

# Asya Pereltsvaig

# Languages of the World

# AN INTRODUCTION

# Languages of the World

## An Introduction

ASYA PERELTSVAIG



CAMBRIDGE  
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town,  
Singapore, São Paulo, Delhi, Tokyo, Mexico City

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](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9780521175777](http://www.cambridge.org/9780521175777)

© Asya Pereltsvaig 2012

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

First published 2012

Printed in the United Kingdom at the University Press, Cambridge

*A catalogue record for this publication is available from the British Library*

*Library of Congress Cataloguing in Publication data*

Pereltsvaig, Asya, 1972–

Languages of the world : an introduction / Asya Pereltsvaig.

p. cm.

Includes bibliographical references and index.

ISBN 978-1-107-00278-4 (hardback)

1. Language and languages – Study and teaching. 2. Multilingualism. I. Title.

P53.P4443 2012

417'.7 – dc23 2011041738

ISBN 978-1-107-00278-4 Hardback

ISBN 978-0-521-17577-7 Paperback

Additional resources for this publication at [www.cambridge.org/pereltsvaig](http://www.cambridge.org/pereltsvaig)

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.

## Languages of the World

What do all human languages have in common and in what ways are they different? How can language be used to trace different peoples and their past? Are certain languages similar because of common descent or language contact? Assuming no prior knowledge of linguistics, this textbook introduces readers to the rich diversity of human languages, familiarizing students with the variety and typology of languages around the world. Linguistic terms and concepts are explained, in the text and in the glossary, and illustrated with simple, accessible examples. Eighteen language maps and numerous language family charts enable students to place a language geographically or genealogically. A supporting website includes additional language maps and sound recordings that can be used to illustrate the peculiarities of the sound systems of various languages. "Test yourself" questions throughout the book make it easier for students to analyze data from unfamiliar languages. This book includes fascinating demographic, social, historical, and geographical information about languages and the people who speak them.

ASYA PERELTSVAIG is a lecturer in the Department of Linguistics at Stanford University.

To my mother, Freyda Pereltsvayg, ז"ל

# Figures

|      |   |               |
|------|---|---------------|
| 1.1  | Partial tree of family relationships amongst Indo-European languages.                             | <i>page 9</i> |
| 2.1  | Branches within the Indo-European language family.  | 30            |
| 3.1  | Partial tree of the Finno-Ugric language family.  | 42            |
| 3.2  | The Finno-Ugric language family.  | 47            |
| 3.3  | The nominative–accusative case system of Latin and the ergative–absolutive case system of Basque. | 51            |
| 4.1  | The Northeast Caucasian language family.  | 68            |
| 4.2  | The Mxedruli alphabet.  | 73            |
| 5.1  | The Turkic language family.   | 88            |
| 5.2  | The Semitic language family.  | 96            |
| 5.3  | Wiki subjects in Maltese.   | 101           |
| 6.1  | The Niger-Congo language family.  | 110           |
| 7.1  | The cycle of morphological type change.   | 140           |
| 8.1  | The Malayo-Polynesian branch of the Austronesian language family (simplified).                    | 148           |
| 11.1 | Pedersen’s Nostratic proposal.  | 219           |
| 11.2 | Illyč-Svityč and Dolgopolsky’s Nostratic proposal.  | 221           |
| 11.3 | Greenberg’s Eurasiatic proposal.  | 222           |
| 11.4 | Nostratic vs. Eurasiatic.   | 223           |
| 11.5 | Eurasiatic as a branch of Nostratic.  | 224           |
| 11.6 | Starostin’s Dené-Caucasian hypothesis.  | 226           |
| 11.7 | The Sino-Austic hypothesis.   | 226           |

