

Cambridge Textbooks in Linguistics

Indo-European Linguistics

An Introduction

James Clackson

CAMBRIDGE

CAMBRIDGE

www.cambridge.org/9780521653138

This page intentionally left blank

Indo-European Linguistics

The Indo-European language family comprises several hundred languages and dialects, including most of those spoken in Europe, and south, south-west and central Asia. Spoken by an estimated 3 billion people, it has the largest number of native speakers in the world today. This textbook provides an accessible introduction to the study of the Indo-European proto-language. It clearly sets out the methods for relating the languages to one another, presents an engaging discussion of the current debates and controversies concerning their classification, and offers sample problems and suggestions for how to solve them. Complete with a comprehensive glossary, almost 100 tables in which language data and examples are clearly laid out, suggestions for further reading, discussion points and a range of exercises, this text will be an essential toolkit for all those studying historical linguistics, language typology and the Indo-European proto-language for the first time.

JAMES CLACKSON is Senior Lecturer in the Faculty of Classics, University of Cambridge, and is Fellow and Director of Studies, Jesus College, University of Cambridge. His previous books include *The Linguistic Relationship between Armenian and Greek* (1994) and *Indo-European Word Formation* (co-edited with Birgit Anette Olson, 2004).

CAMBRIDGE TEXTBOOKS IN LINGUISTICS

General editors: P. AUSTIN, J. BRESNAN, B. COMRIE, S. CRAIN,
W. DRESSLER, C. EWEN, R. LASS, D. LIGHTFOOT, K. RICE,
I. ROBERTS, S. ROMAINE, N. V. SMITH

Indo-European Linguistics

An Introduction

In this series:

- J. ALLWOOD, L.-G. ANDERSON and Ö. DAHL *Logic in Linguistics*
D. B. FRY *The Physics of Speech*
R. A. HUDSON *Sociolinguistics* Second edition
A. J. ELLIOT *Child Language*
P. H. MATTHEWS *Syntax*
A. REDFORD *Transformational Syntax*
L. BAUER *English Word-Formation*
S. C. LEVINSON *Pragmatics*
G. BROWN and G. YULE *Discourse Analysis*
R. HUDDLESTON *Introduction to the Grammar of English*
R. LASS *Phonology*
A. COMRIE *Tense*
W. KLEIN *Second Language Acquisition*
A. J. WOODS, P. FLETCHER and A. HUGHES *Statistics in Language Studies*
D. A. CRUSE *Lexical Semantics*
A. RADFORD *Transformational Grammar*
M. GARMAN *Psycholinguistics*
G. G. CORBETT *Gender*
H. J. GIEGERICH *English Phonology*
R. CANN *Formal Semantics*
J. LAVER *Principles of Phonetics*
F. R. PALMER *Grammatical Roles and Relations*
M. A. JONES *Foundations of French Syntax*
A. RADFORD *Syntactic Theory and the Structure of English: A Minimalist Approach*
R. D. VAN VALIN, JR, and R. J. LAPOLLA *Syntax: Structure, Meaning and Function*
A. DURANTI *Linguistic Anthropology*
A. CRUTTENDEN *Intonation* Second edition
J. K. CHAMBERS and P. TRUDGILL *Dialectology* Second edition
C. LYONS *Definiteness*
R. KAGER *Optimality Theory*
J. A. HOLM *An Introduction to Pidgins and Creoles*
G. G. CORBETT *Number*
C. J. EWEN and H. VAN DER HULST *The Phonological Structure of Words*
F. R. PALMER *Mood and Modality* Second edition
B. J. BLAKE *Case* Second edition
E. GUSSMAN *Phonology: Analysis and Theory*
M. YIP *Tone*
W. CROFT *Typology and Universals* Second edition
F. COULMAS *Writing Systems: An Introduction to their Linguistic Analysis*
P. J. HOPPER and E. C. TRAUGOTT *Grammaticalization* Second edition
L. WHITE *Second Language Acquisition and Universal Grammar*
I. PLAG *Word-Formation in English*
W. CROFT and A. CRUSE *Cognitive Linguistics*
A. SIEWIERSKA *Person*
D. RADFORD *Minimalist Syntax: Exploring the Structure of English*
D. BÜRING *Binding Theory*
N. HORNSTRIN, J. NUÑES and K. GROHMANN *Understanding Minimalism*
B. C. LUST *Child Language: Acquisition and Growth*
M. BUTT *Theories of Case*
G. G. CORBETT *Agreement*
J. C. L. INGRAM *Neurolinguistics: An Introduction to Spoken Language Processing and its Disorders*
J. CLACKSON *Indo-European Linguistics: An Introduction*

Indo-European Linguistics

An Introduction

JAMES CLACKSON

University of Cambridge



CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press

The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521653138

© James Clackson 2007

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

First published in print format 2007

ISBN-13 978-0-511-36368-9 eBook (Adobe Reader)

ISBN-10 0-511-36368-0 eBook (Adobe Reader)

