

OXFORD
LINGUISTICS

דייווירסיטי שפראך
đa ngôn ngữ

bahubahasa

多语言用

An Introduction to Multilingualism

Language in a Changing World

Florian Coulmas

다언어주의
flerspråklighet

Dil çeşitliliği
OXFORD TEXTBOOKS IN LINGUISTICS

An Introduction to Multilingualism

OXFORD TEXTBOOKS IN LINGUISTICS

PUBLISHED

- | | |
|--|--|
| The Grammar of Words
<i>An Introduction to Linguistic Morphology</i>
THIRD EDITION
by Geert Booij | Pragmatics
SECOND EDITION
by Yan Huang |
| A Practical Introduction to Phonetics
SECOND EDITION
by J. C. Catford | Compositional Semantics
<i>An Introduction to the Syntax/Semantics Interface</i>
by Pauline Jacobson |
| An Introduction to Multilingualism
<i>Language in a Changing World</i>
by Florian Coulmas | The History of Languages
<i>An Introduction</i>
by Tore Janson |
| Meaning in Use
<i>An Introduction to Semantics and Pragmatics</i>
THIRD EDITION
by Alan Cruse | The Lexicon
<i>An Introduction</i>
by Elisabetta Jezek |
| Natural Language Syntax
by Peter W. Culicover | A Functional Discourse Grammar for English
by Evelien Keizer |
| Principles and Parameters
<i>An Introduction to Syntactic Theory</i>
by Peter W. Culicover | Diachronic Syntax
by Ian Roberts |
| A Semantic Approach to English Grammar
by R. M. W. Dixon | Cognitive Grammar
<i>An Introduction</i>
by John R. Taylor |
| Semantic Analysis
<i>A Practical Introduction</i>
by Cliff Goddard | Linguistic Categorization
THIRD EDITION
by John R. Taylor |

IN PREPARATION

- | | |
|---|---|
| Translation
<i>Theory and Practice</i>
by Kirsten Malmkjaer | Linguistic Typology
<i>Theory, Method, Data</i>
by Jae Jung Song |
| Grammaticalization
by Heiko Narrog and Bernd Heine | Cognitive Grammar
<i>An Introduction</i>
2ND EDITION
by John R. Taylor |
| Speech Acts and Sentence Types in English
by Peter Siemund | |

An Introduction to Multilingualism

Language in a Changing World

FLORIAN COULMAS

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

© Florian Coulmas 2018

The moral rights of the author have been asserted

First Edition published in 2018

Impression: 1

All rights reserved. No part of this publication may be reproduced, stored in
a retrieval system, or transmitted, in any form or by any means, without the
prior permission in writing of Oxford University Press, or as expressly permitted
by law, by licence or under terms agreed with the appropriate reprographics
rights organization. Enquiries concerning reproduction outside the scope of the
above should be sent to the Rights Department, Oxford University Press, at the
address above

You must not circulate this work in any other form
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2017939644

ISBN 978-0-19-879110-2 (Hbk)

ISBN 978-0-19-879111-9 (Pbk)

Printed and bound by
CPI Group (UK) Ltd, Croydon, CRO 4YY

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

Contents

<i>Preface</i>	xi
<i>List of figures and tables</i>	xiii
<i>Introduction</i>	xvii
1. The polyphonic world	1
1.1 Numbers	1
1.2 Family affairs	8
1.3 Richness of languages and the wealth of nations	11
1.4 A complex system	16
1.5 Counting	20
1.6 Conclusions	23
<i>Problems and questions for discussion</i>	24
<i>Further reading</i>	24
2. Multilingualism is ...: Twenty definitions and more	25
2.1 Dictionary definitions	26
2.2 Expert assessments	31
2.2.1 Capacity	31
2.2.2 Practice	33
2.2.3 Attitude and ideology	36
2.2.4 Object of theorizing	38
2.3 Conclusions	41
<i>Problems and questions for discussion</i>	41
<i>Further reading</i>	42
3. Descriptive and theoretical concepts	43
3.1 Concepts and definitions	44
3.2 Technical terms	47
3.3 Conclusions	58
<i>Problems and questions for discussion</i>	59
<i>Further reading</i>	60