# Maps

|  |                |
|--|----------------|
| 2.1 Indo-European languages of Europe.     | <i>page</i> 28 |
| 2.2 Indo-European languages of Asia.       | 29             |
| 3.1 Non-Indo-European languages of Europe. | 40             |
| 3.2 Finno-Ugric languages.                 | 41             |
| 3.3 Dravidian languages.                   | 53             |
| 4.1 Languages of the Caucasus.             | 65             |
| 5.1 Turkic languages.                      | 87             |
| 5.2 Semitic languages.                     | 92             |
| 6.1 Languages of sub-Saharan Africa.       | 106            |
| 6.2 Languages of Western Africa.           | 107            |
| 6.3 Languages of Southern Africa.          | 107            |
| 7.1 Languages of eastern Asia.             | 125            |
| 8.1 Austronesian languages.                | 144            |
| 9.1 Languages of Papua New Guinea.         | 169            |
| 9.2 Languages of Australia.                | 174            |
| 10.1 Languages of North America.           | 185            |
| 10.2 Languages of Meso-America.            | 195            |
| 10.3 Languages of South America.           | 198            |

# Tables

|     |   |                |
|-----|---|----------------|
| 2.1 | James Parsons' list of cognate numerals (abridged).                   | <i>page</i> 14 |
| 2.2 | Numerals 1–10 in unrelated languages.                                 | 15             |
| 2.3 | Partial paradigm of Sanskrit and Latin cases.                         | 16             |
| 2.4 | English and Latin cognates.   | 18             |
| 2.5 | Pronunciation of cognates in several Romance languages.               | 18             |
| 2.6 | Some cognates in English, Latin, Greek and Irish.                     | 19             |
| 2.7 | Case forms of the word for 'wolf' in several Indo-European languages. | 21             |
| 2.8 | Romani, Hindi, Bengali, Punjabi numerals.                             | 36             |
| 3.1 | Central European numerals.  | 41             |
| 3.2 | Some Finno-Ugric cognates.  | 42             |
| 3.3 | Additional Hungarian/Finnish cognates.                                | 43             |
| 3.4 | Ugric numerals.   | 46             |
| 3.5 | Numbers in Dravidian languages and Bengali.                           | 54             |
| 4.1 | Location and direction case markers in Avar.                          | 72             |
| 4.2 | Location and direction case markers in Lezgin.                        | 72             |
| 4.3 | Case forms of the noun <i>kali</i> 'woman' in Georgian.               | 76             |
| 4.4 | Subject and object agreement markers in Georgian.                     | 77             |
| 4.5 | Cognates in Ossetian and other Indo-European languages.               | 78             |
| 4.6 | Talysh and Persian cognates.  | 80             |
| 5.1 | Cognates in several Turkic languages.                                 | 89             |
| 5.2 | Cognates in Semitic languages.  | 95             |
| 5.3 | Cognates in Uzbek and some other Turkic languages.                    | 103            |
| 6.1 | Noun classes in Swahili.  | 115            |
| 6.2 | Symbols for click sounds in Khoisan.                                  | 119            |
| 7.1 | Tones in Mandarin.  | 129            |
| 7.2 | Tones in Cantonese.   | 129            |
| 7.3 | Tones in Vietnamese.  | 134            |
| 7.4 | Tones in Thai.  | 137            |
| 8.1 | Cognates in Western and Eastern Malayo-Polynesian.                    | 146            |
| 8.2 | Numerals in Austronesian.   | 147            |
| 8.3 | Some Polynesian cognates.   | 149            |
| 8.4 | Some cognates in Malagasy and Ma'anyan.                               | 163            |



|      |  |     |
|------|--|-----|
| 9.1  | Noun classes in Dyirbal.                               | 179 |
| 11.1 | Cognate consonantal roots in Afroasiatic.              | 208 |
| 11.2 | Proposed cognates in Altaic.                           | 213 |
| 11.3 | Proposed cognates between Altaic, Korean and Japanese. | 216 |
| 11.4 | Proposed Nostratic cognates.                           | 221 |

# Words, words, words ...

Words are often the first thing that fascinate us about a language, whether our own or a more exotic one. The words on the cover of this book are those that at some point have captivated me. Some of these words have exquisite, intriguing or plain weird meanings; others mesmerize by their beautiful or otherwise unusual sound or attract by the look of their orthography.