ISBN-13 978-0-521-65313-8 hardback

ISBN-10 0-521-65313-4 hardback

ISBN-13 978-0-521-65367-1 paperback

ISBN-10 0-521-65367-3 paperback

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

Contents

<i>List of figures</i>	<i>page</i> viii
<i>List of tables</i>	ix
<i>Preface</i>	xii
<i>Transliteration conventions</i>	xiv
1 The Indo-European language family	1
2 Phonology	27
3 Morphophonology	64
4 Nominal morphology	90
5 Verbal morphology	114
6 Syntax	157
7 Lexicon and lexical semantics	187
<i>Glossary</i>	216
<i>References</i>	230
<i>Word index</i>	246
<i>Language index</i>	257
<i>Person index</i>	258
<i>Subject index</i>	259

Figures

1.1	A language family tree	<i>page 6</i>
1.2	Schleicher's IE family tree	10
1.3	The New Zealand family tree	11
1.4	The Pennsylvania family tree	12
1.5	Meid's Space-Time model	17
1.6	The New Zealand family tree with dates	19
7.1	A basic taxonomy for animals in PIE	207
7.2	Taxonomy of PIE terms following Benveniste	208
7.3	Gamkrelidze and Ivanov's taxonomy of animals in PIE	208
7.4	Animal taxonomy in modern European languages	209

Tables

1.1	IE languages by date and place of first attestation	page 8
1.2	Instrumental plural markers in various IE languages	10
1.3	Nostratic and PIE stop comparisons	22
2.1	A correspondence set for English <i>t</i> and German <i>ss</i>	29
2.2	Extended correspondence set for medial <i>*t</i> in Germanic	30
	Table for exercise 2.1	31
2.3	Six sound-laws and a rule of IE	32
2.4	Phonological inventory of PIE	34
2.5a–c	Comparative IE phonology	37-39
2.6	Frequency of reconstructed phonemes in PIE roots in <i>LIV</i>	41
	Table for exercise 2.2	42
2.7	Reconstructed four-way stop system of PIE	43
2.8	Glottalic and traditional PIE reconstructed stop system	45
	Table for exercise 2.5	49
2.9	The PIE dorsal series	50
	Table for exercise 2.8	53
2.10	‘Irregular’ ablaut series	55
2.11	Sanskrit nasal infix verbs	56
2.12	Laryngeal developments in some early IE languages	57
2.13	The triple reflex of laryngeals in Greek	59
	Table for exercise 2.9	61
3.1	‘Foot’ in various IE languages	72
3.2	Ablaut grades of <i>*ped-</i>	72
3.3	The morphological structure of inflected words in PIE	73
3.4	Accent and ablaut in a PIE paradigm	75
3.5	Putative correlations of accent and ablaut	78
3.6	Strong and weak forms of nouns and verbs	79
3.7	Accent / ablaut paradigms	80
	Tables for exercise 3.3	82
3.8	Generating accent paradigms for inherent accents	85
4.1	The singular paradigm for ‘father’ in PIE	93
4.2	The singular paradigm for ‘sky / god’ in PIE	94
4.3	The paradigm for ‘cloud’ in PIE	94
4.4	The paradigm for ‘liver’ in PIE	95
	Table for exercise 4.1	95

Table for exercise 4.2	95
Table for exercise 4.3	96
4.5 PIE feminine stems in <i>*-eh₂</i>	97
4.6 PIE feminine stems in <i>*-ih₂</i>	97
4.7 The paradigm for 'wolf' in PIE	98
4.8 Possible ablaut grades for PIE 'sleep, dream'	98
4.9 The paradigm for 'yoke' in PIE	99
Table 1 for exercise 4.4	99
Table 2 for exercise 4.4	100
Table for exercise 4.5	100
4.10 Nouns with two plurals	102
4.11 Number in PIE nouns	103
4.12 Feminine abstract nouns derived from verbal stems	106
4.13 The paradigm for 'woman' in PIE	108
4.14 i-motion in Luwian participles	110
Table for exercise 4.6	111
5.1 Greek verbal stems	116
5.2 Vedic Sanskrit verbal stems	117
5.3 Latin verbal stems	118
5.4 Gothic verbal stems	118
5.5 Comparison of stem formations in Greek and Vedic Sanskrit	119
5.6 The Greco-Aryan model of the PIE verb	120
5.7 Mergers of the inherited preterite formations	122
5.8 Active personal endings in Sanskrit	123
5.9 Athematic active primary endings: PIE 'to be'	124
5.10 Athematic active secondary endings: PIE 'to be'	125
Table for exercise 5.1	125
Table for exercise 5.2	125
Table for exercise 5.3	126
5.11 Reconstructed thematic primary and secondary endings	127
Tables for exercise 5.4	127
5.12 The reconstructed perfect: PIE 'know'	128
5.13 Exact matches between Anatolian and Sanskrit verbal forms	130
5.14 The injunctive compared with indicative tenses	131
5.15 PIE eventive endings	132
5.16 Eventive formations in early PIE	135
5.17 The athematic optative of PIE 'to be'	136
5.18 An 'improved' Greco-Aryan model of the PIE verb	138
5.19 An alternative 'improved' Greco-Aryan model of the PIE verb	138
5.20 The Hittite <i>-hi</i> and <i>-mi</i> active conjugations	139

