnecessary), as well as to the often wide disagreement among scholars on the reconstructions themselves, especially those involving polymorphemic structures. But these objections ignore the essential soundness of the method and the continuous fund of confirmed results that it yields when properly applied. The comparative method, despite its restrictions and limitations, provides the most reliable method for reconstructing linguistic history. The method may have some built-in circularity, but the consistent results of the past century make this circularity a reasonable price to pay.

1.3.1. Challenges to the comparative method and the regularity principle

The strongest theoretical challenge to the regularity doctrine is the theory of lexical diffusion, as originally developed by Wang (1969). Wang's claim is that sound changes spread gradually across the lexicon ("words change"), and that phonemes frequently are not the minimal units of historical phonological change. This contrasts with the classic (Neogrammarian) regularity position that change affects all words which include the sound in the appropriate phonetic environment ("sounds change"), and that change proceeds mechanically. Wang points to lexical doublets present in the speech of individuals as evidence for the gradual drift of some words. Such coexisting forms as [kǽtəlag] vs. [kǽtəlɔg] 'catalog' and [sərprájz] vs. [səprájz] 'surprise' seem to indicate that at least some phonological changes are not abrupt and can leave behind variant forms in individual words which may persist for some time, even in the speech of a single person.

Labov (1981, 1994) has argued against the strong form of lexical diffusion in which all change is carried out lexically, without phonetically motivated sound changes. Through a reexamination of some of the principal test cases for lexical diffusion, plus an instrumental analysis of data from urban American changes in progress, Labov has concluded that "there appear to be far more substantially documented cases of Neogrammarian sound change than of lexical diffusion" (1994: 471), confirming the status of mechanical sound change as the dominant mechanism. Labov's conclusions strengthen the operational principle that historical linguists should carry out their research as if all change were regular. See further Wang and Lien (1993), who argue for an interaction between contact-induced change and internal change, and Kiparsky (1995), who defends the Neogrammarian position.

In addition to the disagreement over the integrity of the comparative method itself, questions have been raised for decades over its applicability to languages of widely divergent typologies, or to languages with social, cultural, and geographical characteristics different from those of the IE languages. Doubt seems to flow along several different lines.

First is the matter of the phonological basis of linguistic change which is so much in evidence in the IE languages. The comparative method relies on sound correspondences. Can it be applied to language families in which sound correspondences are irregular and inconsistent? Boretzky (1984), for example, has argued that change in the Arandic languages of central Australia seems to proceed more by abrupt lexical replacement through borrowing than by gradual phonological change. Moreover, these lexical replacements are frequent and cut deep into the core vocabulary commonly held to be resistant to borrowing. Furthermore, the conditions according to which these replacements take place are difficult to imagine in large language families with many speakers, considerable geographical spread, and complex stratified societies. These conditions include borrowing/replacement because of death taboos, which proscribe the use of common nouns contained in the names of tribe members who have died. Also at work are such phenomena as secret languages, mother-in-law languages, and language loyalty and shift according to marriage patterns, all of which reflect social processes in this highly unusual linguistic setting.

A second area of disagreement over the comparative method concerns its usefulness in the establishment of long-distance linguistic relationships and remote protolanguages. Whether linguists can reliably reconstruct distant genetic relationships and establish phylum-level protolanguages is a matter not only of method, but of philosophy of science. At issue are such questions as the following: What counts as data? What are the requirements on semantic relatability? What are the ultimate determinants of cognacy? How much strict regularity is required in phonological equations? How much internal reconstruction should be done prior to comparative reconstruction? What is an acceptable time-depth? Is it legitimate to use a reconstructed protolanguage as an intermediate stage in the reconstruction of an even

Table 1.5. Multilateral comparison[a]

	'one'	'two'	'three'	'hand'	'eye'	'ear'	'mouth'
1	sang	•su	soti	yung	•sing		•a
2	wate	◊iba	•tati	ju(le)	no(do)	•to(go)	yabodo
3	•toro	•ču	agozo	•daho	•samo	◊sumo	◊či
4	ili	◊iwa	•ita	ilɔ	ewu	•ɔtɔ	♦ɛnu
5	mwe	bali	•tato	◊(li-)to	(le-)iso	•(ku-)toi	♦(mu-)nywa
6	•tilo	ndi	yasko	kela	•sim	◊sumo	◊či
7	kiet	◊iba	•ita	◊ɛte	enyin	•utong	♦inua
8	lakoi	•swe	we	•taha	i	kebbe	•a

a Languages 1, 3, 6, 8 = Saharan; 2, 4, 5, 7 = Niger-Congo. From Greenberg (1954: 408).

more remote protolanguage?

