



# CONCISE ENCYCLOPEDIA OF LANGUAGES OF THE WORLD

---

KEITH BROWN • SARAH OGILVIE



# CONCISE ENCYCLOPEDIA OF LANGUAGES OF THE WORLD

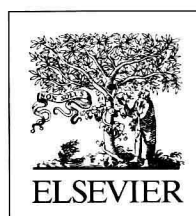
---

COORDINATING EDITOR

KEITH BROWN  
University of Cambridge  
Cambridge, UK

CO-EDITOR

SARAH OGILVIE  
University of Oxford  
Oxford, UK



Amsterdam Boston Heidelberg London New York Oxford  
Paris San Diego San Francisco Singapore Sydney Tokyo

Elsevier Ltd., The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK

© 2009 Elsevier Ltd.

*All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic, or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publishers.*

*Permissions may be sought directly from Elsevier's Rights Department in Oxford, UK: phone (+44) 1865 843830; fax (+44) 1865 853333; e-mail [permissions@elsevier.com](mailto:permissions@elsevier.com).*

*Requests may also be completed online via the homepage  
(<http://www.elsevier.com/locate/permissions>).*

First edition 2009

Library of Congress Control Number: 2008934269

A catalogue record for this book is available from the British Library

ISBN 978-0-08-087774-7

09 10 11 12 13 10 9 8 7 6 5 4 3 2 1

This book is printed on acid-free paper  
Printed and bound in China

Cover image: Adapted from *Orbis Terrarum Nova* (1594) by Petro Plancio

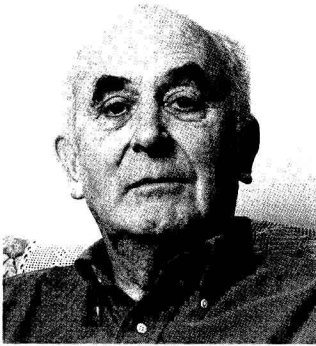
Working together to grow libraries in developing countries		
<a href="http://www.elsevier.com">www.elsevier.com</a>   <a href="http://www.bookaid.org">www.bookaid.org</a>   <a href="http://www.sabre.org">www.sabre.org</a>		
ELSEVIER	BOOK AID International	Sabre Foundation

CONCISE ENCYCLOPEDIA OF  
LANGUAGES  
OF THE WORLD

---

## THE EDITORS

---



**Keith Brown** was Editor-in-Chief of the second edition of the *Encyclopedia of Language and Linguistics* (Elsevier, 2006). He is now an Associate Lecturer in the Faculty of English at Cambridge. From 2007 he has been President of the Philological Society. From 1990 to 1994 he was President of the Linguistics Association of Great Britain, and he has been a Member of Council of the Philological Society since 1998. He is author of *Linguistics Today* (Fontana, 1984) and co-author, with Jim Miller, of *Syntax: A Linguistic Introduction to Sentence Structure* and *Syntax: Generative Grammar* (Hutchinson, 1981). Keith was joint editor of *Concise Encyclopedia of Linguistic Theories* and *Concise Encyclopedia of Grammatical Categories* (Pergamon Press, 1997 and 1998), *Common Denominators in Art and Science* (Aberdeen University Press, 1983) and *Language, Reasoning and Inference* (Academic Press, 1986).



**Sarah Ogilvie**, Trinity College, Oxford, is a linguist and lexicographer who specializes in words that enter English from non-European languages. She was Languages of the World section editor of the second edition of the *Encyclopedia of Language and Linguistics* (Elsevier, 2006), a former editor of the *Oxford English Dictionary*, and was Etymologies Editor of the *Shorter Oxford Dictionary* (6th ed., 2007).