5.21	Lexical assignments of roots to active or middle paradigms	143
5.22	Middle endings in IE languages	144
	Table for exercise 5.5	144
5.23	Reconstruction from archaic middle endings	145
5.24	Vedic Sanskrit third singular middle forms without <i>t</i>	146
5.25	Vedic Sanskrit third plural middle forms with <i>r</i>	147
5.26	Reconstruction of PIE middle endings	147
5.27	Comparison of PIE middle endings, perfect endings and the <i>-hi</i> conjugation	148
5.28	Present and aorist formations from three roots	151
	Table for exercise 5.7	152
5.29	Aorist formations from nasal infix presents	154
6.1	Comparison of negative particles in prohibitions and statements	163
6.2	Position of enclitics in early IE languages	171
6.3	Summary of relative clause types in early IE languages	175
6.4	Nominative and ergative alignment systems	177
6.5	Case-marking in active systems	178
6.6	Active and inactive verbs in PIE	179
7.1a	Derivatives of the root <i>*k'lew-</i> 'hear'	192
7.1b	Derivatives of the root <i>*h₂ek'-</i> 'sharp'	193
7.2	Selected derivatives of the root <i>*med-</i>	194
7.3	IE numerals	198
7.4	PIE formation of decads	199
7.5a	IE kinship terms: relations by blood	202
7.5b	IE kinship terms: relations by marriage	204
7.6	Some PIE animal names	206
7.7	'Life' in PIE and Indo-Iranian	211

Preface

Do we need another introduction to Indo-European linguistics? Since 1995 four have been published in English (Beekes 1995, Szemerényi 1996, Meier-Brügger 2003, Fortson 2004) and the ground seems to be pretty well covered. This book, however, aims to be an introduction of a different sort. Whereas the works mentioned give up-to-date and (usually) reliable information on the current thinking on what is known in Indo-European studies, here the aim is to present rather areas where there currently is, or ought to be, debate and uncertainty. Whereas previous introductions have aimed for the status of handbooks, reliable guides to the terrain presented in detail, this one aspires more to the status of a toolkit, offering up sample problems and suggesting ways of solving them. The reader who wants to know the details of how labio-velar consonants developed in Indo-European languages or the basis for the reconstruction of the locative plural case ending will not find them here; instead they will be able to review in detail arguments about the categories of the Indo-European verb or the syntax of relative clauses. The result is that this book has shorter chapters on areas such as phonology, where there is now more general agreement in the field, and correspondingly longer sections on areas which are passed by more summarily in other introductions. Memory athletes may be disappointed by the reduction in data, but I hope that others will welcome the increase in argumentation.

This book contains a number of exercises and discussion questions within and at the end of each chapter, designed to help readers get to grips with some of the issues in Indo-European linguistics and prompt further discussion. Answers to some of the exercises, hints and tips for others can be found at www.classics.cam.ac.uk/clacksonindoeuropean.

This book originates in the suggestion made some years ago by Geoff Horrocks, Professor of Comparative Philology at the University of Cambridge, to the two young (then) lecturers in his department that they should write a book which concentrated on explaining some of the issues of Indo-European linguistics. Torsten Meißner and I embarked together on the project with enthusiasm, and the framework of the book, and an early draft of Chapter 3 and parts of Chapters 1 and 2 were originally written in collaboration. Other duties and commitments delayed the completion of the work, however, and after a lapse of a few years I took up the project again on my own, partly through freedom allowed me by the generous award of a Philip Leverhulme Prize. Throughout the whole writing process Torsten has been an invaluable colleague and friend, he has commented

on drafts of the entire work at various stages and I have learnt from him more than it is possible to express about all aspects of Indo-European. Without his input, this book would be half of what it is now. He is of course, like all the others mentioned here, exculpated of any responsibility for my own errors or misjudgements.

I have also benefited greatly from the help of many other colleagues and students. Parts of the work in draft were read by Andreas Bartholomä, Dr Michael Clarke, Dr Coulter George, Dr Antonia Ruppell and Dr Sheila Watts. I owe particular gratitude to Alex Mullen, who worked as research assistant and as an exacting copy-editor over one summer and saved me from innumerable errors and anacolutha. Professor Peter Matthews, one of the series editors for the CUP Textbooks in Linguistics, read and commented on the work in draft and Andrew Winnard, the CUP linguistics editor, was always ready to offer help and advice. Chris Jackson acted as an assiduous and attentive copy-editor for CUP and plugged many gaps in my knowledge. Sarah Clackson gave love and support in the initial stages of writing. Véronique Mottier enabled me to continue and complete the manuscript.

Finally, my first teacher in Indo-European linguistics was the late and much missed Bob Coleman, sometime Professor of Comparative Philology at the University of Cambridge. This book is dedicated to his memory.