Greenberg has claimed (1987) that the native languages of the Americas can be classified into three main genetic groups, with ninety percent of them in what he calls "Amerind" and the other ten percent in the groups he calls "Na-Dene" and "Eskimo-Aleut". Though most specialists agree on the isolation of Na-Dene and Eskimo-Aleut as separate families, the postulation of one super-group (Amerind) containing all the rest contrasts with the more general view, which holds that perhaps as many as one hundred fifty or even two hundred independent families can be established, and according to which deeper relationships are so far poorly demonstrated. Greenberg's method is the same one he used for his classification of the languages of Africa (Greenberg 1963a). His classification methodology is based not on the application of the orthodox comparative method, but rather on the technique of "multilateral comparison". The multilateral approach considers a relatively small number of words across a relatively large number of languages and attempts to expose similarities in form and meaning which reveal genetic relatedness among the languages. The technique of multilateral comparison is illustrated in table 1.5, where eight languages are compared (the vertical column) across seven core lexical items (the horizontal column). Purported cognates, based on superficial phonetic similarities, are marked in each column by the identifying symbols •, ◊, and ♦. Words without identifying marks are considered to be non-cognate forms. The resulting overall pattern is claimed to reveal a distribution of cognate forms such that languages 1, 3, 6, and 8 form one group, while languages 2, 4, 5, and 7 form another.

The large number of languages used in the comparison (though not in this illustration) typically increases the chance for similarities, and the lack of a phonological correspondence requirement across cognate forms makes the procedure considerably more flexible than the comparative method as described earlier. Of course, it is significant that the goal of multilateral comparison is not reconstruction, but rather classification. Greenberg rejects the traditional idea that regular phonological correspondences are the only certain way to establish linguistic relationships. For a dis-

Table 1.6. Examples of reconstructed Proto-Nostratic[a]

1.	PN **bur-/bor-* 'to bore, to pierce'	PAA **bər-/bar-* 'to bore, to pierce'; PIE **b^{h}or-/b^{h}ṛ-* 'to bore, to pierce'; PU **pura* 'borer, auger'; PD **pur-* 'gimlet, borer, to bore, to perforate'; PA **bur-* 'to bore a hole'; S *bùr* 'to bore through, to pierce'.
2.	PN **t'ar-/t'er-* 'to tear, to rend, to cut, to sever'	PIE **t'er-/t'or-/t'ṛ-* 'to tear, to rend, to flay'; PD **tar-* 'to cut, to cut off, to lop off, to strip off', **ter-* 'to burst asunder, to split, to cut, to cut off'; S *dar* 'to split'.
3.	PN **tl^{h}unk^h-/tl^{h}onk^h-* 'to hook up, to hang, hanging, dangling, peg, hook'	PAA **tl^{h}ənk^h-/tl^{h}ank^h-* 'to hook up, to hang, peg, hook'; PIE **k^{h}onk^h-* 'to hook up, to hang, peg, hook'; PD **cuṅk-* 'piece hanging out or dangling'.
4.	PN **wig^y-/weg^y-* 'to carry, to convey'	PAA **wəg^y-/wag^y-* 'to carry, to weigh'; PIE **ueg^h-/uog^h-* 'to carry, to convey, to weigh'; PFU **weye-/wiye-* 'to bring, to carry, to convey'.

a Abbreviations: PN = Proto-Nostratic; PA = Proto-Altaic; PAA = Proto-Afro-Asiatic; PD = Proto-Dravidian; PFU = Proto-Finno-Ugric; PIE = Proto-Indo-European; S = Sumerian. From Bomhard (1990: 346–353).

cussion of the method, see now Ringe (1995), who argues that the predictions of multilateral comparison are successful only at the level of chance. For a defense of the Greenbergian approach, see Ruhlen (1994a, 1994b). For one of the few analyses of the issues which works from primary data on the remote relations of some languages (Nostratic, Sino-Caucasian, and Dene-Caucasian), see Shevoroshkin and Manaster Ramer (1991). For a review of Greenberg's classification of the languages of the Americas, see the reviews by Chafe, Darnell, Goddard, Golla, Hymes, Rogers, and Sapir in *Current Anthropology* 28 (1987): 647–667, as well as Campbell (1988), and the extensive discussion in McMahon and McMahon (1995).

There are reconstructions in other long-distance work, however. Of particular note are efforts to reconstruct (one version of) Proto-Nostratic as a parent language of the Afroasiatic, Kartvelian, Indo-European, Uralic, Dravidian, and Altaic language families, and possibly Sumerian as well. Though no less controversial than Greenberg's classification of the Americas,[7] the reconstruction of Proto-Nostratic has proceeded along more traditional methodological lines, including the use of sound correspondences. The Nostratic issue has attracted much attention, not only in the scholarly community (see Bomhard and Kerns 1994, Ringe 1995, Bomhard

7 Greenberg himself (to appear) has proposed an even more inclusive membership for Nostratic. In Greenberg's view, the Nostratic macrofamily comprises the following language families: Afroasiatic, Kartvelian, Sumerian, Elamo-Dravidian, Indo-European, Uralic-Yukaghir, Altaic, Chukchi-Kamchatkan, Gilyak, and Eskimo-Aleut. The latter six (from Indo-European to Eskimo-Aleut) are members of the Eurasiatic subgroup of Nostratic.

1996, Joseph and Salmons to appear), but also in the general press.[8] A few examples of Nostratic reconstructions appear in table 1.6.

Most scholars of traditional IE linguistics reject the method of multilateral comparison, as well as the concept of Nostratic. No particular side on the issue is taken in this book, though it is clear that for the reconstruction of PIE proper a method like that of multilateral comparison is superfluous.