# ALPHABETICAL LIST OF ARTICLES

---

Abkhaz	Balto-Slavic Languages
Adamawa-Ubangi	Bantu Languages
Africa as a Linguistic Area	Bashkir
Afrikaans	Basque
Afroasiatic Languages	Belorussian
Ainu	Bengali
Akan	Benue-Congo Languages
Akkadian	Berber
Albanian	Bikol
Algonquian and Ritwan Languages	Bislama
Altaic Languages	Brahui
Amharic	Breton
Anatolian Languages	Bulgarian
Ancient Egyptian and Coptic	Burmese
Andean Languages	Burushaski
Arabic	Caddoan Languages
Arabic as an Introflecting Language	Cape Verdean Creole
Arabic Languages, Variation in	Cariban Languages
Aramaic and Syriac	Catalan
Arawak Languages	Caucasian Languages
Areal Linguistics	Cebuano
Armenian	Celtic
Arrernte	Central Siberian Yupik as a Polysynthetic Language
Artificial Languages	Central Solomon Languages
Assamese	Chadic Languages
Australia: Language Situation	Chibchan
Australian Languages	Chimakuan Languages
Austic Hypothesis	Chinantec: Phonology
Austroasiatic Languages	Chinese
Austronesian Languages	Chinese as an Isolating Language
Austro-Tai Hypotheses	Choco Languages
Avestan	Chorasmian
Aymará	Chukotko-Kamchatkan Languages
Azerbaijani	Church Slavonic
Bactrian	Chuvash
Balinese	Classification of Languages
Balkans as a Linguistic Area	Cornish
Balochi	Cree

- Creek  
Crow  
Cupeño  
Cushitic Languages  
Czech  
Danish  
Dardic  
Dhivehi  
Diachronic Morphological Typology  
Dinka  
Dogon  
Domari  
Dravidian Languages  
Dutch  
Eblaite  
Efik  
Elamite  
Endangered Languages  
English in the Present Day  
English, African-American Vernacular  
English, Early Modern  
English, Later Modern (ca. 1700–1900)  
English, Middle English  
English, Old English  
English, Variation in Nonnative Varieties  
English: World Englishes  
Eskimo–Aleut  
Esperanto  
Estonian  
Ethiopia as a Linguistic Area  
Ethiopian Semitic Languages  
Ethnologue  
Etruscan  
Europe as a Linguistic Area  
Evenki  
Ewe  
Fanagalo  
Fijian  
Finnish  
Finnish as an Agglutinating Language  
Flores Languages  
Formosan Languages  
Franglais  
French  
Fulfulde  
Galician  
Gamilaraay  
Gəʼəz  
Georgian  
German  
Germanic Languages  
Gikuyu  
Goidelic Languages  
Gondi  
Gothic  
Greek, Ancient  
Greek, Modern  
Guaraní  
Gujarati  
Gullah  
Gur Languages  
Guugu Yimithirr  
Hausa  
Hawaiian  
Hawaiian Creole English  
Hebrew, Biblical and Jewish  
Hebrew, Israeli  
Highland East Cushitic Languages  
Hiligaynon  
Hindi  
Hindustani  
Hiri Motu  
Hittite  
Hmong-Mien Languages  
Hokan Languages  
Hopi  
Hungarian  
Hurrian  
Ijọ  
Ilocano  
Indo-Aryan Languages  
Indo-European Languages  
Indo-Iranian  
Inupiaq  
Iranian Languages  
Iroquoian Languages  
Italian  
Italian as a Fusional Language  
Italic Languages  
Japanese  
Javanese  
Jèrriais  
Jewish Languages  
Jiwarli  
Kalkutungu  
Kannada  
Kanuri  
Kapampangan  
Karen Languages  
Kashmiri  
Kayardild  
Kaytetye  
Kazakh  
Keres  
Ket  
Khasi  
Khmer  
Khoesaaan Languages  
Khotanese  
Kinyarwanda

