Languages of the World
Third Edition

Are you curious to know what all human languages have in common and in what ways they differ? Do you want to find out how language can be used to trace different peoples and their past? Then this book is for you! Now in its third edition, it guides beginners through the rich diversity of the world’s languages. It presupposes no background in linguistics, and introduces the reader to linguistic concepts with the help of problem sets, end-of-chapter exercises, and an extensive bibliography. Charts of language families provide geographical and genealogical information, and engaging sidebars with demographic, social, historical, and geographical facts help to contextualize and bring languages to life. This edition includes a fully updated glossary of all linguistic terms used, new problem sets, and a new section on cartography. Supplementary online materials include links to all websites mentioned and answers to the exercises for instructors.

ASYA PERELTSVAIG is a native speaker of Russian and a fluent speaker of three other languages. For the last fifteen years, she has taught courses in linguistics at Yale, Cornell, Stanford, and several other universities around the world. Her academic interests include languages, history, genetics, and the relationship between the three.
LANGUAGES OF THE WORLD
An Introduction
Third Edition

ASYA PERELTSVAIG
Independent scholar
To my parents, Freyda and Michael Pereltsvaig, ðŸ“
Contents

List of Figures  page x
List of Tables xi
List of Textboxes xiii
List of “Focus on” Topics xiv
Preface to the Third Edition xv
Acknowledgments xvii
Abbreviations Used in the Glosses xviii

1 Introduction  1
1.1 Languages, Dialects, and Accents  5
1.2 Language Families  10
1.3 How Language Families Are Established: Comparative Reconstruction  14
1.4 Linguistic Diversity  22
1.5 How Do Languages Diversify?  26
1.6 Field Linguistics  28
1.7 Focus on: Language Maps  32
Do It Yourself  37

2 Languages of Europe  43
2.1 The Indo-European Language Family  45
2.2 The Indo-European Controversy  55
2.3 Non-Indo-European Languages of Europe  60
2.4 Focus on: Endangered Languages of Europe  65
Do It Yourself  71

3 Languages of Iran and South Asia  75
3.1 Indo-European Languages of Iran and South Asia  76
  3.1.1 Iranian Languages  78
  3.1.2 Indo-Aryan Languages  81
3.2 Dravidian Languages  87
3.3 Other Languages of South Asia  91
3.4 Focus on: Universals and the Parametric Theory of Language  95
Do It Yourself  103