Two other issues are frequently raised in connection with the goals and integrity of the comparative method and the reconstructions which result from its application. The first of these concerns the linguistic reality to be assigned to the reconstructed proto-units, while the second concerns the uniformity of protolanguages. On the matter of reality, one school of thought (Pulgram 1959) holds that a reconstruction is little more than an abstract fiction, algebraically constructed, which does not reflect any linguistic reality. On this view, a reconstruction is a formula which logically, not perceptually, embodies a collection of proto-sounds (phonemes) which are assembled into meaning units (morphemes), and then into grammatical constructions (syntax). There is no claim that the reconstructions which result actually represent elements of a real language. Opposing this position (Hall 1960, 1974) are those who claim that a reconstructed form bears a strict phonetic relation to actually occurring historical forms, so much of a relation, in fact, that vanished speakers of the language in question might be able to understand it if they were somehow brought back to hear the reconstructed forms uttered.

The resolution of the reality question depends on two principal factors, namely time-depth and the nature of the evidence. It goes without saying that the potential reality of a reconstruction decreases as the time-depth increases. Thus reconstructed classical PIE has a greater chance of being phonetically real than does early PIE, which in turn has more potential reality than Proto-Nostratic. Indeed, Schleicher even went so far as to reconstruct a Proto-Indo-European fairy tale in 1868. As for the evidence, Hall's realist position is surely more reasonable for the reconstruction of Proto-Romance, a system which is not only chronologically closer to the oldest Romance language documents than even the older IE languages are to PIE, but also one which can be "checked" against the evidence of attested Latin.[9]

As is usually the case in such debates, the best approach lies somewhere between the two extremes (see Lass 1993 for discussion). Some aspects of PIE phonology, for example laryngeal consonants, seem beyond phonetic perceptibility, while others, for example the voiceless stops, seem phonetically secure. A similar state of affairs holds in the reconstruction of Proto-Romance. The consonants, for example,

---

8 For example, popular articles have appeared in such sources in the US as *The Atlantic Monthly* and *Scientific American*, both in April 1991, in *Scientific American* in January 1994, and in the *New York Times* on June 27, 1995; in France, in *Le Monde* on January 23, 1997, and in *Le Nouvel Observateur*, also in January, 1997.

9 As Hall states clearly (see his chart 1974: 14–15), the data provided by the written classical language can be used as a template to verify reconstructions of Proto-Romance.

are relatively straightforward, while the vocalic systems are quite unstable in reconstructions and cannot be posited with any phonetic guarantees.[10] In the final analysis, phonetic information has surely figured in PIE reconstructions, though verifiability remains elusive.

The second issue concerns the status of protolanguages as uniform, dialect-free entities without recoverable synchronic variation. This is a particularly troublesome question which defies simple resolution. It is easily demonstrated that natural languages always exhibit dialect differences. These differences may result from an intersection of social and geographical factors, with consequent variation in the speech community. Generally speaking, the more social stratification and geographical spread, the more variation. In the reconstruction of a protolanguage it is difficult to capture synchronic variation. This is because the data from language to language are not consistent, are often from widely divergent geographical areas with different cultural and external influences, and often exhibit considerable chronological disparities. These factors alone make the postulation of a single dialect-free system somewhat hazardous. At the same time, the comparative method demands a certain homogenization and idealization of the data in order to achieve proper matchings and defensible reconstructions. The result is typically the reconstruction of a dialect-free proto-system which may bear up to methodological scrutiny but probably does not reflect the linguistic realities of any given period. It appears to be an unavoidable problem. For discussion see Anttila (1989: 282–284).

For the present it can be said that among scholars of historical-comparative linguistics, there is a common opinion: the comparative method works (see the papers in Baldi 1990, Fox 1995). While it is true that the method has flaws, it nonetheless allows for a general replicability which validates its internal integrity. Furthermore, the comparative method has a wide applicability, not only to IE languages and languages of similar typological structure, but to languages and language families which differ markedly from these. This includes the languages of Australia, in contrast to what Boretzky has claimed (Dixon 1990). Finally, on the matter of the limits of the method, there is less agreement, but still consensus. The comparative method has limits, both practical and methodological. The people who use this method are restricted by human constraints, and the limitations on availability and accessibility of all the data required to make remote speculations are severe.

---

10 Comrie (1993: 96 n. 2) asserts that reconstructions in Pulgram's sense, namely convenient formulaic summaries of the relations among the attested genetically related languages, are without theoretical interest. Inasmuch as reconstructions are based on phonetic and phonological observations about the behavior of sounds in known languages, he is surely right. However, the postulation of abstract segments like laryngeals will not withstand close phonetic scrutiny; their analysis depends purely on logical distributions.

1.4. The method of internal reconstruction

After the comparative method, the method of internal reconstruction provides the most powerful procedure of historical analysis, supplementing comparative analysis to achieve reliable reconstructions. While the comparative method relies on patterned relations across different languages, the method of internal reconstruction relies on the patterned relations of linguistic units within a single language which point the way to earlier structures of the language in question. The principle behind the method of internal reconstruction is that individual languages contain within themselves evidence of their own history. This evidence takes several shapes, namely language-internal alternations and exceptions.

The primary source of internal evidence is found in alternations between linguistic units in recognizable distributions. In phonology, complementary correspondences of a given phoneme within a language can be taken to represent reflexes of a single proto-phoneme (Fox 1995: 146–147). The alternations often arise from historical sound changes which resulted in phonological splits, creating different surface realizations of the same proto-unit. Consider the following Latin data.