- 
- Kirghiz  
 Kordofanian Languages  
 Korean  
 Krio  
 Kru Languages  
 Kurdish  
 Kurukh  
 Kwa Languages  
 Lahnda  
 Lak  
 Lakota  
 Lao  
 Latin  
 Latvian  
 Lithuanian  
 Long-Range Comparison: Methodological Disputes  
 Louisiana Creole  
 Luganda  
 Luo  
 Luxembourgish  
 Macedonian  
 Macro-Jê  
 Madang Languages  
 Madurese  
 Malagasy  
 Malay  
 Malayalam  
 Malayo-Polynesian Languages  
 Maltese  
 Malukan Languages  
 Mambila  
 Manambu  
 Mande Languages  
 Maori  
 Mapudungan  
 Marathi  
 Mayan Languages  
 Michif  
 Misumalpan  
 Mixe-Zoquean Languages  
 Mobilian Jargon  
 Mon  
 Mongolic Languages  
 Mon-Khmer Languages  
 Morphological Typology  
 Morrobalama  
 Munda Languages  
 Muskogean Languages  
 Na-Dene Languages  
 Nahuatl  
 Native American Languages  
 Native Languages of North America, Variation in  
 Navajo  
 Nenets  
 Nepali  
 Ngan'gi  
 Niger-Congo Languages  
 Nilo-Saharan Languages  
 Niuean  
 Nivkh  
 Norse and Icelandic  
 North Philippine Languages  
 Norwegian  
 Nostratic Hypothesis  
 Nuristani Languages  
 Nuuchahnulth  
 Nyanja  
 Occitan  
 Old Church Slavonic  
 Omaha-Ponca  
 Omotic Languages  
 Oneida  
 Oromo  
 Ossetic  
 Oto-Mangean Languages  
 Pahlavi  
 Palenquero  
 Pāli  
 Panoan Languages  
 Papiamentu  
 Papuan Languages  
 Pashto  
 Persian, Modern  
 Persian, Old  
 Phoenician  
 Pictish  
 Pidgins and Creoles  
 Pidgins and Creoles, Variation in  
 Pitjantjatjara/Yankunytjatjara  
 Polish  
 Pomoan Languages  
 Portuguese  
 Punjabi  
 Quechua  
 Rhaeto Romance  
 Riau Indonesian  
 Romance Languages  
 Romani  
 Romanian  
 Russenorsk  
 Russian  
 Ryukyuan  
 Saami  
 Salishan Languages  
 Samar-Leyte  
 Sango  
 Sanskrit  
 Santali  
 Scots  
 Scots Gaelic



Semitic Languages  
Serbian-Croatian-Bosnian Linguistic Complex  
Shona  
Sign Languages  
Sign Language: Morphology  
Sign Languages of the World  
Sindhi  
Sinhala  
Sino-Tibetan Languages  
Siouan Languages  
Skou Languages  
Slavic Languages  
Slovak  
Slovene  
Sogdian  
Somali  
Songhay Languages  
Sorbian  
South Asia as a Linguistic Area  
South Philippine Languages  
Southeast Asia as a Linguistic Area  
Southern Bantu Languages  
Spanish  
Sumerian  
Swahili  
Swedish  
Syriac  
Tagalog  
Tahitian  
Tai Languages  
Tajik Persian  
Tamambo  
Tamil  
Tanoan  
Tariana  
Tatar  
Telugu  
Thai  
Tibetan  
Tigrinya  
Tiwi

Tocharian  
Toda  
Tohono O'odham  
Tok Pisin  
Torricelli Languages  
Totonacan Languages  
Trans New Guinea Languages  
Tsotsi Taal  
Tucanoan Languages  
Tungusic Languages  
Tupian Languages  
Turkic Languages  
Turkish  
Turkmen  
Ugaritic  
Ukrainian  
United States of America: Language Situation  
Uralic Languages  
Urdu  
Uto-Aztecan Languages  
Uyghur  
Uzbek  
Vietnamese  
Vurës  
Wa  
Wakashan  
Wambaya  
Warlpiri  
Welsh  
West Greenlandic  
West Papuan Languages  
Wolaitta  
Wolof  
Xhosa  
Yakut  
Yanito  
Yiddish  
Yoruba  
Yukaghir  
Zapotecan  
Zulu