CONTENTS

4. Power, inequality, and language	61
4.1 Restricting choice	61
4.1.1 Whose language choices are restricted?	65
4.1.2 Who decides on restricting language choice?	68
4.1.3 How are restrictions on language choice justified?	70
4.1.4 Who enforces restrictions on language use and how?	72
4.1.5 Are restrictions on language choice contested and, if so, by whom?	74
4.2 Conclusions	78
<i>Problems and questions for discussion</i>	79
<i>Further reading</i>	79
5. The polyglot individual	80
5.1 Preparing the ground	80
5.2 Language skills and critical age: when is a multilingual?	85
5.3 Is there enough room for several languages in one brain, and if so, for how many?	89
5.4 Simultaneous or sequential: how important is the order of acquisition?	91
5.5 The foam of the brainwave: do bilinguals always have a dominant language?	93
5.6 The social setting: is growing up multilingual a burden?	95
5.7 Conclusions	98
<i>Problems and questions for discussion</i>	99
<i>Further reading</i>	100
6. Multilingual (international) institutions	101
6.1 Bilingual education	102
6.2 Language in international institutions	110
6.3 Unité dans la diversité: multilingualism in European institutions	112
6.4 Language management of other international institutions	118
6.5 Conclusions	125
<i>Problems and questions for discussion</i>	126
<i>Further reading</i>	127

CONTENTS

7. Talk of the town: Language in super-diverse cities	128
7.1 Cosmopolis	128
7.2 Urbanization and language	132
7.3 City language profiles	137
7.4 Conclusions	152
<i>Problems and questions for discussion</i>	153
<i>Further reading</i>	153
8. Multilingual (multiethnic) countries	154
8.1 Background	154
8.2 Community relations	155
8.3 Countries, nations, languages	161
8.4 Formative factors of national multilingualism	167
8.4.1 Age of country	167
8.4.2 Official status of language(s)	168
8.4.3 Demographic strength of languages	170
8.4.4 Minority languages	173
8.4.5 The wealth of nations	177
8.5 Conclusions	179
<i>Problems and questions for discussion</i>	179
<i>Further reading</i>	180
9. Diversity in cyberspace: The multilingual internet	181
9.1 Offline	181
9.2 Online	183
9.3 The prevalence of different languages on the internet	184
9.4 Writing in cyberspace: online literacy and quasi-orality	189
9.5 Curse or blessing for minority languages?	198
9.6 Online tools of multilingualism	201
9.7 Conclusions	204
<i>Problems and questions for discussion</i>	205
<i>Further reading</i>	205

CONTENTS

10. Integration and separation: Language	206
10.1 Clouds	206
10.2 Integration: enrichment or contamination?	207
10.3 Separation: isolation or break-up?	217
10.4 Fluidity and distinction	223
10.5 Conclusions	228
<i>Problems and questions for discussion</i>	229
<i>Further reading</i>	229
 11. Integration and separation: Society	 230
11.1 The world today	230
11.2 Territoriality	231
11.3 Code of conduct	233
11.4 Essentialism and ideology	234
11.5 Migration	237
11.6 Contact	239
11.7 Segregation and fusion	241
11.8 Minority protection	244
11.9 Stratification	247
11.10 Integration	249
11.11 Globalization	253
11.12 Conclusions	254
<i>Problems and questions for discussion</i>	255
<i>Further reading</i>	256
 12. Research methods for investigating multilingualism	 257
12.1 Individual language behaviour: naturalistic data	258
12.2 Individual language behaviour: experimental data	260
12.3 Societal multilingualism	262
12.4 Fieldwork	262
12.5 Secondary data	264

CONTENTS

12.6 Written language data	265
12.7 Concluding remarks	267
<i>Multilingual corpora</i>	268
<i>Online Resources</i>	269
<i>Bibliography</i>	279
<i>Index of Names</i>	309
<i>Index of Subjects</i>	315

Preface

The global transformations that mark our age include, among others, rapid urbanization, forced displacement of people by war and armed conflict, labour migration from south to north, demographic ageing, and the intensified spread of digital technologies. While these developments work as transformative forces each in their own right, there is also interactivity between them. For instance, mobile telephony allows refugees and migrants to communicate with people in their homeland much more than would have been possible a generation ago. Similarly, more students take part in study abroad programmes and labour markets have become more international and permeable.