	'flock'	'king'
nom.	grex	rēx
gen.	gregis	rēgis

At issue is the status of the final consonant of the base form. It can be seen that the nominative of these two words shows *x*, i.e. [k] followed by the ending [s], while the genitives contain *g*, i.e. [g] followed by the ending [is]. This suggests a phonological rule of Latin by which voiceless consonants are voiced intervocalically, i.e. [k] → [g], making the unitary form of the bases /grek-/ and /rēk-/. Now consider the following two partial paradigms.

	'leader'	'wild'
nom.	dux	ferōx
gen.	ducis	ferōcis

If the proposed voicing rule were in effect, the expected genitive forms of 'leader' and 'wild' would be **dugis* and **ferōgis*, respectively. In fact the opposite distribution is in effect, namely that voiced sounds like *g* are devoiced before voiceless consonants like *s*, viz. [g] → [k], indicating the base forms /greg-/, /rēg-/, /duk-/, and /ferōk-/. The original voiceless [k] of 'leader' and 'wild' is left unaffected by this rule. The existence of the verbal forms *(con)gregō* 'I gather', *regō* 'I rule', and especially *dūcō* 'I lead' supports at least the first three internally reconstructed root forms.

A second source of evidence which can be utilized in internal reconstruction is found in exceptional structures, especially morphological structures. For example, the irregular, non-productive alternation of the vowels in such words as Eng.

steal/stealth, *creep/crept*, and *sleep/slept* results from a process which was productive at one time in English but which has ceased to operate. In Latin, perfect forms such as reduplicated *cecidī* 'I fell' (pres. *cadō*) and *tetigī* 'I touched' (pres. *tangō*) represent archaic morphological processes which are marginal in the synchronic system. Such forms represent historical residue which for one reason or another failed to be remodeled analogically to the productive patterns, making them morphological anomalies and affording them historical primacy (Fox 1995: 150).

The results of internal and comparative reconstruction are frequently the same, reflecting the unity of reconstruction methodology. What is different are the data employed by each. The two methods are typically combined in the process of reconstruction, with internal reconstruction being used before or after the comparative method, and in a variety of combinations. Together these two methods constitute the most powerful tools available for the reconstruction of ancestral systems.

1.5. Genetic classifications and family trees

In the family tree representation of the IE language family (figure 1.5), languages are grouped under branching nodes according to the degree of similarity which they exhibit. The protolanguages from which they stem are reconstructed, hypothetical entities based on the compilation, analysis, and evaluation of shared features and differences among attested daughter languages. A Common or Proto-Germanic is quite easily defended, since Germanic shows a clearly defined set of features which distinguish it from the other IE subgroups. Similar sets of commonalities are found to one degree or another in Italic, Slavic, Baltic, and Indo-Iranian, all groups with complex internal subdivisions, and all of which make the family tree an attractive mode of representation for a group of genetically related languages. However, the family-tree illustration can sometimes lead to misleading representations of language relationships and change. Branching schemes do not properly show geographical contiguity, dialect mixing, bilingualism, or interaction with speakers of non-related languages. They represent language development in discrete divisions with abrupt splits, not allowing for dual or multiple parentage (creoles). Furthermore, they typically do not represent time-depth, such as the more-than-3,000-year difference in the attestation of documents from Hittite and Lithuanian. In short, what the family tree model does is to display metaphorically the results of linguistic change, not change itself, as though languages are born of one parent and move through history in cultural and linguistic isolation. Thus, it cannot be held to be a universally adequate model of linguistic change.[11] Despite these deficiencies, how-

11 The standard alternative to the family tree model is the "wave theory", which allows for a spatially more realistic representation of genetic relationships on the basis of shared features among different language groups. See Anttila (1989: 300–309) for theoretical discussion, illustration, and bibliography.