Transliteration conventions

Words and texts cited in this book generally follow established conventions of transliteration or citation and are not given in IPA transcription. The following notes are intended to guide the reader to the pronunciation of forms cited in this book. Since in many cases the languages are no longer spoken, there is often uncertainty about the precise realisation of certain sounds, and the pronunciations given here can only at best be approximate. It should be noted that we have not attempted to give comprehensive accounts of the phonologies of the languages concerned, but merely to aid readers to understand how a particular sign is used. In general we have avoided giving details of signs which are not used in this book. Where no information is given on the pronunciation of a sign, the reader can assume that it has a value approximately equivalent to its IPA equivalent. In all cases we have tried to follow the standard orthography used in the scholarly literature, except in the case of Greek, for which we have not used the Greek alphabet, but a transliteration which should make it accessible to all and enable readers who know Greek to recognise the original words.

Albanian

Albanian is written in the Latin alphabet.

c ç q gj x xh represent affricates or palatals: *c* = [ts] *ç* = [tʃ] *q* = [c] *gj* = [ʝ] *x* = [dʒ] *xh* = [dʒ]

dh th sh zh represent fricatives: *dh* = [ð], *th* = [θ] *sh* = [ʃ] *zh* = [ʒ]

ë is the central unrounded mid vowel [ə]

Armenian

Armenian is written in its own alphabet. The transliteration here follows that used in most modern scholarly accounts of the language, for example Schmitt (1981).

c j č ĵ represent affricates: *c* = [ts] *j* = [dz] *č* = [tʃ] *ĵ* = [dʒ]

š ž represent postalveolar sibilants: š = [ʃ] ž = [ʒ]

p' t' c' č' k' represent aspirates: p' = [p^h] etc.

ł represents a velarised lateral [ɫ]

ř represents a trilled [r], whereas r represents the approximant [ɹ]

y represents the palatal approximant [j]

ê represents a close-mid unrounded front vowel, e an open-mid unrounded front vowel.

Avestan

Forms cited are transliterated from the Avestan alphabet, following the practice of Hoffmann and Forssman (1996: 41).

Vowels written with a macron, such as ā, are conventionally pronounced long.

ā represents a nasalised vowel.

â represents a vowel in between [a] and [o], probably the unrounded low back vowel [ɑ].

c j represent affricates: c = [tʃ] j = [dʒ]

š ž represent post-alveolar sibilants: š = [ʃ] ž = [ʒ]

ŋ represents an unreleased nasal.

y represents the palatal approximant [j] in word initial position; the same sound is generally represented by *ii* within a word.

v represents the labio-velar approximant [w] in word-initial position; the same sound is generally represented by *uu* within a word.

Palatalised consonants are denoted with a superscript acute accent, for example *ń* represents a palatalised *n*.

Labialised consonants are written with following superscript *v*, for example *x^v* represents a labialised velar fricative.

Etruscan

Etruscan is written in an alphabet adapted from the Greek alphabet. The transliteration follows standard scholarly practice.

c represents an unvoiced velar plosive.

z probably represents an affricate [ts]

θ represents an unvoiced dental consonant, distinguished in some way from the unvoiced plosive *t*.

χ represents an unvoiced velar consonant, distinguished in some way from the unvoiced plosive *c*.

Gothic

The Gothic alphabet is an adaptation of the Greek alphabet, with reuse of some letters to correspond to sounds present in Gothic but not in Greek. The transcription here follows the standard scholarly conventions, as given, for example, in Rauch (2003: 6).

There is dispute about what sounds the digraphs *ai au* represent. Etymologically, and in transcriptions of foreign words into Gothic, these digraphs correspond both to short and long vowels and diphthongs; *ai* thus appears to represent all of [ɛ], [ɛ:] and [ai] and *au* appears to represent all of [ɔ], [ɔ:] and [au]. We have not used here the convention of using the notation *aí* for [ɛ] and *ái* for [ai], since this corresponds to no difference in the actual written texts.

Long vowels are not marked separately to short vowels except in the case of a long *i*, which is written with a digraph *ei*. The vowels *e* and *o* are only used to represent the long close-mid front and back vowels [ɛ:] and [ɔ:], for which there are no short counterparts.

g written before another velar consonant represents the velar nasal.

q represents a voiceless glottal stop with simultaneous lip-rounding [k^w]

hw represents the glottal fricative with simultaneous lip-rounding [h^w]

þ represents the voiceless dental fricative [θ]

Greek

Mycenaean Greek is originally written in a syllabic script. In the transcription, syllabic signs are identified through writing hyphens between them (= is used to indicate a syllabic boundary which is also a clitic boundary). The syllabic script does not represent voiced stops other than [d], or aspirated stops, and *r* represents both [r] and [l]. The syllabary only has signs for open syllables, and frequently sounds which occur in the coda of syllables are omitted in the script.

q represents a voiceless stop, a voiced stop and an aspirated stop which have various outcomes in later Greek, and are usually understood to be [k^w], [g^w] and [k^{wh}].

Alphabetic Greek is written in the Greek alphabet, which has many different local variants. Forms cited are generally taken from the Attic dialect. The transliteration used here transliterates Greek letters by single letters, except in the cases of the so-called double consonants, where *zd* represents Greek ζ, *ps* represents Greek ψ and *ks* represents Greek ξ.

ph th kh represent aspirated consonants, so that *ph* = [p^h] etc.