4 Languages of Northern Eurasia  108
4.1 Finno-Ugric Languages  110
4.2 Other Uralic Languages  119
4.3 Turkic Languages  122
4.4 Other Languages of Siberia  128
4.5 Focus on: Evidential Markers  136
Do It Yourself  140
<table>
<thead>
<tr>
<th>Language Region</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 Languages of the Caucasus</td>
<td>146</td>
</tr>
<tr>
<td>5.1 Northwest Caucasian Languages</td>
<td>152</td>
</tr>
<tr>
<td>5.2 Northeast Caucasian Languages</td>
<td>157</td>
</tr>
<tr>
<td>5.3 Kartvelian Languages</td>
<td>163</td>
</tr>
<tr>
<td>5.4 Indo-European Languages in the Caucasus</td>
<td>168</td>
</tr>
<tr>
<td>5.5 Focus on: Head-Marking vs. Dependent-Marking</td>
<td>172</td>
</tr>
<tr>
<td>Do It Yourself</td>
<td>176</td>
</tr>
<tr>
<td>6 Languages of the Greater Middle East</td>
<td>181</td>
</tr>
<tr>
<td>6.1 Afroasiatic Languages</td>
<td>183</td>
</tr>
<tr>
<td>6.2 Semitic Languages</td>
<td>190</td>
</tr>
<tr>
<td>6.3 Berber Languages</td>
<td>203</td>
</tr>
<tr>
<td>6.4 Focus on: Language Contact</td>
<td>206</td>
</tr>
<tr>
<td>Do It Yourself</td>
<td>213</td>
</tr>
<tr>
<td>7 Languages of Sub-Saharan Africa</td>
<td>216</td>
</tr>
<tr>
<td>7.1 Nilo-Saharan Languages</td>
<td>219</td>
</tr>
<tr>
<td>7.2 Niger-Congo Languages</td>
<td>224</td>
</tr>
<tr>
<td>7.3 Khoisan Languages</td>
<td>234</td>
</tr>
<tr>
<td>7.4 Focus on: Official Languages, Trade Languages, and Creole Languages in Sub-Saharan Africa</td>
<td>240</td>
</tr>
<tr>
<td>Do It Yourself</td>
<td>245</td>
</tr>
<tr>
<td>8 Languages of Eastern Asia</td>
<td>249</td>
</tr>
<tr>
<td>8.1 Sino-Tibetan Languages</td>
<td>251</td>
</tr>
<tr>
<td>8.2 Austro-Asiatic Languages</td>
<td>261</td>
</tr>
<tr>
<td>8.3 Tai-Kadai Languages</td>
<td>265</td>
</tr>
<tr>
<td>8.4 Japanese and Korean</td>
<td>270</td>
</tr>
<tr>
<td>8.5 Focus on: Isolating Morphology and Language Change</td>
<td>275</td>
</tr>
<tr>
<td>Do It Yourself</td>
<td>280</td>
</tr>
<tr>
<td>9 Languages of the South Sea Islands</td>
<td>287</td>
</tr>
<tr>
<td>9.1 Discovery of the Austronesian Family and the Austronesian Homeland</td>
<td>292</td>
</tr>
<tr>
<td>9.2 Today’s Austronesian Languages and the Internal Classification of the Family</td>
<td>295</td>
</tr>
<tr>
<td>9.3 Linguistic Properties of Austronesian Languages</td>
<td>301</td>
</tr>
<tr>
<td>9.4 Focus on: The Mystery of Malagasy</td>
<td>311</td>
</tr>
<tr>
<td>Do It Yourself</td>
<td>319</td>
</tr>
</tbody>
</table>
10 Aboriginal Languages of New Guinea and Australia 324

10.1 Languages of New Guinea 328
  10.1.1 Overview of Papuan Languages 329
  10.1.2 Areal Papuan Features 331
  10.1.3 Tok Pisin 336

10.2 Languages of Australia 339

10.3 Focus on: Is There Such a Thing as a Primitive Language? 347

Do It Yourself 355

11 Native Languages of the Americas 359

11.1 Languages of North America 362
11.2 Languages of Meso-America 381
11.3 Languages of South America 386

11.4 Focus on: The Pirahã Controversy 394

Do It Yourself 400

12 Remaining Issues: Macro Families, Sign Languages, and Constructed Languages 406

12.1 Macro Families 406
  12.1.1 The Dene-Yeniseian Hypothesis 408
  12.1.2 Altaic and Ural-Altaic Macro Families 409
  12.1.3 The Nostratic and Eurasian Hypotheses 414
  12.1.4 Other Hypothesized Macro Families 421

12.2 Sign Languages 425

12.3 Constructed Languages 434

Do It Yourself 441

Glossary 443

Bibliography 449

Index of Languages 471

Index of Terms 481
Figures

1.1 Map of the chapters of this book page 5
1.2 Partial tree of family relationships among some Indo-European languages 12
2.1 Surviving branches within the Indo-European language family 46
2.2 The Insular Celtic Hypothesis 51
2.3 The P-Celtic Hypothesis 51
2.4 The nominative-accusative case system of Latin and the ergative-absolutive case system of Basque 65
4.1 Partial tree of the Finno-Ugric language family 112
4.2 The Finno-Ugric language family 117
4.3 The Turkic language family 124
5.1 The Northwest Caucasian language family 153
5.2 The Northeast Caucasian language family 158
5.3 The Mxedruli alphabet 164
6.1 The Semitic language family 194
7.1 The Nilo-Saharan language family (according to Greenberg 1963) 220
7.2 The Niger-Congo language family 225
8.1 The cycle of morphological type change 277
12.1 Pedersen’s Nostratic proposal 416
12.2 Illyč-Svityč and Dolgopolsky’s Nostratic proposal 417
12.3 Greenberg’s Eurasiatic proposal 418
12.4 Nostratic vs. Eurasiatic 419
12.5 Eurasian as a branch of Nostratic 419
12.6 Starostin’s Dene-Caucasian hypothesis 422
12.7 The Sino-Austric hypothesis 422
### Tables