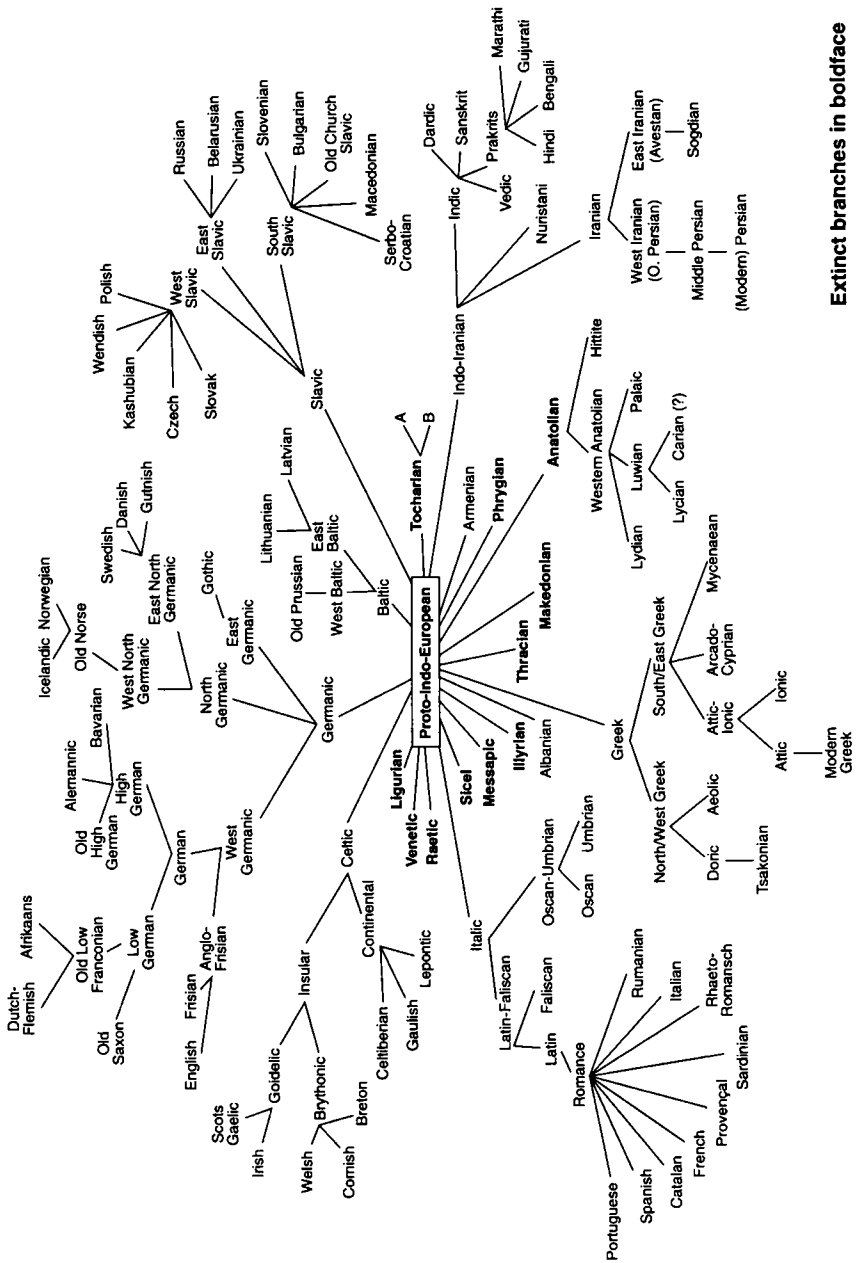

Figure 1.5. A family tree of the Indo-European languages

ever, the family tree model still has great value in that it provides a useful way of representing linguistic relationships, though not language change itself. It remains the most graphically consistent means of depicting bundles of commonalities among related languages. The family tree is not a model of linguistic change; it is a model of its results.

1.6. The Indo-European language family

The term Indo-European (Fr. "indo-européen", Germ. "indogermanisch") refers to a family of languages which by about 1000 B.C.E. were spoken over a large part of Europe and parts of southwestern and southern Asia. "Indogermanique" was first used by Conrad Malte-Brun, a Danish geographer living in Paris, in 1810. The English counterpart "Indo-European" was apparently coined by the Englishman Thomas Young in 1813. Though the concept of Indo-European is essentially linguistic, the term is originally a geographical one. It refers to the easternmost (India) and westernmost (Iceland) known expansion of the family (excluding post-colonial developments) at the time it was proven to be a historical entity by scholars of the eighteenth and nineteenth centuries. Post-colonial growth which has taken IE languages to the Americas, Africa, Hawaii, Australia, Hong Kong, and elsewhere around the world might now suggest another name for the family, but the term "Indo-European" is well rooted in the scholarly tradition.

The IE languages are classified into eleven major groups (ten if Baltic and Slavic are considered together as Balto-Slavic). Some of these groups have many members, while some others have only one. Of the eleven major groups, nine have modern spoken representatives while two, Anatolian and Tocharian, have no modern descendants. Additionally, there is a significant number of extinct "minor" branches of the IE family, most of which are fragmentarily known. A summary review follows, in which a sample of the languages of each group is mentioned. For further information, see Lockwood (1972), Baldi (1983), and especially the papers in Giacalone Ramat and Ramat (1998). For written specimens of a number of IE languages, see Beekes (1995: 318–347).

1.6.1. Indo-Iranian

After the Aryan migrations into India in the second millennium B.C.E., the earlier Indo-Iranian linguistic community was split into two main subdivisions, Indic (Indo-Aryan) and Iranian. The similarities between the two subdivisions are so consistent that there is no question about the status of Indo-Iranian as an intermediate unity between PIE and the Indic and Iranian subgroups. According to Payne (1987: 519), Indic and Iranian share fifty-seven isoglosses which link the two, a convinc-

ingly large number of shared characteristics to justify the Indo-Iranian unity. Indo-Iranian dominates the IE family numerically: of the 144 IE languages listed in Ruhlen (1987: 325), 93 are Indo-Iranian.