Both increasing multilingualism and growing medial and scholarly occupation with it are effects of these developments, which, however, is not to say that something that used to be small is now big. While multilingualism can be understood as an outgrowth of the said trends, it is at the same time itself undergoing transformations as one aspect of the ever-changing system of the world's languages.

Multilingualism cannot be understood as a phenomenon that waxes and wanes with changing circumstances, while staying substantially what it is. Just like multilingualism in 2017 London is not the same as multilingualism in 2017 Kolkata, the multilingualism of today's Western European cities is not what it was a hundred years ago in the same cities.

Multilingualism is not only an observable objective fact, but a condition that is subject to evaluations, policies, and ideologies that shape our perception as well as the reality we create. It is not in any way my intention to advocate unconditional cultural relativism; there are many facts that can be positively established about the coexistence and interaction of multiple languages. But it is necessary to stress that, regardless of whether it is examined as an individual, societal, or political condition, multilingualism is a phenomenon about which every age produces its own truths. Wherever possible it should be conceptualized as a process rather than a state of affairs.

This book is about change. While writing it, it became clear to me that change is what most prominently characterizes multilingualism today. One conspicuous change is that, because the number of people who live

PREFACE

with two or more languages in their everyday lives—estimated at constituting more than half the world’s population—is still increasing, and that, therefore, multilingualism is today less frequently seen as a rare phenomenon than used to be the case. This book examines the many faces of multilingualism and the reasons why they keep changing.

Many colleagues have helped me to develop this perspective, and some of them have actively contributed to the enterprise. As the reader will find out in Chapter 2, they have a say on the matter, which I gratefully acknowledge here.

F. C.
Venice, February 2017

List of figures and tables

Figures

1.1	Percentage of number of languages spoken on the five continents (data from <i>Ethnologue</i> 2013)	2
1.2	Gaussian distribution of languages by number of native speakers	6
1.3	Family tree of Sino-Tibetan languages, greatly simplified	9
1.4	The world language system	20
1.5	The distribution of languages across continents	23
6.1	'Teachers call for more dialect in schools.'	105
6.2	Simple coordination game between A, column, and B, row: Four options of language choice	111
7.1	Institute for Civic and Municipal Affairs, Macau	143
7.2	Multilingual street signs in four cities: Hohhot: Mongolian, Chinese, Mandarin Pinyin; Macau: Chinese, Portuguese; New Delhi: Hindi, English, Punjabi, Urdu; Nanjing: Chinese, English, Japanese, Korean	145
7.3	Prohibition signs at a park in Tokyo. A more appropriate translation of the Japanese text would be: 'In addition, please refrain from any activities that may disturb others.'	148
7.4	Information about bulky waste collection in Düsseldorf in German, English, Turkish, Russian, and Arabic	149
8.1	Matrix for classifying languages in terms of political power and demographic strength	171
8.2	England and Wales, residents aged 3 and over whose main language is not English	173
9.1	Languages used on the internet by number of speakers, 1997 (in per cent)	186
9.2	Languages used on the internet by number of speakers, 2016, in millions	187

10.1	Loanwords and loan graphs in Japanese. Katakana: k ^ā do r ^ō n ‘card loan’, terebi ‘TV’; hiragana: native grammatical morphemes; Chinese characters: content words like madoguchi ‘teller’ and the here underlined parts of the name of the bank, Mitsubishi Tokyo UFG ginkō; roman letters: the other part of the bank’s name, acronym UFG, and BANQUIC	214
10.2	<i>Afrikaanse Taalmonument</i> (language monument) dedicated to Afrikaans	221
11.1	A protest against bilingual signage in Carinthia: Die Ortsschilder ihn Unterkärnten werden Deutsch geschriben (‘Town signs in Lower Carinthia are written in German’). The author’s proficiency in German is limited, as appears from three spelling mistakes (in bold) in the sentence.	242
11.2	Percentage of immigrant population in the fifty most populous countries, 2015	252