Among the words with beautiful meanings are the Brazilian Portuguese *cafuné* meaning 'to soothe someone by tenderly running one's fingers through their hair'; my mother used to do that and this book is dedicated to her memory. Similarly delicate is the Japanese word (儼い) (pronounced [hakanai]) which refers to 'something that, in its ephemeral nature, reminds you of how short and beautiful life is, for example, cherry blossoms' (more on Japanese in Chapter 11).

Two words with heart-warming meanings made my list: the Dutch *gezelligheid* 'time spent with loved ones in a cozy atmosphere of togetherness' and the Yiddish מאמע'לשען (pronounced [mame'loshen]) meaning literally 'mother tongue' but referring most often specifically to Yiddish itself. In addition to being the speakers' native tongue, Yiddish was literally the language of mothers, while fathers prayed and traded in Hebrew (more on Yiddish in Chapter 12).

And some words have a meaning I wish we could express in English but a good translation is sorely lacking: one example of this is the Scots word *kilfuddoch* meaning roughly 'a meeting and discussion' but then also so much more (Scots is discussed in Chapter 2).

Other words fascinate me not so much by their meaning but by the way the meaning derives from the parts of the word. Take, for example, the Tok Pisin word *manmeri* 'people': it is composed from two parts, *man* (from the similar-sounding English word) plus *meri* from the English *Mary*, extended in Tok Pisin to mean all women, regardless of their name (I discuss Tok Pisin in Chapter 12). Or consider the Russian word *лоботряс* (pronounced [laba'trjas]), which comes from the words *lob* 'forehead' and *trjasti* 'shake' and means 'a lazy person, a do-nothing' – I hope you are not such a person and will read this book!

There are also words with beautiful or peculiar sound to them. One of my favorite examples is the Hebrew word בקבוק meaning 'bottle'. It is pronounced [baq'buq]; say it over and over again and it sounds like wine being poured from a bottle, doesn't it? (Turn to Chapter 5 for a further discussion of Hebrew.)

But perhaps the most strange sounds that have captivated the popular imagination in recent years are the click sounds, found in languages of southern Africa.

To illustrate, here is a Xhosa word *ukúk!<sup>h</sup>ola* meaning ‘perfume’: the sound represented here as *k!<sup>h</sup>* is a voiceless aspirated alveopalatal click (now, that’s a mouthful!). How exactly click sounds like this are pronounced is explained in Chapter 6 of this book.

Among words with unusual (for us, English speakers) sound are the Tagalog *ngilo* meaning ‘to have tooth-edge pain’, but also ‘the physical sensation of nails scratching the chalkboard’ and the Warlpiri *ngarrka* ‘man’; both of these words start with the [ŋ] sound that in English is found only at the end of words (e.g. *sing*) and in the middle of words (e.g. *singing*). Tagalog is discussed in Chapter 8 and Warlpiri – in Chapter 9.

Speaking of word-initial consonants, some languages allow us to pile them up in a way that exceeds our English-based expectations. Two examples of that are the Polish word *chrząszcz* ‘beetle’ and the Georgian word ზგვრკვნი ‘he peeled us’ pronounced [gvprskvni], with eight consonants in the beginning! (Georgian is one of the languages discussed in Chapter 4.)

Notice also that this Georgian word expresses a meaning that we would render with a three-word sentence. It is not unusual in the world’s languages to have words with sentence-like meaning; an additional example comes from Cherokee, where the word ႠႣႪ (pronounced [hinvsɪ]) means ‘hand him something flexible (like clothes, rope, etc.)’. Native American languages such as Cherokee are discussed in Chapter 10.