1.6.1.1. Indic (or Indo-Aryan)

The Indic languages are classified into three historical periods. They are Old Indic (1500–600 B.C.E.), Middle Indic (600 B.C.E. – 1000 C.E.), and New or Modern Indic (1000 C.E.–); see Masica (1991). Within the Old Indic group the most ancient language is Vedic. The oldest Vedic material is the *Ṛgveda* (*ṛg-* 'stanza'), a set of ten books of religious hymns which is usually dated to 1200–1000 B.C.E., though not in written form until much later. Subsequent Vedic texts include the *Atharvaveda* and the *Brāhmaṇas*. The later language is called (classical) Sanskrit (*saṁskṛta* 'adorned, elaborated, refined'), a literary language codified in the grammar of Pāṇini known as the *Aṣṭādhyāyī* written ca. 500 B.C.E. Its two most prominent literary texts are the epics called the *Mahābhārata* and the *Rāmāyaṇa*. Sanskrit is distinguished from the vernaculars commonly known as Prakrits (*prākṛta* 'unrefined'), which evolved into the later Indic dialects.

The oldest of the Middle Indic Prakrits is Pāli, with Buddhist religious texts datable to the sixth to fifth century, and the inscriptions of Aśoka (ca. 250 B.C.E.). There are also some Jainist religious writings in Jaina-Prakrit, and lyric and epic poetry in Mahārāṣṭrī from the Maharashtra region of Western India, all from roughly the same period.

Modern Indic is one of the most diffuse subgroups of the IE family, with scores of languages spread over the Indian subcontinent. Among the best known are Hindi, Urdu, Marathi, Panjabi, Gujarati, and Bengali. Also contained within the Modern Indic subgroup are the so-called "Dardic" languages (e.g. Shina, Khowar) spoken in the Upper Indus region in northern Pakistan and northwestern India. The exact relation of the Dardic languages to the other Modern Indic languages is unclear. Beekes (1995: 41–42) reports the existence of an IE language known as Bangani from the north of India in the foothills of the Himalayas which was first thought to be Indo-European, but not Indo-Iranian. Subsequent investigation has however established that this language is not only Indo-Iranian, but is specifically from the Indic subgroup. For general information on Indic languages, see Cardona (1987) and Lazzeroni (1998).

1.6.1.2. Iranian

The Iranians spread out perhaps as early as 1500 B.C.E. over a wide area, reaching not only the Iranian plateau, but also parts of China (Xinjiang) and Southern Russia. There are two principal representatives of Ancient Iranian, namely Eastern Iranian

and Western Iranian. Eastern Iranian is represented by Avestan, the language of the Avesta, a collection of sacred Zoroastrian texts. Avestan is comparable to Vedic in structure and has two recognizable dialects. The first of these is Old Avestan (Gathic), which records the writings of the prophet Zarathuštra. The Old Avestan materials are at least as old as 600 B.C.E. and are possibly much older (the first manuscripts, however, are from the thirteenth and fourteenth centuries C.E.). Somewhat later is the dialect known as Late or Younger Avestan.

Western Iranian is represented primarily by Old Persian, which is dated as far back as 500 B.C.E., mostly in cuneiform inscriptions of the Achaemenid emperors, in particular Darius the Great (521–486 B.C.E.) and Xerxes (486–465 B.C.E.).

While Avestan and Old Persian are the main linguistic sources of Ancient Iranian, there are some proper names and toponyms which provide information on Median, the language of the Median Empire in the province of Media in northwest Iran (eighth to sixth centuries B.C.E.), and on the languages of the Scythian and Sarmatian tribes of the south Russian steppes. Modern Ossetic is the only extant representative of these languages.

Western Middle Iranian (ca. 200 B.C.E. – 900 C.E.) is represented by Middle Persian and Parthian. Middle Persian (also known as Pahlavi) is the direct descendant of Old Persian. The earliest Middle Persian monuments, inscriptions on coins, date from the second century B.C.E., but the bulk of the corpus is from the Sassanid Empire (third–seventh centuries C.E.; see Back 1978). The Parthian corpus dates from the Arsacid dynasty (third century C.E.). The Eastern Middle Iranian languages are Sogdian (eighth century C.E. texts from Central Asia), Khotanese (texts from the fifth to tenth centuries C.E. from Chinese Turkestan), Khorasmian, and Bactrian, an Eastern Iranian language written in a Greek script.

Some modern descendants of the Iranian subgroup are Modern Persian, with a great deal of Arabic vocabulary; Pashto; Baluchi; Ossetic; Yaghnobi (the modern descendant of Sogdian); Judeo-Persian, with beginnings in the Middle Persian period; and Kurdish. For general information on Iranian languages, see Payne (1987), Schmitt (1989), and N. Sims-Williams (1998).

1.6.1.3. Nuristani

Also included in the Indo-Iranian subgroup is Nūristānī (formerly called Kafir), a group of dialects from eastern Afghanistan, represented by Ashkun and Kati. Nūristānī may be a subgroup of Iranian rather than a third branch of Indo-Iranian.

1.6.2. Greek (Hellenic)

One of the best known and documented of the ancient IE languages is Greek or Hellenic, which has been continuously spoken in the southern Balkan peninsula since

approximately 2000 B.C.E. The oldest literary monuments in Greek are the *Iliad* and *Odyssey*, traditionally ascribed to Homer and thought by some to have been written down as early as 800 B.C.E. Even older than these are the tablets found at Knossos on the island of Crete and at Mycenae, Pylos, Thebes, Tiryns, and Khania. The inscriptions on these tablets have been shown to be written in an archaic form of Greek known as Mycenaean, the oldest of which have been dated to the fourteenth and thirteenth centuries B.C.E., making it the second oldest representative of the IE languages, after Hittite.