Tables

1.1	Number of languages by number of native speakers and their percentage of the languages of the world	4
1.2	The top twelve languages by number of native speakers	5
1.3	Linguistic diversity index calculated by dividing percentage of languages by percentage of world population per continent	7
1.4	Diversity index of the ten countries with the most languages, C/E	12
1.5	The ten poorest countries, A, and the ten richest countries, D	15
1.6	Languages with official status in several countries (2017)	18
6.1	Potential costs and benefits of bilingual education; some examples	107
6.2	Official (o) and working (w) languages of major international organizations	121
6.3	Official (o) and working (w) languages of major regional international organizations	122

LIST OF FIGURES AND TABLES

7.1 The OECD–EC definition of cities according to population size	132
7.2 Urban language profiles with social and linguistic variables	138
8.1 Permanent resident population by main language(s), 1970–2014, in per cent	163
8.2 Switzerland and Singapore, some social indicators	166
8.3 Types of minorities in modern states	175
9.1 The ten most widely used languages on the internet in millions of users, June 2016	185
9.2 Languages of websites 2016	188
9.3 The languages in which more than one million Wikipedia articles were written by October 2016	197
10.1 Preferred strategies of lexical innovation: calquing in Chinese, borrowing in Japanese	212
10.2 Shared words in Dutch, Portuguese, Indonesian, Frisian, and Tetum	218
10.3 Afrikaans and Dutch words (English for reference)	220
11.1 Common aspects of society and language in the world today	231
11.2 International migrant stock (millions)	237
11.3 Don Handelman's (1977) typology of ethnic incorporation	250

Introduction

Multilingualism is a wide canvas, too wide to be painted by the brush of a single discipline. More able hands from a variety of fields are needed to sketch the outlines and fill in the details of the intricate mosaic of linguistic forms of expression the human mind has produced and to shed light on their functions for the formation of society. Language is a crucial part of human nature, and the multiplicity of languages is part of the human condition. What Oscar Wilde supposedly once quipped about Britain and the USA, that they were divided by a common language, could be said about humanity at large, and Elias Canetti, who lived a life with several languages, called the fact that there are different languages 'the most sinister fact of the world' (Canetti 1993: 18). We are all endowed with language, in the plural, however. United by the faculty of language, a common trait that distinguishes us from other animals, we employ different languages that segregate us more rigidly than almost anything else does. Thus, commenting on the fact that a shared common language is pre-eminently considered the normal basis of nationality, Max Weber (1978: 359) called his time 'the age of language conflicts', a characterization that has hardly lost its aptness in our time.

Because language unites and separates, it is not just linguists who take a serious interest in it, but several scientific disciplines, ranging from the natural sciences, including physical anthropology, to the social sciences, including cultural anthropology, psychology, cognitive science, on to economics, and political science. They all have their specific concerns and look at language from the point of view and by means of the tools of their field. They all produce valuable knowledge about this most human of human properties. However, the apparent commonality of the object of research notwithstanding, the walls that divide scientific disciplines sometimes prove as hard to scale as language boundaries, if not more so.