The last group of words here are interesting for the way they are written. For example, the French word *août* ‘August’ uses four letters to represent just one sound [u] (and you thought English spelling was the champion of perverseness?!). In contrast, the spelling of the Danish word *ø* ‘island’ is so minimal, it looks like the mathematical symbol for an empty set.

Finally, I included a few words to represent several more exotic writing systems: the Greek *αλφάβητο* ‘alphabet’ (after all, it was the Greeks who came up with the concept!), the Sanskrit *ऋतु* ‘season’ (pronounced [rtu], with the initial vowel sound), the Arabic *كتاب* ‘book’, pronounced [kitab] and representing the best-known tri-consonant Semitic root K-T-B (this vowel-less nature of Semitic roots is explained further in Chapter 5).

Last but not least is the Udmurt expression *гажаса өтиськом* ‘welcome’, pronounced [gažasa ötis’kom]; Udmurt is a Finno-Ugric language (see Chapter 3), but like many other languages in the former Soviet Union it uses the Cyrillic alphabet, even though it is not related to Slavic languages for which this alphabet was originally invented.

# Acknowledgements

This book has required a great deal of work collecting information and figuring out the best way to present it. It is my pleasure to express my gratitude to all those who have helped to make this book a reality.

First, the three *sine qua non*:

- Helen Barton, my editor at Cambridge University Press, who first suggested that I write a textbook to go with the course that I was teaching at Stanford University at the time: her continuing support, encouragement and guidance through the publication process made this book possible;
- Martin Lewis, a geographer extraordinaire: our frequent discussions of history, geography and languages shed new light on many issues discussed throughout this book and his help with the matters of cartography was invaluable;
- Karen Sherwood Sotelino, who has read different versions of the manuscript and suggested many improvements to the content, presentation and style: without her, this book would have been full of unexplained terminology, clumsy sentences and missed opportunities to use this or that English article.

I am also very grateful to all those who read parts of the manuscript, asked challenging questions, caught factual and stylistic errors, mentioned tidbits of interesting information about various languages and provided references to useful sources which I would not have found otherwise (in alphabetical order): Doug Ball, Renee A. Berger, David Erschler, Stephane Goyette, Joel Hoffman, Pavel Iosad, Olga Kagan, Ekaterina Lutikova, Johanna Nichols, Sergei Tatevosov and Rory van Tuyl. I am also indebted to the anonymous reviewer for Cambridge University Press, who provided suggestions for significant improvements.

In addition, I'd like to thank my students of the various instantiations of the "Languages of the World" course at Stanford University's Linguistics Department, Stanford Continuing Studies Program, Santa Clara University OLLI, San Francisco State University OLLI and the Forum at Rancho San Antonio, as well as the readers of my Languages of the World blog who took the time to share their comments.

Finally, I wish to express thanks to Byron Ahn, Venelina Dimitrova, Rebecca Greene, Patrycja Jablonska, Kasia Miechowicz-Mathiasen, Stephanie Shih and Vitaliy Rayz for suggesting some great words for the cover of this book.

# Abbreviations used in the glosses

|       |  |
|-------|--|
| 1     | first person                                       |
| 2     | second person                                      |
| 3     | third person                                       |
| A     | agent-like argument of a canonical transitive verb |
| ABL   | ablative   |
| ABS   | absolutive   |
| ACC   | accusative   |
| ADESS | adessive   |
| ANTIP | antipassive  |
| AOR   | aorist   |
| APPL  | applicative  |
| ART   | article  |
| ASP   | aspect   |
| AT    | actor topic  |
| AUX   | auxiliary  |
| CLF   | classifier   |
| COMP  | complementizer                                     |
| CONT  | continuative                                       |
| CT    | circumstantial topic                               |
| DAT   | dative   |
| DEF   | definite   |
| DU    | dual   |
| DUR   | durative   |
| ERG   | ergative   |
| EVID  | evidential   |
| F     | feminine   |
| FACT  | factive  |
| FORM  | formal   |
| FUT   | future   |
| GEN   | genitive   |
| HAB   | habitual   |
| HON   | honorific  |
| INS   | instrumental                                       |
| INTNS | intensifier  |
| IPFV  | imperfective                                       |