# SUBJECT CLASSIFICATION

---

*Note that italicized titles are included for classification purposes only and do not cross-refer to articles.*

## **Introduction**

List of Abbreviations

Classification of Languages

## **Areal Linguistics**

Africa as a Linguistic Area

Balkans as a Linguistic Area

Ethiopia as a Linguistic Area

Europe as a Linguistic Area

South Asia as a Linguistic Area

Southeast Asia as a Linguistic Area

## **Afroasiatic Languages**

Ancient Egyptian and Coptic

Berber Languages

Chadic Languages

Hausa

Cushitic Languages

Highland East Cushitic Languages

Oromo

Somali

Omotic Languages

Wolaitta

Semitic Languages

Eblaite

*Central*

Arabic

Arabic Languages, Variation in

Aramaic and Syriac

Hebrew, Biblical and Jewish

Hebrew, Israeli

Jewish languages

Maltese

Phoenician

Syriac

Ugaritic

*Eastern*

Akkadian

*Southern*

Ethiopian Semitic Languages

Amharic

Gəʿəz

Tigrinya

## **Altaic Languages**

Mongolic Languages

Tungusic Languages

Evenki

Turkic Languages

Azerbaijani

Bashkir

Chuvash

Kazakh

Kirghiz

Tatar

Turkish

Turkmen

Uyghur

Uzbek

Yakut

## **Australian Languages**

Australia: Language Situation

*Mirndi*

Wambaya

*Pama-Nyungan*

Arrernte

Gamilaraay

Guugu Yimithirr

Jiwarli

Kalkutungu

Kaytetye  
Morrobalama  
Pitjantjatjara / Yankunytjatjara  
Warlpiri

*Southern Daly*

Ngan'gi  
*Tangkic*  
Kayardild  
Tiwi

**Austroasiatic Languages**

Mon-Khmer Languages

*Eastern*

Khmer

*Northern*

Khasi

Vietnamese

Wa

*Southern*

Mon

Munda Languages

Santali

**Austronesian Languages**

Formosan Languages

Malayo-Polynesian Languages

*Central*

Flores Languages

Malukan Languages

*Oceanic*

Fijian

Maori

Tahitian

Tamambo

Vures

*Western*

Balinese

Bikol

Cebuano

Hawaiian

Hiligaynon

Ilocano

Javanese

Kapampangan

Madurese

Malagasy

Malay

Niuean

North Philippine Languages

Riau Indonesian

Samar-Leyte

South Philippine Languages

Tagalog

**Caucasian Languages**

Abkhaz

Georgian

Lak

**Chukotko-Kamchatkan Languages**

**Dravidian Languages**

Brahui

Gondi

Kannada

Kurukh

Malayalam

Tamil

Telugu

Toda

**Hmong-Mien Languages**

**Indo-European Languages**

Albanian

Anatolian Languages

Hittite

Armenian

Balto-Slavic Languages

*Baltic Languages*

Latvian

Lithuanian

Slavic Languages

Belorussian

Bulgarian

Church Slavonic

Czech

Macedonian

Old Church Slavonic

Polish

Russian

'Serbian-Croatian-Bosnian Linguistic Complex'