For example, I have often noticed that sociologists and linguists rarely talk to each other, and when they talk about language among themselves, it is hard to recognize that they are talking about the same phenomenon, for their concepts of 'language' are quite different. In the event that sociologists take issue with language, they tend to take

languages as a given, something that has an independent existence. They may use terms such as 'mother tongue' and 'native speaker' as a matter of course, while linguists may feel compelled to define these terms or try to avoid them altogether. Social scientists have little interest in structure, language change, or the coming into existence of new languages. To linguists, on the other hand, the question is of major interest how languages, dialects, and other varieties are related to each other in terms of structure, vocabulary, and pronunciation. Sociologists think that writing is important; linguists usually do not. To linguists all languages are, in principle, equal, that is, equally promising for gaining insights about the architecture of language, whereas sociologists are more interested in the inequality of languages, their prestige, level of cultivation, whether they are used for political or religious purposes, and so on. Psychologists study the acquisition, disorder, and loss of language. Educationalists take issue with measuring the distance between languages in order to improve foreign language teaching and learning. Political scientists are intrigued by the fact that words can be loaded and used more or less skilfully in election campaigns, to inform, or misinform, or manipulate people. Schoolteachers know whether *you may turn on the TV* or *turn the TV on* and generally how to use prepositions correctly. Linguists, by contrast, have serious difficulties with the very notion of correctness. And so on.

Many other questions about language as a natural faculty of *Homo sapiens*, on the one hand, and a cultural product of distinct groups, on the other, have been studied in the past and continue to be researched today. Because language is so central to human life, various disciplines and theories are concerned with it, and I have mentioned only some. By focusing on particular aspects of language they unravel the mysteries of how children acquire language, how language connects us to the world, how it binds communities together, and allows us to absorb and communicate knowledge. In one way or another, they all contribute to our understanding of language and languages. Both the singular and the plural of 'language' seem to be equally important, certainly, if we look at the world today and that of former times. For all we know, the multiplicity of languages has accompanied the human race as long as the most sophisticated scientific tools can look back into the prehistory of the mind. Each and every language is worth studying, but just as people do not live as monads, languages are not isolated from one another. The coexistence of languages is, therefore, a field of study in its own right,

which, however, relies on the insights of many other disciplines, some of which I have mentioned above.

This book describes and explores the consequences of the multiplicity of languages from various points of view. It begins with an overview of the geographic distribution of languages on the planet and then presents a summary account of the complex history interlinking states and languages. Against the background of a discussion about language and power, it offers an account of the world language system as it exists today and from there goes on to examine various explanations grounded in history and political philosophy for where, when, and why multilingualism came to be regarded as a problem, that is, under conditions of the assumed or mandated dominance of a single language.

Subsequent chapters examine the reality of multilingualism with regard to polyglot individuals, international institutions, super-diverse cities, multiethnic countries, and the seemingly borderless cyberspace.

The final part of the book takes up theoretical issues centred upon the integrity of linguistic systems and social systems, raising questions about drawing boundaries, inclusion and exclusion, incorporation and segregation, approval and prohibition. In the light of the examples expounded in the following chapters it will become clear that nowadays multilingualism is not an exotic occurrence, but that it is rather common, and therefore presents a serious challenge for both linguistic and social theory.

1

The polyphonic world

1.1 Numbers	1
1.2 Family affairs	8
1.3 Richness of languages and the wealth of nations	11
1.4 A complex system	16
1.5 Counting	20
1.6 Conclusions	23
<i>Problems and questions for discussion</i>	24
<i>Further reading</i>	24

1.1 Numbers

Pashto, Urdu, Hindi, Nepali, Sinhalese, Dzongkha, Brunei, Armenian, Turkish, Georgian, Persian, Kazakh, Kyrgyz, Tajik, Uzbek, Turkmen, Mongolian, Lao, Khmer, Vietnamese, Thai, Burmese, Indonesian, Malaysian, Tetun, Arabic, Filipino, Iban, Hebrew, Mandarin, Tamil, Japanese, Korean. ¶ Afrikaans, Bari, Chokwe, Dyrema, Eleme, Fulfulde, Gbaya, Hausa, Idakho-Isukha-Tiriki, Jowulu, Kunda, Loko, Mama, Nupe, Ogbia, Pangwa, Pulaar, Qimant, Ronga, Serer, Swahili, Tetela, Urdu, Viduna, Wanji, Wolof, Xhosa, Yoruba, Zimba. ¶ Fijian, Samoan, Nauruan, Palauan, Tok Pisin, Hiri Motu, Ekari, Makasai, Māori, Arrernte, Kala Lagaw Ya, Tongan, Xaracuu, Rukai, Puyuma, Skou, Bislama, Tahitian. ¶ Inuit, Micmac, Navaho, Nahuatl, Yucatec Maya, Sranan, Quiché, Aymara, Apalaí, Bororo, Guaraní, Pomeranian, Quechua, Mapudungun, Aymara. ¶ Spanish, English, Portuguese, Russian, German.