Although Greek does not evidence the same degree of internal dialect divisions as, say, the Italic or Germanic languages, there are already considerable differences from the earliest documentation. Greek has two major dialect divisions, South/East Greek and North/West Greek. In the South/East dialect are Attic-Ionic, Arcado-Cyprian, and Mycenaean. In the North/West dialect are Aeolic and Doric. A major phonological isogloss separating these two divisions is their treatment of PIE *-*ti*, which in North/West Greek is preserved as -τι (cf. δίδωτι 'X gives') but in South/East Greek is changed to -σι (cf. δίδωσι), a distinction also evident in the 3rd pl. verbal endings -οντι/-ουσι, as seen in Dor. δίδοντι 'they give', Ion. διδοῦσι.

1.6.2.1. South/East Greek

Within South/East Greek the principal dialect is Attic, the literary language of fifth century B.C.E. Athens which represents the classical standard of Greek literature. This is the dialect of the great tragedians such as Aeschylus and Sophocles, comic writers such as Aristophanes, historians such as Thucydides, and philosophers such as Plato and Aristotle. The oldest Attic is an eighth century B.C.E. inscription, and the continuous literature in this dialect begins in the sixth century. Ionic is the dialect of the oldest continuous compositions in Greek, Homer's *Iliad* and *Odyssey*. There are three major dialect divisions within Ionic: West Ionic, represented by early inscriptions from Oropos and Euboea, some reaching into the eighth century; Central Ionic, found in the Cyclades and attested in seventh century inscriptions; and East Ionic, which is found in early inscriptions from Asia Minor.

In Arcado-Cyprian there are three principal representatives: Arcadian, from Arcadia in the Peloponnesus, attested in texts as old as the sixth century; Cypriote, from Cyprus, with one inscription from the eleventh century B.C.E. and others from the eighth century; and Pamphylian, which is generally classified as Arcado-Cyprian despite late attestation and a scant corpus from the fourth to first century B.C.E., which makes secure classification difficult. See Brixhe (1976).

Mycenaean is the name of the dialect ascribed to texts written by scribes on clay tablets in Mycenaean palaces at Knossos, Thebes, Pylos, Mycenae, Tiryns, and Khania which were ravaged by fire, baking the tablets and preserving them for posterity. They are written in a syllabic script called Linear B. Linear B appears to be descended from Linear A, one of two syllabic scripts (the other being Cretan Hiero-

glyphic) invented by the Minoans in the third millennium B.C.E. Linear B consists of 87 syllabic signs, about 40 of which are used as logograms, about 150 logograms (only 50 of which are used frequently), 11 symbols for weights and measures, and five numeral symbols. The oldest of the Linear B materials are those from Knossos, dated to 1375 B.C.E.

Most of the Linear B tablets are from Knossos and Pylos. They are primarily bookkeeping and other bureaucratic records dealing with the administration of the Mycenaean palaces. The Linear B script was deciphered in 1952 by the architect and amateur philologist Michael Ventris and the philologist John Chadwick. See Chadwick (1990).

1.6.2.2. North/West Greek

The Aeolic subgroup of North/West Greek (also known as North Achaean) has two major divisions. The first of these is Boeotian, which is attested in a large number of inscriptions and texts from as early as the seventh century B.C.E. from around the region of Boeotia. The second major division of Aeolic is North Aeolic, which is also known as Thessalo-Lesbian. North Aeolic comprises Thessalian, with numerous late texts from Thessaly stemming from the fourth century B.C.E., and Lesbian, from the island of Lesbos, in which the oldest usable inscriptions date from the sixth century B.C.E., though most others are later.

The Doric subgroup of North/West Greek evidences a number of geographically separated dialects. Miller (ms.) postulates a Proto-Doric with four subdivisions: Western Doric, Northern Doric, Elean, and Southeastern Doric.

Western Doric is represented mainly by Laconian, from parts of the Peloponnesus and Sicily, with texts from the late seventh century B.C.E.; and by Messenian, from Messenia, with inscriptions no earlier than the fourth century B.C.E.

Northern Doric splits into a number of historically attested dialects, including Aetolian, West Locrian, Phocian, East Locrian, Corinthian, Megarian, and East Argolic. The oldest material from this group comes in the form of inscriptions from the sixth century B.C.E.

Elean, from the northwest corner of the Peloponnesus, shows fragmentary textual material from the seventh and sixth centuries B.C.E., mostly from Olympia.

Southeastern Doric is divided into West Argolic, Cretan, and the East Aegean dialects of Cyrenaean, Theran, Coan, and Rhodian. West Argolic shows seventh century B.C.E. inscriptions from Argos, Mycenae, and Heraeum. Cretan, from the island of Crete, evidences a trove of materials from as early as the seventh century B.C.E., mostly from Knossos, Dreros, and Gortyn. Of the East Aegean dialects, Theran provides name lists and some graffiti from the eighth to seventh centuries B.C.E., and Cyrenian has inscriptions which can be dated to the seventh century. Rhodian inscriptions are generally late (fourth century), though there is one graffito which may be as old as the eighth century. In Coan, from Cos, most usable inscriptional

material is not earlier than the fifth century.