Slovak

Slovene

Sorbian

Ukrainian

**Celtic**

Breton

Cornish

Goidelic Languages

Pictish

Scots Gaelic

Welsh

**Germanic Languages**

Afrikaans

Danish

Dutch

English, Old English

English, Middle English

English, Early Modern  
 English, Later Modern (ca. 1700–1900)  
 English in the Present Day  
 English, World Englishes  
 English, African-American Vernacular  
 English, Variation in Nonnative Varieties  
 German  
 Gothic  
 Luxembourgish  
 Norse and Icelandic  
 Norwegian  
 Scots  
 Swedish  
 Yiddish  
*Hellenic*  
   Greek, Ancient  
   Greek, Modern  
*Indo-Iranian*  
   Indo-Aryan Languages  
     Assamese  
     Bengali  
     Dardic  
     Kashmiri  
     Dhivehi  
     Domari  
     Gujarati  
     Hindi  
     Hindustani  
     Lahnda  
     Marathi  
     Nepali  
     Nuristani Languages  
     Pāli  
     Punjabi  
     Romani  
     Sanskrit  
     Sindhi  
     Sinhala  
     Urdu  
   Iranian Languages  
     Avestan  
     Bactrian  
     Balochi  
     Chorasmian  
     Khotanese  
     Kurdish  
     Ossetic  
     Pahlavi  
     Pashto  
     Persian, Modern  
     Persian, Old  
     Sogdian  
     Tajik Persian  
*Italic Languages*  
   Latin

Romance Languages  
   Catalan  
   French  
   Galician  
   Italian  
   Jèrriais  
   Occitan  
   Portuguese  
   Rhaeto Romance  
   Romanian  
   Spanish  
   Tocharian  
  
*Khoesaaan Languages*  
*Niger-Congo Languages*  
   Adamawa-Ubangi  
   *Atlantic Congo Languages*  
     Fulfulde  
     Ijo  
     Wolof  
   Benue-Congo Languages  
     Efik  
     Mambila  
     Bantu Languages  
       Gikuyu  
       Kinyarwanda  
       Luganda  
       Nyanja  
       Shona  
       Southern Bantu Languages  
       Swahili  
       Xhosa  
       Zulu  
   Dogon  
   Gur Languages  
   Kru Languages  
   Kwa Languages  
     Akan  
     Ewe  
     Yoruba  
   Kordofanian Languages  
   Mande Languages  
  
*Nilo-Saharan Languages*  
   Dinka  
   Kanuri  
   Luo  
   Songhay Languages  
  
*North American and Middle American Languages*  
   Algonquian and Ritwan Languages  
     Cree  
     Michif  
   Caddoan Languages  
   Chibchan

Chimakuan Languages  
 Eskimo-Aleut  
   Inupiaq  
   West Greenlandic  
 Hokan Languages  
   Pomoan Languages  
 Iroquoian Languages  
   Oneida  
 Keres  
 Mayan Languages  
 Misumalpan  
 Mixe-Zoquean Languages  
 Muskogean Languages  
   Creek  
 Na-Dene Languages  
   Navajo  
 Native American Languages  
   Native Languages of North America, Variation in  
 Oto-Mangean Languages  
   Chinantec: Phonology  
   Zapotecan  
 Salishan Languages  
 Siouan Languages  
   Crow  
   Lakota  
   Omaha-Ponca  
 Tanoan Languages  
 Totonacan Languages  
 United States of America: Language Situation  
 Uto-Aztecan Languages  
   Cupeño  
   Hopi  
   Nahuatl  
   Tohono O'odham  
 Wakashan  
   Nuuchahnulth

**Papuan Languages**  
 Central Solomon Languages  
 Manambu  
 Skou Languages  
 Torricelli Languages  
 Trans New Guinea Languages  
   Madang Languages  
 West Papuan Languages

**Pidgins and Creoles**  
 Bislama  
 Cape Verdean Creole  
 Fanagalo  
 Franglais  
 Gullah

Hawaiian Creole English  
 Hiri Motu  
 Krio  
 Louisiana Creole  
 Mobilian Jargon  
 Palenquero  
 Papamientu  
 Pidgins and Creoles, Variation in  
 Russenorsk  
 Sango  
 Tok Pisin  
 Tsotsi Taal  
 Yanito

**Sign Languages**  
 Sign Languages of the World  
 Sign Language: Morphology

**Sino-Tibetan Languages**  
*Sinitic Languages*  
   Chinese  
*Tibeto-Burman Languages*  
   Burmese  
   Karen Languages  
   Tibetan