Figure 1.1 Percentage of number of languages spoken on the five continents.

Source: Data from *Ethnologue* 2013. This source is used here in spite of the fact that the organization behind the *Ethnologue*'s database, the Summer Institute of Linguistics (SIL), pursues a Christian missionary agenda and, partly because of that, uses a classification system that is beset with some problems, as Kamusella (2012) has demonstrated. The International Organization for Standardization also provides a list of the world's languages, from Afar to Zazaki (ISO 639-3), but since the organization invited SIL to develop the list, it is just as well to refer to *Ethnologue*.

The first paragraph of this chapter that you just read, if you have read it, contains the same information as the map in Figure 1.1, although this is perhaps not immediately apparent. It lists 100 languages representing 100 per cent of the multitude of languages known to be used today somewhere on our planet. The enumeration is divided into five sections by the Greek letter Γ gamma, short for γλῶσσα (*glossa*) 'language'. The five sections are of variable length, the first one encompassing 32 languages, representing 32 per cent of the world's languages, the second 30 languages, representing 30 per cent of the same stock, and so on. The five sections thus refer to the five continents, but the languages in each section are not listed on the basis of the same ordering principle. The first section, includes one language each of 32 Asian countries. In Africa, the second section, the multitude of languages extends from A to Z. Oceania, in the third section, includes very old languages that developed in relative isolation for a long time as well as some very young languages that came into existence in the language-richest country on earth, Papua New Guinea, as if they had not enough already. The fourth section registers fifteen languages that are indigenous to the Americas, with one and a half exceptions (which?). And finally,

the five languages given for Europe are ordered for number of native speakers.

The map in Figure 1.1 looks slightly different than world maps based on the Mercator projection¹ which are likely what the reader is most familiar with. This is because I have adjusted the size of the continents to reflect the percentage of languages found on each. Africa, which accounts for 20 per cent of the land area on planet earth, therefore, is larger than usual, occupying 32 per cent of the surface. Asia is quite close to its conventional size, since the difference between percentage of landmass and share of the world's languages is just 3 per cent. Europe has 7 per cent land area, but only 5 per cent of languages and is therefore slightly smaller, whereas Oceania is considerably bigger. The Americas look small, because the percentage of languages, 15, is noticeably less than the 28 per cent of landmass. Two other features of the map, which are obvious, but still warrant comment are that Australia, New Zealand, and the Pacific islands are lumped together as Oceania, and that Antarctica is not shown. As for the former, this is because the spoken languages of Australia and the Pacifics are often dealt with together; and for the purpose at hand, Antarctica is irrelevant because no language has a permanent footing there, all settlements being temporary only (cf. also Figure 1.5 below).

In passing it may also be noted that the map is Eurocentric, as most world maps are, which is just to remind us of the inescapability of a point of view. The world map is a good metaphor of the inevitable bias of our considerations; for it represents the globe on a flat surface, which, as mathematician Carl Friedrich Gauss (1777–1855) proved in his 'remarkable theorem' (*Theorema Egregium*) almost 200 years ago, is not possible without distortion. We cannot avoid the bias, but we can try to be aware of it. A European point of view informs all social sciences, subtly or openly, even though we are living in the age of global networks and global-everything. When talking about language, ridding ourselves of biases and prejudices is extra difficult, because languages arouse emotions and almost everyone has opinions about language; about the 'beauty' and 'ugliness' of certain pronunciations; about some languages being harsh and aggressive and others poetic and sweet; worse, some languages not being languages at all but barbaric gibberish. When dealing

¹ This projection is so called for the Flemish cartographer Gerardus Mercator (1512–1594). It has been used for nautical purposes since the seventeenth century and, reflecting the European expansion, thus became something like the world standard of world maps.

with the multiplicity of the world's languages, questions of liking and loathing should best be set aside.