---

|        |  |
|--------|--|
| LOC    | locative   |
| M      | masculine  |
| NEG    | negation   |
| NFUT   | non-future   |
| NMLZ   | nominalizer/nominalization                           |
| NOM    | nominative   |
| OBJ    | object   |
| OBL    | oblique  |
| P      | patient-like argument of a canonical transitive verb |
| PART   | particle   |
| PASS   | passive  |
| PFV    | perfective   |
| PL     | plural   |
| POSS   | possessive   |
| POTENT | potential  |
| PROG   | progressive  |
| PRS    | present  |
| PRTV   | partitive  |
| PST    | past   |
| PUNC   | punctual   |
| Q      | question particle/marker                             |
| REFL   | reflexive  |
| REL    | relative   |
| SBJ    | subject  |
| SG     | singular   |
| SGL    | singulative  |
| TMA    | tense-mood-aspect                                    |
| TR     | transitive   |
| TT     | theme topic  |

# Contents

|  |                |
|--|----------------|
| <i>List of figures</i>                   | <i>page ix</i> |
| <i>List of maps</i>                      | <i>x</i>       |
| <i>List of tables</i>                    | <i>xi</i>      |
| <i>Words, words, words . . .</i>         | <i>xiii</i>    |
| <i>Acknowledgements</i>                  | <i>xv</i>      |
| <i>Abbreviations used in the glosses</i> | <i>xvi</i>     |

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>  | <b>1</b>  |
| 1.1      | Languages, dialects and accents                                      | 4         |
| 1.2      | Language families  | 7         |
| 1.3      | Linguistic diversity   | 11        |
| <b>2</b> | <b>Indo-European languages</b>                                       | <b>13</b> |
| 2.1      | Discovery of the Indo-European family and comparative reconstruction | 13        |
| 2.2      | The Indo-European controversy  | 21        |
| 2.3      | The Indo-European realm  | 27        |
| 2.4      | Focus on: Lesser-known Indo-European languages                       | 32        |
| <b>3</b> | <b>Non-Indo-European languages of Europe and India</b>               | <b>39</b> |
| 3.1      | Finno-Ugric languages  | 39        |
| 3.2      | Basque   | 48        |
| 3.3      | Dravidian languages  | 52        |
| 3.4      | Focus on: Universals and the parametric theory of language           | 56        |
| <b>4</b> | <b>Languages of the Caucasus</b>                                     | <b>64</b> |
| 4.1      | Northwest Caucasian languages  | 66        |
| 4.2      | Northeast Caucasian languages  | 67        |
| 4.3      | Kartvelian languages   | 73        |
| 4.4      | Indo-European languages in the Caucasus                              | 77        |
| 4.5      | Focus on: Field linguistics  | 81        |
| <b>5</b> | <b>Languages of Northern Africa, Middle East and Central Asia</b>    | <b>86</b> |
| 5.1      | Turkic languages   | 86        |
| 5.2      | Semitic languages  | 92        |
| 5.3      | Focus on: Language contact   | 99        |