**South American Languages**  
 Andean Languages  
   Aymará  
   Quechua  
 Arawak Languages  
   Tariana  
 Cariban Languages  
 Choco Languages  
 Macro-Jê  
 Mapudungan  
 Panoan Languages  
 Tucanoan Languages  
 Tupian Languages  
   Guarani

**Tai Kadai Languages**  
 Lao  
 Thai

**Uralic Languages**  
 Estonian  
 Finnish  
 Hungarian  
 Nenets  
 Saami

**Language Isolates and Languages of Disputed Affiliation**

Ainu  
Basque  
Burushaski  
Elamite  
Etruscan  
Hurrian  
Japanese  
Ket  
Korean  
Nivkh  
Ryukyuan  
Sumerian  
Yukaghir

**Artificial Languages**

Esperanto

*Issues in Language Typology and Classification*

Endangered Languages

Ethnologue

Morphological Typology

Diachronic Morphological Typology

Arabic as an Introflexing Language

Chinese as an Isolating Language

Central Siberian Yupik as a Polysynthetic Language

Finnish as an Agglutinating Language

Italian as a Fusional Language

Long-Range Comparison: Methodological Disputes

Austic hypothesis

Austro-Tai hypothesis

Nostratic hypothesis

# INTRODUCTION

---

In this volume, the world's leading experts describe many of the languages of the world. It is estimated that there are more than 250 established language families in the world, and over 6800 distinct languages, many of which are threatened or endangered. This volume provides the most comprehensive survey available on a large proportion of these. It contains 377 articles on specific languages or language families drawn from the two editions of the *Encyclopedia of Language and Linguistics* (ELL). The articles describe the sounds, meaning, structure, and family relationships of the languages, and have been chosen to illustrate the range and diversity of human language.

The *Concise Encyclopedia of Languages of the World* is unrivalled in its scope and content. We include articles on all the large language families, such as *Austronesian* by Tony Crowley, *Niger-Congo* by John Bendor-Samuel, and *Indo-European* by Neville Collinge; on many smaller families, like the North American *Iroquoian* by Marianne Mithun and *Caddoan* by David Rood; and on many 'language isolates', languages with disputed genetic affiliation to any other language, such as *Burushaski* by Greg Anderson, *Basque* by José Hualde, and *Japanese* by Masayoshi Shibatani. We have included a few languages which are no longer spoken but which have been important for historical linguistics, like *Ancient Egyptian* by John Ray, *Hittite* by J G McQueen, and *Pictish* by William Nicolaisen. There are also articles on pidgins and creoles spoken all over the world, from an article by Suzanne Romaine on *Tok Pisin* in Papua New Guinea to another by Raj Mesthrie on *Fanagalo* in southern Africa; as well as various articles on *Sign languages* by Wendy Sandler, Ulrike Zeshan, and Trevor Johnston respectively.

All the world's major languages are covered with articles on *Chinese* by Yueguo Gu, *Arabic* by Stephan Procházka, *Hindi* by Shaligram Shukla, and *Spanish* by Roger Wright. *English* is thoroughly described with articles on all its periods by Cynthia Allen (*Old English*), Jeremy J Smith (*Middle English*), Helena Raumolin-Brunberg (*Early Modern English*), Joan Beal (*Later Modern English*), Michael Swan (*English in the Present Day*), and Braj Kachru (*World Englishes*). Inevitably some of the languages described in this volume have very small numbers of speakers and hence are in danger of being overwhelmed and lost altogether. Some linguists estimate that as many as 50–80% of the world's languages may be at risk of extinction in the next century. Many communities and linguists around the world are working together to develop innovative ways of passing on their languages to future generations. The article *Endangered Languages* by Lenore Grenoble describes some of the reasons for language loss and proposes practical means of assessing language vitality.