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Cognates in English and other Germanic languages</td>
<td>13</td>
</tr>
<tr>
<td>1.2</td>
<td>Cognate numerals 1 through 10 in Eurasian languages</td>
<td>18</td>
</tr>
<tr>
<td>1.3</td>
<td>Some cognates in English, Latin, Greek, and Irish</td>
<td>18</td>
</tr>
<tr>
<td>1.4</td>
<td>Some cognates in Maori, Tongan, Samoan, and Hawaiian</td>
<td>19</td>
</tr>
<tr>
<td>1.5</td>
<td>Case forms of the word ‘wolf’ in several Indo-European languages</td>
<td>22</td>
</tr>
<tr>
<td>2.1</td>
<td>Cognates from Albanian, Greek, Armenian, and other Indo-European languages</td>
<td>52</td>
</tr>
<tr>
<td>2.2</td>
<td>Cognates from Baltic, Slavic, and other Indo-European languages</td>
<td>53</td>
</tr>
<tr>
<td>2.3</td>
<td>Cognates from three branches of Slavic languages</td>
<td>53</td>
</tr>
<tr>
<td>3.1</td>
<td>Cognates in modern Indo-Aryan languages</td>
<td>85</td>
</tr>
<tr>
<td>3.2</td>
<td>Numbers in Dravidian languages and Bengali</td>
<td>89</td>
</tr>
<tr>
<td>4.1</td>
<td>Central European numerals</td>
<td>111</td>
</tr>
<tr>
<td>4.2</td>
<td>Some Finno-Ugric cognates</td>
<td>111</td>
</tr>
<tr>
<td>4.3</td>
<td>Additional Hungarian/Finnish cognates</td>
<td>113</td>
</tr>
<tr>
<td>4.4</td>
<td>Ugric numerals</td>
<td>117</td>
</tr>
<tr>
<td>4.5</td>
<td>Cognates in several Turkic languages</td>
<td>125</td>
</tr>
<tr>
<td>5.1</td>
<td>Location and direction case markers in Avar</td>
<td>162</td>
</tr>
<tr>
<td>5.2</td>
<td>Location and direction case markers in Lezgian</td>
<td>162</td>
</tr>
<tr>
<td>5.3</td>
<td>Case forms of the noun <em>kali</em> ‘woman’ in Georgian</td>
<td>167</td>
</tr>
<tr>
<td>5.4</td>
<td>Subject and object agreement markers in Georgian</td>
<td>168</td>
</tr>
<tr>
<td>5.5</td>
<td>Cognates in Ossetian and other Indo-European languages</td>
<td>170</td>
</tr>
<tr>
<td>5.6</td>
<td>Talysh and Persian cognates</td>
<td>171</td>
</tr>
<tr>
<td>6.1</td>
<td>Cognate consonantal roots in Afroasiatic</td>
<td>188</td>
</tr>
<tr>
<td>6.2</td>
<td>Prefixal verb conjugation in Afroasiatic languages</td>
<td>189</td>
</tr>
<tr>
<td>6.3</td>
<td>Cognates in Semitic languages</td>
<td>195</td>
</tr>
<tr>
<td>6.4</td>
<td>Person/number preterite forms of the regular Spanish <em>-ar</em> verb <em>hablar</em> (‘to speak’)</td>
<td>198</td>
</tr>
<tr>
<td>6.5</td>
<td>Person/number/gender past-tense forms of the Modern Hebrew verb <em>\K-T-V</em> (‘to write’)</td>
<td>198</td>
</tr>
<tr>
<td>6.6</td>
<td>Person/number/gender past-tense forms of the Modern Standard Arabic verb <em>\K-T-B</em> (‘to write’)</td>
<td>199</td>
</tr>
<tr>
<td>7.1</td>
<td>Noun classes in Swahili</td>
<td>231</td>
</tr>
<tr>
<td>7.2</td>
<td>Symbols for click sounds in Khoisan</td>
<td>236</td>
</tr>
<tr>
<td>8.1</td>
<td>Tones in Mandarin</td>
<td>257</td>
</tr>
<tr>
<td>8.2</td>
<td>Tones in Vietnamese</td>
<td>264</td>
</tr>
</tbody>
</table>
8.3 Tones in Thai .............................................. 267
9.1 Cognates in Western and Eastern Malayo-Polynesian .......... 293
9.2 Numerals in Austronesian .................................. 295
9.3 Some Polynesian cognates ................................ 303
9.4 Some cognates in Malagasy and Ma’anyan .................... 317
10.1 Noun classes in Dyirbal .................................... 350
12.1 Proposed cognates in Altaic ............................... 410
12.2 Proposed cognates between Altaic, Korean, and Japanese .... 412
12.3 Proposed Nostratic cognates ............................. 417