|  |     |
|--|-----|
| <b>6 Languages of sub-Saharan Africa</b>   | 105 |
| 6.1 Nilo-Saharan languages   | 106 |
| 6.2 Niger-Congo languages  | 109 |
| 6.3 Khoisan languages  | 117 |
| 6.4 Focus on: Official languages, trade languages and creole languages in sub-Saharan Africa | 120 |
| <b>7 Languages of eastern Asia</b>   | 124 |
| 7.1 Sino-Tibetan languages   | 124 |
| 7.2 Austro-Asiatic languages   | 132 |
| 7.3 Tai-Kadai languages  | 136 |
| 7.4 Focus on: Isolating morphology and language change                                       | 139 |
| <b>8 Languages of the South Sea Islands</b>  | 143 |
| 8.1 The Austronesian realm   | 143 |
| 8.2 The Austronesian prototype   | 146 |
| 8.3 The Polynesian controversy   | 155 |
| 8.4 Focus on: The mystery of Malagasy  | 160 |
| <b>9 Aboriginal languages of Australia and Papua New Guinea</b>                              | 166 |
| 9.1 Languages of Papua New Guinea  | 168 |
| 9.2 Languages of Australia   | 173 |
| 9.3 Focus on: Is Dyirbal a primitive language?   | 177 |
| <b>10 Native languages of the Americas</b>   | 183 |
| 10.1 Languages of North America  | 184 |
| 10.2 Languages of Meso-America   | 194 |
| 10.3 Languages of South America  | 197 |
| 10.4 Focus on: The Pirahã controversy  | 201 |
| <b>11 Macro families</b>   | 205 |
| 11.1 Afroasiatic languages   | 205 |
| 11.2 Uralic languages  | 209 |
| 11.3 Altaic languages  | 211 |
| 11.4 The Nostratic and Eurasiatic hypotheses   | 218 |
| 11.5 Other hypothesized macro language families  | 225 |
| <b>12 Pidgins, creoles and other mixed languages</b>   | 230 |
| 12.1 Pidgins   | 231 |
| 12.2 Creoles   | 237 |
| 12.3 Mixed Jewish languages  | 244 |
| <i>Glossary</i>  | 254 |
| <i>References</i>  | 260 |
| <i>Index of languages</i>  | 271 |
| <i>Index of terms</i>  | 277 |



# 1 Introduction

April 26, 1607. After a long 144 day voyage, three ships belonging to the Virginia Company of London and led by Captain John Smith make landfall at the southern edge of the mouth of the Chesapeake Bay, which they name Cape Henry. Shortly thereafter, they are forced to move their camp along the estuary, to a new location eventually known as “James Towne” or Jamestown, Virginia. Almost immediately, they encounter a group of “American Indians” who communicate with each other in what surely sounds like language, only it is quite different from the English of the Virginia Company settlers. Sure enough, there are words in this language and just like back home, people from different places pronounce the same words somewhat differently. But to an English ear, these words are unrecognizable: not only are there words for things unfamiliar to the English settlers (some of which will later be taken into the English language, like *racoon*, *moccasin*, *opossum* and others), but even words for familiar objects and concepts sound different: for example, the word for ‘sun’ is either *nepass* or *keshowghes* and the word for ‘copper’ is either *matassen* or *osawas*. And it is not just the words that are different, but so is the way the words are put together: for example, grammatical objects in this language typically precede rather than follow the verb. This pattern would not have surprised the settlers, had they come from the Basque country or Turkey or Japan, or even had they arrived 700 years earlier, but for the Virginia Company men it must have been a striking pattern. The differences between the language of these “American Indians” and English are so ear-grating that the English settlers start compiling lists of “American Indian” words: Captain John Smith himself compiles a list of about fifty words and William Strachey publishes a “dictionary” of the language containing about 1,000 words. Today, most of what we know about this language – called Powhatan and attributed to the Eastern Algonquian branch of the Algonquian language family (see Chapter 10) – derives from the descriptions of seventeenth- and eighteenth-century writers, as the language died in the 1790s when its speakers switched to English.

At about the same time as the English are colonizing the eastern seaboard of what will become the United States, the Russians are pushing into Siberia. Twenty-five years after the first encounter between the Virginia Company men and the Powhatan Indians, a Russian company of twenty or so men led by Pyotr Beketov land on the shores of the Lena River and, on September 25, 1632, found the fortified town of Yakutsk. As they settle the frozen expanse of Northeastern Siberia, the Russians too come into contact with people who