Vowels written with a macron, such as *ā*, are long.

ei and *ou* in Attic Greek represent front and back long close-mid vowels, but in other dialects and earlier Greek these are front and back rising diphthongs. *u* and *ū* in Attic Greek represent close front rounded vowels, but in other dialects and in earlier Greek these are close back rounded vowels.

Three accent marks are used. The acute is reckoned to indicate a rising pitch on the vowel, the circumflex a rise and fall in pitch on a long vowel or diphthong, and the grave is a modification of the acute accent when it stands before another accented word.

Hittite

Hittite is written in a form of the cuneiform syllabic script employed also for the Semitic language Akkadian. As well as using signs to represent syllables, the script also employs various conventional ideograms and classificatory signs, and sometimes scribes use Akkadian words in place of Hittite ones. We have followed the conventional means of transcribing these, which sometimes gives the text a confusing appearance, with capital and superscript letters alongside lower-case. For our purposes it may suffice to state here that only the text in lower-case reproduces Hittite words and endings. The reader who wishes to know more is advised to consult Friedrich (1960: 21–5).

In our transcription of Hittite we have followed current scholarly practice in using a broad transcription which reproduces the likely shape of the Hittite word. We have avoided using diacritics in the transcription as far as possible (thus we write *s* and *h*, not *š* and *ḥ*, in line with current practice).

ku before a following vowel probably indicates a labialised velar plosive [k^w]
z probably represents an affricate [ts]

Vowels written with a macron, such as *ā*, represent the combination of two syllabic signs, and are usually reckoned to have been long vowels.

= is used to indicate a clitic boundary.

Latin

i represents both the vowel [i] and the consonant [j].

u represents both the vowel [u] and the consonant [w]. By convention when in upper case this sign is written *V*.

qu represents a combination of [k] and [w].

Vowels written with a macron, such as *ā*, are pronounced long.

Lithuanian

Lithuanian is written in the Latin alphabet, with some extra characters.

č represents an affricate, [tʃ]

š ž represent post-alveolar sibilants: *š* = [ʃ] *ž* = [ʒ]

ė represents a long close-mid front vowel; *e* represents an open-mid front vowel.

y represents a long unrounded high front vowel.

a ę į ū represent vowels which were originally nasalised, but which have now lost their nasalisation and are pronounced long.

Accented short vowels are marked with a grave accent. On accented long vowels or diphthongs (which include combinations of vowel and *l, m, n, r*) two signs are used to represent different pitch contours: the acute accent signifies a falling pitch, the circumflex a rising pitch.

Luwian

Luwian is written either in the cuneiform syllabary employed for Hittite (see above) or in a hieroglyphic syllabic script. We here follow the transcription of the hieroglyphic script as employed in Hawkins (2000 and 2003).

Lycian

Lycian is written in its own alphabet, adapted from the Greek. The transcription of the Lycian alphabet here follows that used in Melchert (2004).

ā and *ē* are nasalised vowels.

m̄ and *n̄* may represent unreleased nasals.

q represents some sort of voiceless velar consonant.

x represents some sort of voiceless velar consonant; it is transcribed as *χ* in earlier works.

z can represent the affricate [ts]

Old Church Slavonic

Old Church Slavonic is written in the Cyrillic and Glagolitic alphabets. There are many competing systems of transliteration of the Cyrillic alphabet; the one we use here follows Comrie and Corbett (1993), except in the use of the signs *ǐ* and *ǔ*.

c č represent affricates: *c* = [ts] *č* = [tʃ]

š ž represent post-alveolar sibilants: š = [ʃ] ž = [ʒ]

ě represents an open-mid unrounded front vowel (*e* is a close-mid unrounded front vowel).

y represents a close unrounded back vowel.

ř represents a mid central unrounded vowel.

ů represents a mid central rounded vowel.

ě and ů are nasalised vowels.

Old English

Old English is written in the Latin alphabet (see above) with additional letters.

þ and *ð* are used to represent voiceless and voiced interdental fricatives [θ] and [ð].

æ represents an open unrounded front vowel, *a* an open unrounded back vowel.

y and *æ* represent rounded close and mid front vowels.

Old High German

Old High German is written in the Latin alphabet. Long vowels are denoted with a macron.

ch represents the voiceless velar fricative [x]

Old Irish

Old Irish is written in the Latin alphabet (see above) with a number of orthographic innovations.

Long vowels are indicated by an acute accent, for example *á*.

Palatalisation of syllable-final consonants is indicated by writing *i* before the consonant.

p t c represent voiceless stops word-initially; elsewhere they stand for voiced stops.

pp tt cc represent voiceless stops word-medially or word-finally.

b d g represent voiced stops word-initially; elsewhere they stand for voiced fricatives [β] [ð] [ɣ]

ph th ch represent voiceless fricatives, [f] [θ] [x] respectively.

= is used to indicate a clitic boundary.

Old Norse

The Old Norse cited in this book is taken from texts originally written in a form of the Latin alphabet, with added letters, diacritics and digraphs. Long vowels are denoted with the acute accent: for example, *á* and *é* are the lengthened counterparts to *a* and *e*.

þ represents the voiceless dental fricative [θ]

Old Persian

Old Persian is written in a syllabic script. The transcription used here follows Brandenstein and Mayrhofer (1964: 17–24). Vowels written with a macron, such as *ā*, represent the combination of two syllabic signs and are pronounced long.

y represents the palatal approximant [j]

= is used to indicate a clitic boundary.