© in this web service Cambridge University Press & Assessment
www.cambridge.org
1.1 Knights, Riders, and False Friends 15
1.2 Shared Innovations Are More Important Than Shared Retentions 20
1.3 Being a Small Language: Does Size Matter? 24
2.1 Word Order in the World’s Languages 48
2.2 The Balkan Sprachbund 54
2.3 The Indo-European Homeland: Alternative Theories 58
2.4 What Do Geneticists Say? 59
3.1 A New Language Discovered in India 92
3.2 Vowel Classification 93
3.3 Which Way Are You Headed? 101
4.1 Vowel Harmony 114
4.2 “Language Nest” Program to Nurture the Enets Language 120
4.3 Agglutinative Morphology 126
4.4 Why Is the Itelmen Language Disappearing? 133
5.1 Caucasus Linguistic Expeditions 150
5.2 Ejectives 155
6.1 Buried Treasures 190
7.1 Unusual Sounds of African Languages 225
7.2 Hadza and Sandawe 235
7.3 Sound Inventories and the “Out-Of-Africa” Theory 237
7.4 Indigenous African Writing Systems 242
8.1 Hmong-Mien Languages 269
8.2 Kusunda, a Language Like No Other? 274
9.1 Pioneers of the Pacific 289
9.2 Hanunó’o Color Categories 299
10.1 Language Diversity and Density 327
10.2 In-Law Language and Other Taboos 334
10.3 Light Warlpiri: A New “Mixed” Language Discovered in Northern Australia 345
11.1 Saving Native American Languages 379
11.2 Reconstructing the Lifestyles of Three Prehistoric Amazonian Tribes 389
11.3 Unusual Number Systems 396
12.1 The Borean Proposal 423
12.2 Gestures in Oral Languages 427
12.3 Adamorobe Sign Language 432
"Focus on" Topics

1.7  Language Maps  32
2.4  Endangered Languages of Europe  65
3.4  Universals and the Parametric Theory of Language  95
4.5  Evidential Markers  136
5.5  Head-marking vs. Dependent-Marking  172
6.4  Language Contact  206
7.4  Official Languages, Trade Languages, and Creole Languages in Sub-Saharan Africa  240
8.5  Isolating Morphology and Language Change  275
9.4  The Mystery of Malagasy  311
10.3  Is There Such a Thing as a Primitive Language?  347
11.4  The Pirahã Controversy  394
Preface to the Third Edition

Ancient Greeks were among the first ones to make the observation – still relevant today – that languages differ from one another. Languages use different sounds to make up different words and combine words into phrases and sentences in very different ways. When I was growing up in Saint Petersburg, Russia, my family spoke Russian. In this language, the same word is used for ‘house’ and ‘home,’ while two different words describe what Russian speakers perceive as two different colors (синий and голубой), which English speakers group as ‘blue.’ Russian has soft (palatalized) consonant sounds but lacks the English “th-sounds” (θ and ð, in a linguistic notation). English distinguishes a blue house from the blue house, while Russian has no counterpart to English articles a or the. Unlike English, Russian stacks all question words at the beginning of a sentence, as in Кто что где когдакупил? (literally ‘Who what where when bought?’). The list goes on, and each language presents its own peculiar twists. Yet, as you will discover in this book, there are many ways in which languages are similar and use the same patterns in widely dispersed parts of the world.