The *Concise Encyclopedia of Languages of the World* is the definitive resource on the languages of the world in one compact volume. Each language article gives a brief description of the language and its speakers, together with any known or hypothesized genetic relationships, and highlights interesting phonological, semantic, and syntactic features. Similarly, the articles on language families outline the membership and distribution of the family and highlight any particular phonological, semantic, or syntactic features common to the family. There is a list of useful references for further reading at the end of each article. The articles are ordered alphabetically by language, so the reader who wishes to see the overall coverage in a particular family or area will find it helpful to consult the subject classification in the front of the volume. Many languages are known in the literature under different names or spellings. Authors have highlighted these differences, and, in some cases, explained why they have chosen one name or spelling over another. For ease of reference, all variant language

names and spellings are listed in the index. Just because a language does not have its own article, does not mean that it is not discussed in another article, so users of this volume are encouraged to work from the index in order to find information on the language they want.

## The Notion 'Language'

The identification of different languages is not a straightforward matter. Every language is characterized by variation within the speech community that uses it. If the resulting speech varieties are sufficiently similar as to be considered merely characteristic of a particular geographic region or social grouping they are generally referred to as dialects, so Cockney and Norfolk are usually considered to be dialects of English. Sometimes social, political and historical pressures are such that the varieties are considered to be distinct enough to be treated as separate languages, like *Swedish* and *Norwegian* or *Hindi* and *Urdu*. Often the question of whether two languages are varieties of a single language or distinct languages is much argued over, like *Macedonian* and *Bulgarian*, or *English* and *Scots*. The naming of a language is another point of possible contention. While most linguists estimate around 6800 languages in the world, they also recognise four or five times that number of language names. A particular language may be known by one name to scholarship and another to its speakers; thus the name '*Akan*' is not generally used by speakers of the language since Akan speech forms constitute a dialect continuum running from north to south in Ghana and different communities refer to their tongue by different names – *Asante*, *Fante*, *Tivi*, *Akuapem*, *Brong*, *Akyem* or *Kwahu*.

## Language Classification

Languages can be classified in a number of different ways and for a number of different purposes. The most common classification is 'genetic', which classifies languages into families on the basis of descent from a presumed common ancestor. 'Areal' classification groups languages together either on the basis of structural features shared across language boundaries within a geographical area, or more straightforwardly simply within a geographical area. A 'lexicostatistic' classification uses word comparisons as evidence of language relationships. A 'typological' classification supposes a small set of language types, traditionally word types (isolating, agglutinating, fusional, polysynthetic), to which languages can be assigned.

**Genetic classification** The article *Classification of Languages* by Barry Blake describes the principles underlying the classification of languages adopted in ELL2 and hence in this work. It is accompanied by a map showing the location of major language groupings worldwide. This approach is one in which languages are classified into families, based on divergence from a presumed common ancestor. Good examples are the *Dravidian languages* of Southern India and *Indo-European*. The Indo-European family includes most of the languages of Europe, Iran, Afghanistan, and the northern part of South Asia. These languages can be shown to descend from a common ancestor, a common protolanguage. There are no records of the ancestral language, but it can be reconstructed from records of daughter languages such as *Sanskrit*, *Ancient Greek*, and *Latin* by using what is known as the 'comparative method'. The method is briefly explained in the article. The comparative method relies on the existence of historical records and while this is possible for Indo-European and Dravidian languages, it is not possible in the same way for other proposed language families – the indigenous languages of the Americas or of Australia for example.

More speculative classifications, far from universally accepted, relate more language families together and hence try to explore language further back in time. These efforts are discussed in Lyle Campbell's article *Long-Range Comparison: Methodological Disputes*. One of the boldest and most controversial is the *Nostratic hypothesis*, which proposes a macrofamily consisting of Indo-European, Semitic, Berber, Kartvelian, Uralic, Altaic, Korean, Japanese, and Dravidian. Similarly ambitious is the proposed *Austro-Tai hypothesis* combining Hmong-Mien (Miao-Yao), the Tai-Kadai (or Daic) family, and Austronesian. The *Austrian hypothesis* extends this proposal to include Austroasiatic.