Oscan

Oscan is written both in the Latin alphabet and a native alphabet. We have not followed the standard practice of differentiating between the two alphabets through the use of bold script, since all the forms in this work are originally written in the native script.

Any doubled vowel, such as *aa* and *ii*, represents a long vowel.

Palaic

The very small corpus of the Anatolian language Palaic is written in the same cuneiform script as Hittite.

Russian

Russian is written in the same Cyrillic script as is used for Old Church Slavonic (with the abandonment of a few signs). The transcription here used is the same as for Old Church Slavonic, except for the use of the soft sign '.

c *č* represent affricates: *c* = [ts] *č* = [tʃ]

š *ž* represent post-alveolar sibilants: *š* = [ʃ] *ž* = [ʒ]

y represents a close unrounded back vowel.

' written after a consonant denotes that the consonant is palatalised.

Sanskrit

Sanskrit forms are generally cited from the earliest texts, the Vedic hymns and associated texts, the language of which is sometimes called Vedic. The transliteration of the devanagari script adopted here is the one used in modern scholarly treatments of the language, (for example, Mayrhofer 1986–2001).

ph th ṭh ch kh represent aspirated consonants, so that *ph* = [p^h] etc.

bh dh ḍh jh gh represent consonants traditionally described as voiced and aspirated.

c ch j jh represent palatal stops, so that *c* = [c] *j* = [j]

ṭ ṭh ḍ ḍh ṣ ṣ represent retroflex consonants, so that *ḍ* = [ḍ] *ṣ* = [ṣ] etc.

ś represents the palatal fricative [ç]

h represents the voiced glottal fricative [ɦ]

ñ represents the palatal nasal [ɲ]

y represents the palatal approximant [j]

r represents a syllabic *r* [r̥]

Vowels written with a macron, such as *ā*, are pronounced long.

The acute accent indicates a rise in pitch on the syllable.

Serbian

Serbian is written in the Cyrillic alphabet, and the transcription used here is the same as that for Russian.

Tocharian

Tocharian uses a version of the same script as Sanskrit (see above). It also has an additional vowel sign transcribed *ä* which is taken to represent a mid central unrounded vowel [ə]. In Tocharian *ts* represents an affricate [ts].

Umbrian

Umbrian is written both in the Latin alphabet and a native alphabet. We have not followed the standard practice of differentiating between the two alphabets through the use of bold script, since all the forms cited in this work are originally written in the Latin alphabet, except for one, *utur*, written in the native script. In the word *utur*, *t* may represent a voiced dental stop, since the Umbrian alphabet has no separate sign for this sound. For the other forms, see the notes for the Latin alphabet given above.

Welsh

Welsh is written in the Latin alphabet, and the forms cited here are in the modern orthography.

ch th represent the unvoiced fricatives [x] and [θ] respectively.

dd represents the voiced dental fricative [ð]

f represents the voiced labio-dental fricative [v], *ff* its unvoiced counterpart [f].

u represents the close unrounded central vowel [ɨ].

w represents either the close rounded back vowel [u] or the consonant [w].

y represents an unrounded central vowel, either [ə] or [ɨ].

1 The Indo-European language family

1.1 Introduction

Indo-European (IE) is the best-studied language family in the world. For much of the past 200 years more scholars have worked on the comparative philology of IE than on all the other areas of linguistics put together. We know more about the history and relationships of the IE languages than about any other group of languages. For some branches of IE – Greek, Sanskrit and Indic, Latin and Romance, Germanic, Celtic – we are fortunate to have records extending over two or more millennia, and excellent scholarly resources such as grammars, dictionaries and text editions that surpass those available for nearly all non-IE languages. The reconstruction of Proto-Indo-European (PIE) and the historical developments of the IE languages have consequently provided the framework for much research on other language families and on historical linguistics in general. Some of the leading figures in modern linguistics, including Saussure, Bloomfield, Trubetzkoy and Jakobson, were Indo-Europeanists by training, as were many of those who taught in newly founded university departments of linguistics in the second half of the twentieth century. Despite this pedigree, IE studies are now marginalised within most university linguistics courses and departments. In most US and European institutions, Indo-Europeanists with university posts do not teach in linguistics departments but in classics, oriental studies, celtic studies or the like. Historical linguistics courses may include a section on PIE, or Saussure's work on laryngeals as an example of internal reconstruction, but few students will engage in any current work on IE in any depth.

The intention of this book is not to convert general linguists to IE studies, or to restore the discipline to the central position in linguistics that it had a hundred years ago. Rather it aims to set forth some of the areas of debate in IE studies. In recent years a number of grammars and handbooks of PIE have been published in English (Gamkrelidze and Ivanov 1984 (English translation 1995), Sihler 1995, Beekes 1995, Szemerényi 1996, Meier-Brügger 2000 (English translation 2003), Fortson 2004). Most of these works are excellent, but sometimes the apodeictic style of the presentation leaves the reader uncertain about whether what is presented is actually hypothesis or 'fact'. One explanation for a historical change may be preferred over another, but the author may not make clear what is at stake in the choice between the alternatives. This book takes a different approach. It is

deliberately not intended to be a grammar of IE, or a survey of the developments that have taken place between PIE and the daughter IE languages, but rather to be a survey of some current debates and topics of more general interest in the reconstruction of PIE, and a guide to the ways in which some of these issues have been addressed. The material throughout the book is selective and illustrative, and the reader who wants to find out more will be advised to follow the further reading sections at the end of each chapter.