This book combines geolinguistics (what languages are spoken where and by whom) with typology (in what ways languages differ and in what ways they are similar). It is organized geographically (you can see the breakdown of the world into chapters in Figure 1.1): as you travel through the world’s regions, you will learn not only what languages you might encounter there (this information is easily obtainable online) but how these languages are related to one another, what peculiarities they share, and what unique features they exhibit. You will also gain an appreciation of what is “exotic” in languages, not from the perspective of English alone, but from a more global point of view. You may even appreciate how exotic English itself is, when compared to other languages of the world.

While intended primarily as an undergraduate textbook, and complete with pedagogical features described below, this book can serve as an introduction to the world of language and the world’s languages for a general reader as well. No knowledge of linguistics or of any particular language (besides English, of course!) is presupposed, and all technical terminology is explained in the text and in the glossary at the end of the book. Numerous examples are included throughout the book to illustrate the linguistic phenomena being discussed, as are URLs of websites where you can hear the sounds of various languages, explore their grammatical properties, or investigate their geographical
Whether or not you are using this book for a college class, I highly recommend that you roll up your sleeves and work through at least some of the “Do It Yourself” sections at the end of each chapter. Linguistics is a little like ballet: you cannot really learn it properly without doing it yourself. All the assignments included in this book are designed to be approachable – and solvable! – with just the information provided here or in the websites that are incorporated into specific assignments. (This third edition has been enriched by the addition of many new “Do It Yourself” problems. Answers to the problems are available for course instructors only.)

Each chapter also includes a number of features whose goal is to add interesting information, ease the reading, or keep you, the reader, on your toes. Each chapter opens with a short narrative introduction, which includes a textbox listing all the language families spoken in the region, with the numbers of chapters/sections where these languages are discussed. Each chapter includes one or more longer textboxes that cover tangential topics, like indigenous African writing systems and taboo languages, and several short “Did you know?” boxes that present tidbits and factoids about the material discussed in the main text. Each chapter also includes a “Fascinating Language” sidebar that provides a very brief overview of one of the corresponding region’s languages. Some sections are partitioned from the rest of the text as “Advanced”; these segments, concerned with slightly more complicated phenomena, are meant to still be readable without any prerequisite knowledge, but you will not hurt my feelings if you decide to skip them. In addition to being updated and revised, this third edition includes several entirely new sections, including ones discussing sign languages, constructed languages, language maps, and more.

Happy exploring!
Acknowledgments

We all stand on the shoulders of giants, and this book is no exception. It is my pleasure to express my gratitude to all those whose work has helped make this book a reality.

First and foremost, I would like to thank my linguistics teachers who taught me the wonders of linguistic diversity: Mark Baker, Lisa Travis, the late Edit Doron, and especially the late Yehuda Falk – many examples from a variety of languages used throughout this book are drawn from Yehuda’s extensive handouts and my notes from his lectures and seminars.

I am greatly indebted to my colleagues, Stéphane Goyette, Elena “Helen” Koulidobrova, and Martin W. Lewis, who read parts or all of the manuscript, asked challenging questions, caught factual and stylistic errors, and made useful suggestions for improvement.