The extensive and continuous early documentation of Greek, as well as its complex structure, not only in texts but in thousands of inscriptions and loanwords into other languages, make it an invaluable resource in the understanding of PIE and the IE family of languages.

Standard Modern Greek and the other modern dialects arose from the Hellenistic Koine (ca. 300 B.C.E. to 300 C.E.), which is considered to be an Atticized Ionic at its core, with an admixture from other dialects. Thus Standard Modern Greek fits under the Attic-Ionic branch of South/East Greek. One exception to the Attic-Ionic lineage among the varieties of Modern Greek is the Tsakonian dialect, spoken in the eastern part of the Peloponnesus, which derives from the ancient Doric dialect. For details on Greek and its dialects, see Buck (1955), Palmer (1980), Joseph (1987), Hoenigswald (1998), and Miller (ms.).

1.6.3. Italic

The Italic group is constituted of IE languages of ancient Italy which are genetically linked by a large number of shared characteristics inherited from Proto-Indo-European. Around the turn of the first millennium B.C.E., ancient Italy was occupied by many Indo-European- and non-Indo-European-speaking peoples. Among the non-IE speakers, the most important are the Etruscans, who occupied central and later southern Italy; another group of non-IE speakers are the Iapygians in the southeast. Certain IE speakers include Veneti in the northeast, Lepontians and Ligurians in the northwest, Latinians, Umbrians, and South Picenes in the center, and Greeks, Messapians, Oscans, and Sicels in the south. There is some uncertainty over the linguistic status of the Raetians in the north and the North Picenes in the center, though it is most likely that they were IE speakers. Of the IE languages of ancient Italy, the principal ones are Venetic, Latin, Umbrian, and Oscan.

There are two main subdivisions within the Italic group. The first, Latin-Faliscan, is represented chiefly by Latin, by far the most important member. Apart from some early inscriptions (the oldest, short inscriptions incised on wine containers, have been dated to 620–600 B.C.E.; see Wallace 1989: 124), the first continuous Latin texts of any length are from the third century B.C.E. Faliscan, which is the language most closely related to Latin in the Italic group, is attested in about one hundred fifty inscriptional fragments and is imperfectly known. The oldest inscriptions in Faliscan are from the early seventh and sixth centuries B.C.E.

The second Italic subdivision is the Oscan–Umbrian (also known as Sabellic or Sabellian), whose chief members are Oscan and Umbrian. Oscan, the linguistically more conservative of the two, is attested in some coin legends from as early as the fifth century B.C.E., though most of the literature and other material is much later. The total Oscan corpus is about 200 inscriptions, the longest being the Tabula Ban-

tina (ca. 80 B.C.E.) from Lucania. Umbrian is contained largely in a single source, the Iguvine Tables from ancient Iguvium in Umbria, which have been dated to the period of approximately 300 or 200–90 B.C.E. and altogether contain about 5,000 words. There are also a few short inscriptions from north central Italy, near Bologna. Minor languages such as South Picene, Paelignian, Marrucinian, Vestinian, Marsian, Aequian, Sabine, and Volscian all share certain affinities with Oscan or Umbrian.

The Italic group of IE languages survives in the modern Romance languages (French, Italian, Portuguese, Spanish, Catalan, Rumanian, Sardinian, Rhaeto-Romansch, and the extinct Dalmatian). These languages descend from many geographical and social varieties of spoken Latin and cannot be traced directly back to the classical language. For details of the Italic group, see Coleman (1987) and Silvestri (1998). For Latin in particular, see Vineis (1998).

1.6.4. Anatolian

Now extinct, the Anatolian group comprises Hittite, Luwian, Lycian, Palaic, Lydian, and Carian, the last four surviving only in fragments. The Anatolian languages were unknown to modern scholars until archaeological excavations during the first part of this century in Boğazköy, Turkey, uncovered the royal archives of Hattušaš, the capital city of the Hittites. These excavations yielded about 10,000 texts written on clay tablets (continued excavation has brought the number to about 25,000, though not all have been published). The texts, which date from approximately the seventeenth to the thirteenth centuries B.C.E., were composed in a cuneiform script known from the study of Akkadian. They were written primarily in Hittite, but also contained material in Akkadian, Palaic, Luwian, Sumerian, Hattic, Hurrian, and Mitanni, a language related to Hurrian with some Indic loans. Decipherment proceeded quickly, and as the material from Boğazköy became plentiful, an Assyriologist called Bedřich Hrozný asserted in 1917 that the Hittite in the texts represented an IE language, an assertion that could not be denied. The Hittite empire fell around 1200 B.C.E., though the Anatolian family survived in the form of Luwian and Lycian into the last centuries B.C.E. until they were overrun by Greek colonists (Mallory 1989: 25).

From a philological analysis, three periods can be identified in the Hittite materials: Archaic Hittite (ca. 1700–1500 B.C.E.), Middle Hittite (ca. 1500–1350 B.C.E.), and Neo- or Late Hittite (ca. 1350–1200 B.C.E.). It was demonstrated soon after decipherment that Hittite contained a large number of apparently archaic features not found in other IE languages. Among the archaisms found in Hittite which are now considered by many to be features of the protolanguage are laryngeal consonants (also in Palaic); the two-gender system of animate and inanimate as opposed to the three-gender system of masculine/feminine/neuter common in the "classical"