1.2 The IE languages

The IE language family is extensive in time and space. The earliest attested IE language, Hittite, is attested nearly 4,000 years ago, written on clay tablets in cuneiform script in central Anatolia from the early second millennium BC. We have extensive textual remains, including native-speaker accounts of three more IE languages from 2,000 years ago: Ancient Greek, Latin and Sanskrit. Also from the beginning of the Christian Era we have much more limited corpora of many more IE languages. The stock of recorded IE languages further increases as we move forward in time. In 2003, over 2.5 billion people spoke an IE language as their first language, and there were at least seventy codified varieties, each spoken by a million or more native speakers. Four hundred years ago nearly all speakers of IE lived in Europe, Iran, Turkey, Western Asia and the Indian sub-continent, but migrations have now spread speakers to every part of the world. The wealth of historical material makes IE the best-documented language family in the world.

What is it that makes an IE language IE? What does it mean to be classed as an IE language? It is usual at the opening of books on IE to repeat the famous words of Sir William Jones in 1786 which are traditionally taken to have inaugurated the discipline. Jones remarked on the similarity of Sanskrit to Latin and Greek, stating that they all bore ‘a stronger affinity, both in the roots of verbs and in the forms of grammar, than could possibly have been produced by accident; so strong, indeed, that no philologer could examine them all three, without believing them to be sprung from some common source, which, perhaps, no longer exists’. Jones also noted that Gothic, Celtic and Persian could be added to the same family. Since 1786, a considerable methodology has been established to qualify and quantify Jones’ notion of ‘affinity’ between the grammars and lexicons of the IE languages, and to work out a hypothetical model of the ‘common source’, PIE. But there has been no advance on Jones’ criterion for relatedness between languages of the family: greater similarity in verbal roots and morphological paradigms than might be expected by chance. Languages which belong to the IE family do so either because the similarity between them and other IE languages is so strong as to be self-evident, or because they can be clearly related to languages which do obviously belong to the family. For a language which has textual remains sufficient

for the linguist to extract lexical and grammatical information, it is possible to apply the techniques of reconstruction, such as the comparative method, to build a picture of its development from PIE. However, the operation of the comparative method does not guarantee a language's place in the family; only the initial recognition that two or more languages are related can do that. (We shall return to examine the implications of this point more fully in section 1.6.)

When does a linguist decide that there is enough material to relate a language to the IE family? There is no absolute set of criteria beyond the general rule that the evidence must convince both the individual linguist and the majority of the scholarly community. A language which only survives in a very limited corpus may contain sufficient IE features to be generally agreed to be IE. As an example, take the case of Lusitanian. Lusitanian is known from a handful of inscriptions from the west of the Iberian peninsula, written in the Latin alphabet around the first century of the Christian Era. One of these inscriptions, from Lamas de Moledo in Portugal, reads as follows (the slash / signals the end of the line in the original inscription):

RVFINVS. ET
 TIRO SCRIP/SERVNT
 VEAMINICORI
 DOENTI
 ANGOM
 LAMATICOM
 CROVCEAIMAGA
 REAICOI. PETRANIOI. T
 ADOM. PORGOM IOVEAI
 CAELOBRIGOI

The first four words are Latin: 'Rufus and Tiro wrote (this).' But the remainder of the inscription is not Latin. The inscription is taken to refer to the sacrifice of animals by a people called the *Veaminicori* to gods who are also addressed with their cult titles. Not all the words are understood, although the structure is clear: *Veaminicori* is nominative plural, *doenti* is a verb meaning 'they give'. The rest of the inscription has nouns in the accusative singular, denoting what is given: *angom lamaticom*, *tadom porgom*; and the names of the recipients in the dative singular: *petranioi*, *caelobrigoi*. This is not much, but enough that no Indo-Europeanist doubts that Lusitanian is a member of the IE family. Several of the word-forms are very similar to Latin. For example, the dedicated item *porgom* is very likely to mean 'pig' (Latin accusative singular *porcum* 'pig'), and *angom* to mean 'lamb' (Latin accusative singular *agnum* 'lamb'). The verb-form *doenti* 'they give' contains the root *do-* 'give', familiar from the equivalent forms in Greek (*dō-*), Latin (*da-*) and Sanskrit (*dā-* / *d-*). More importantly, it shows a third person plural ending *-enti* which is also found in these languages (dialectal Greek *-enti*, Archaic Latin *-nti* and Sanskrit *-anti*). Furthermore, the ending *-oi* coincides with a dative singular marker elsewhere (Greek *-ōi*, Archaic Latin *-oi*

and Sanskrit *-ai*), and the nominative plural ending *-i* accords with the nominative plural *-i* of one Latin noun declension. The interpretation of this inscription rests entirely on the identification of its language as IE, but most scholars have found it hard to believe all these similarities are entirely due to chance.