I am also grateful to Faruk Akkuş, David Erschler, Bert Vaux, and Ruth Kramer for suggesting literature about Berber and Cushitic languages; to Fannie Goldman and Slava Suris for inspiring me to write a much needed section about sign languages; to Maria Esipova for fascinating work on gestures accompanying speech in oral languages; and to my colleagues who shared their stories about working with lesser-studied languages, language documentation, and revitalization efforts around the world (in alphabetical order): Petr Arkadiev, Marina Chumakina, Heidi Harley, Paulina Hurwitz, Rob Munro, Charles Nydorff, Rob Pensalfini, Andrey Shluinsky, and Ghil’ad Zuckermann.

I am also obliged to the anonymous reviewers for Cambridge University Press, who provided numerous suggestions for significant improvements on the first and second editions; to the Press staff, especially Helen Barton, who has made this book (in all its three editions!) possible and who worked with me on making this third edition as effective a tool for exploration and learning as possible; and to the students in the various instantiations of my “Languages of the World” course: at NYI Summer School in Saint Petersburg, LOT Summer School in Utrecht, Stanford’s Linguistics Department, Stanford Continuing Studies Program, Santa Clara University OLLI, San Francisco State University OLLI, the Forum at Rancho San Antonio, and the Nueva High School, all of whom helped me develop the material for this book.

Last but not least, I am most grateful to Kirill Glebovitskiy for his love and support.
Abbreviations Used in the Glosses

1 û rst person
2 second person
3 third person
a agent-like argument of a canonical transitive verb
abl ablative
abs absolutive
acc accusative
ad additive
adess adessive
adv adverbial clause marker (complementizer)
all allative case
antip antipassive
aor aorist
appl applicative
art article
asp aspect
at Actor-topic
aux auxiliary
bt Beneficiary-topic
caus causative
clf classifier
cmpl completive
comp complementizer
cond conditional
cont continuative
cs construct state
ct Circumstantial-topic
dat dative
dcl declarative
def definite
dem demonstrative
direct direct case
dt direction-topic
du dual
dur durative
dyn dynamic tense
erg ergative
evid evidential
exc exclamatory
f feminine
fact factitive
fin finite
foc focus
form formal
fut future
gen genitive
hab habitual
hon honorific
indef indefinite
indic indicative mood
indirect indirect case
ins instrumental
intns intensifier
ipfv imperfective
it Instrument-topic
loc locative
m masculine
mid middle voice
neg negation
nfut non-future
nmlz nominalizer/nominalization
nom nominative
nsg non-singular
obj object
obl oblique
p patient-like argument of a canonical transitive verb
part particle
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PASS</td>
<td>passive</td>
</tr>
<tr>
<td>PFV</td>
<td>perfective</td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>POSS</td>
<td>possessive</td>
</tr>
<tr>
<td>POTENT</td>
<td>potential</td>
</tr>
<tr>
<td>PROG</td>
<td>progressive</td>
</tr>
<tr>
<td>PRS</td>
<td>present</td>
</tr>
<tr>
<td>PRTV</td>
<td>partitive</td>
</tr>
<tr>
<td>PST</td>
<td>past</td>
</tr>
<tr>
<td>PTCPL</td>
<td>participle</td>
</tr>
<tr>
<td>PUNC</td>
<td>punctual</td>
</tr>
<tr>
<td>Q</td>
<td>question particle/markers</td>
</tr>
<tr>
<td>REL</td>
<td>relative</td>
</tr>
<tr>
<td>REPET</td>
<td>repetitive</td>
</tr>
<tr>
<td>SBJ</td>
<td>subject</td>
</tr>
<tr>
<td>SBJNCT</td>
<td>subjunctive</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>SGL</td>
<td>singulative</td>
</tr>
<tr>
<td>SIM</td>
<td>simple tense</td>
</tr>
<tr>
<td>TAM</td>
<td>tense-aspect-modality</td>
</tr>
<tr>
<td>TOP</td>
<td>topic</td>
</tr>
<tr>
<td>TR</td>
<td>transitive</td>
</tr>
<tr>
<td>TT</td>
<td>Theme-topic</td>
</tr>
<tr>
<td>UNPOSS</td>
<td>unpossessed noun</td>
</tr>
<tr>
<td>VOC</td>
<td>vocative</td>
</tr>
</tbody>
</table>