**Areal classification** There is a broader and a looser sense in which an areal classification can be useful. The looser sense simply groups languages together regionally. Here genetic affiliations are not firmly established but shared lexicon and similar structural features suggest that the languages in question have been in contact with each other over a long period of time. In the stricter sense, areal linguistics is concerned with the diffusion of structural features across language boundaries within a geographical area. The term 'linguistic area' refers to a geographical area in which, due to borrowing and language contact, languages of a region come to share certain structural features – not just loanwords, but also shared phonological, morphological, syntactic, and other



traits. The central feature of a linguistic area is the existence of structural similarities shared among languages where some of the languages are genetically unrelated, like Turkish and Greek in the Balkans. It is assumed that the reason the languages of the area share these traits is through contact and borrowing. In addition to a general article on *Areal Linguistics* by Lyle Campbell, this volume also includes articles on areas which have been particularly studied from an areal point of view: *Africa as a Linguistic Area* by Bernd Heine; *Balkans as a Linguistic Area* by Victor Friedman; *Ethiopia as a Linguistic Area* by Joachim Crass; *Europe as a Linguistic Area* by Thomas Stolz; *South Asia as a Linguistic Area* by Karen Ebert; *Southeast Asia as a Linguistic Area* by Walter Bisang.

**Lexicostatistic classification** Word comparisons were thought for a long time to be evidence of language family relationship, but, given a small collection of likely-looking words, it is difficult to determine whether they are really the residue of common origin and not due to chance or some other factor. Lexical comparisons by themselves are seldom convincing without additional support from other criteria. Most scholars require that basic vocabulary be part of the supporting evidence for any distant genetic relationship. Basic vocabulary is generally understood to include terms for body parts, close kinship, frequently encountered aspects of the natural world (mountain, river, cloud), and low numbers. Basic vocabulary is generally resistant to borrowing, so comparisons involving basic vocabulary items are less likely to be due to diffusion and stand a better chance of being inherited from a common ancestor than other kinds of vocabulary. Still, basic vocabulary can also be borrowed – though infrequently – so that its role as a safeguard against borrowing is not foolproof. Lexicostatistics are often used as partial evidence in discussing relationships between Southern American and African languages where there are few historical records: see for example the articles by Constenla Umaña on *Misumalpan* and *Chibchan*, and the article by David Dwyer on *Mande*.

**Typological classification** At the beginning of the nineteenth century, morphological studies identified a small set of language types related primarily to word structure. The main types were isolating (words are monomorphemic and invariable, as explained in the article on *Chinese as an Isolating Language* by Jerome Packard) agglutinating (words are formed by a root and a clearly detachable sequence of affixes, each of them expressing a separate item of meaning, as exemplified in the article *Finnish as an Agglutinating Language* by Fred Karlsson), fusional (words are formed by a root and (one or more) inflectional affixes, which are employed as a primary means to indicate the grammatical function of the words in the language; see *Italian as a Fusional Language* by Claudio Iacobini) and polysynthetic (the base is the lexical core of the word; it can be followed by a number of postbases e.g. *Central Siberian Yupik as a Polysynthetic Language* by Willem de Reuse). Further types have been added as explained in *Arabic as an Introflexing Language* by Janet Watson. This morphological typology is still of some relevance but with advances in grammatical and semantic description typological classification is nowadays refined. It extends to a range of other linguistic features and to an interest in ‘universal’ linguistic properties. Syntactic features such as word order differences between languages, case marking systems, tense and aspect distinctions, modal markers, for instance evidentiality, and serial verb construction. Phonological features such as consonant types, like ejectives or clicks, vowel or nasal harmony and stressmarking. It also includes discourse phenomena including topic marking, reference chaining, and switch reference. Features like these can be found in the index.

The articles in this volume provide fascinating insights into the structure, history, and development of language families and individual languages. They highlight the diversity of the world’s languages, from the thriving to the endangered and extinct. No other single volume matches the coverage of languages or the authority of the contributors of the *Concise Encyclopedia of Languages of the World*.

Keith Brown and Sarah Ogilvie