Compare with Lusitanian the case of Tartessian, another language from Ancient Spain which is known only from short inscriptions. Tartessian is better attested than Lusitanian, and from a period 600–800 years earlier. Unfortunately, we are not confident about our reading of the Tartessian script, and we do not have the helpful marks which are usually present in the Lusitanian inscriptions indicating where words begin or end. We consequently do not know a lot about the morphology of the language. However, some scholars have identified in Tartessian repeated patterns of (what they take to be) verbal endings. Consider the following inscription, reproduced in its entirety:

botieanakertorobatebarebanarkenti

The final nine letters, *narkenti*, occur elsewhere in the inscriptional corpus, as do the similar forms *narken*, *narkenii*, *narke*, *narkenai*. Here again we see a final element *-nti* that could represent the third person plural of a verbal ending in an IE language, just as in Lusitanian above. However, there is no obvious connection in the older IE languages to what would appear to be the verbal ‘stem’ *nark-*. Moreover, if we try to use what we know of IE morphology and vocabulary to interpret the rest of the inscription, we do not get very far. In Lusitanian, the assumption that the language was IE yielded vocabulary and morphology. In Tartessian, we have nothing more than the ending *-enti*. We do not even know enough about the morphological structure of the language to be confident that *narkenti* should be analysed as stem *nark-* + affix *-enti*. Accordingly, the general consensus is that Tartessian should not be included among the IE family.

The status of languages as IE or not may change in the light of an increase in our knowledge of the family. This is the case with the languages Lydian and Lycian, spoken in Anatolia in the first millennium BC, and known from inscriptions written in modified forms of the Greek alphabet. Before the discovery and accurate description of older IE languages in the Anatolian family, Hittite and Luwian, written in cuneiform and hieroglyphic scripts hundreds of years earlier, Lydian and Lycian could not be securely included in the IE family. However, their affinity to the earlier Anatolian languages is now patent, and since these show clear morphological and vocabulary similarities with the rest of IE, there is no doubt that Lydian and Lycian belong in the family as well. If we did not have any Anatolian languages other than Lydian and Lycian, we would not now be so certain of their ancestry. Indeed, we would not be able to make much sense of them at all, since it is only through the knowledge of how Anatolian languages are structured that headway has been made with the interpretation of the surviving inscriptions. It is, consequently, conceivable that a language such as Tartessian could come into the IE fold, if we were to have some intermediate steps to show the link between the rest of the family and the inscriptional remains that we have.

1.3 The branches of the IE tree

It follows from the remarks about Lydian and Lycian that the sub-families of IE are vitally important in determining the membership of the family. Whereas the affinity of the oldest IE languages declares itself as stronger than could be produced by chance (to most of those who study them), the affinity of languages attested more recently is sometimes only discernible through first relating them to sub-families of IE. Thus, to take an example of two languages at the far ends of the historical IE speech area, Modern Irish and Sinhala would not strike a linguist who was fluent in each, but unacquainted with their history, as *necessarily* related. It is only through relating Modern Irish to Old Irish, and Sinhala to Sanskrit, that the connection between the two languages becomes clear.

The majority of IE languages currently spoken belong to six large sub-groups of IE. Modern Irish and Old Irish are members of the Celtic sub-group, which also includes Welsh, Scots Gaelic, Breton, Cornish and Manx. Sinhala is part of the large Indic family, comprising most of the languages currently spoken in North India and Pakistan, Sanskrit and the Middle Indian Prakrits. English is a member of the Germanic branch; this includes Dutch, German and the Scandinavian languages among living languages, as well as earlier stages of these languages, such as Old English, Old High German and Old Norse, and other extinct varieties such as Gothic, once spoken in south-east Europe and southern Russia. The other large sub-groups are Romance and Slavic in Europe, and Iranian in Asia. All of these sub-groups of IE were themselves recognised as linguistic families before Jones' identification of the larger IE family cited above. The traditional criterion for grouping these languages was, in general terms, analogous to the criterion Jones used for IE. The members of a sub-group are so much more similar to each other than they are to other IE languages that the similarity cannot be put down to chance. Now, however, there are firmer criteria for membership of a sub-group. Two languages grouped together in a sub-group are assumed to have derived from a language, the 'sub-group parent language', which is chronologically earlier than either of the grouped languages, but which was spoken after PIE. The relationship can be represented diagrammatically as a family tree, with the historically prior languages situated at higher nodes in the tree. In figure 1.1, languages A and B constitute a sub-group, since they derive from a single language intermediate between them and the parent. Languages C and D do not constitute a sub-group between each other or with either A or B.

The family tree model has been very influential in IE studies, and we shall consider it in more detail below. In some cases, as in the Romance language sub-group of IE or the Indic sub-group, we have records of an early language variety which either can be identified with the sub-group parent, or which is very close to the sub-group parent (Latin and Sanskrit in the two cases respectively). But for some other sub-groups we do not have an attested parent, and it has to be reconstructed using the comparative method. It is now generally agreed among