The map in Figure 1.1 reflects the geographic distribution of languages. Dividing the world population—7 billion, give or take a few hundred million—by the number of known languages—7,000, give or take a few hundred— $7^9/7,000$, is a simple calculation that gives us a neat round number: one million Γ . Had the world been designed by a mathematician, this would be the average number of speakers per language with little variance, but if indeed a mathematician (mathematical laws) had been involved in its creation, the Platonic order of the universe has been all messed up by human intervention. The average number of speakers per language says nothing about the real world, for the size of language groups varies widely, testifying to the migration and settlement of humans from their earliest dwellings in Africa to cover all continents where they shaped their own ways of speaking. In the course of history, some languages expanded, others stayed small, and many fell by the wayside. The distribution of languages across continents thus contains implicit stories about population dynamics, expansion, conquest, and growth propelled by advances of civilization. Several interesting facts can be inferred by examining the speaker populations of the world's languages. A rough numerical grouping is given in Table 1.1.

There are 134 languages with fewer than 10 speakers constituting some 2 per cent of all languages, and so on. On the whole, there are

Table 1.1 Number of languages by number of native speakers and their percentage of the languages of the world.

Language group size	Number	Percentage
<10	134	2
≤ 100	340	5.1
≤ 1000	1054	15.9
$\leq 10K$	1984	29.9
$\leq 100K$	1798	27.1
$\leq 1M$	928	14
$\leq 10M$	308	4.7
$\leq 100M$	77	1.2
>100M	8	0.1

Source: *Ethnologue* 2013.

1.1 NUMBERS

Table 1.2 The top twelve languages by number of native speakers.

Language	Number of native speakers in millions
Chinese	982
Hindi	460
English	375
Spanish	330
Portuguese	216
Bengali	215
Arabic	206
Russian	165
Japanese	127
German	105
French	79
Korean	78

Data source: [statista.com/statistics/266808/](https://www.statista.com/statistics/266808/)

many languages with few speakers and few languages with many speakers. Table 1.2 lists in descending order the twelve languages with the most native speakers. Twelve of some 7,000. Taken together their speakers account for almost half the world population, 47.6 per cent, to be exact, although striving for exactitude must be a futile endeavour here. Every figure cited in the table can be contested. Are there really 982 million native speakers of Chinese, rather than 981 million or 983 million? Not to mention the millions of Chinese babies that will be born before this book goes to press and who will grow up to be native speakers of Chinese. And what Chinese? We will come to that. At this point, suffice it to note that data on languages collected through censuses—and there are few other ways—are fraught with problems. They depend on the design of questionnaires, the purposes of the agencies that commission and execute the census, the understanding of respondents and their willingness to respond. All international comparative statistics are faced with similar difficulties which, however, in regards to language are compounded by often politically sensitive issues of language proficiency. The figures (absolute numbers even more so than percentages) must therefore be used with caution. For the purpose at hand, the order of magnitude is what counts: Less than a dozen languages, however

defined, with more than 100 million speakers and more than 4,000 languages with fewer than 100,000 speakers. Many ramifications of this disproportion for the social and political existence of languages will occupy us in subsequent chapters.

Numbers do not tell us everything and cannot always be trusted, but even if they suggest a higher degree of precision than in fact can be established, they often allow us to see relationships that would otherwise be hidden. My above speculation that no mathematician was involved in the creation of the world, especially not when it comes to the distribution of languages, may have been rash. Dieter Wunderlich (2015: 37) has pointed out that if we transform Table 1.1 into a bar graph it looks surprisingly like a Gaussian normal distribution of random variables. In this case the random variables are the languages of the world, and the values they can take are the numbers of mother tongue speakers of each. A Gaussian distribution, also called 'bell curve' (Figure 1.2), is symmetric about its mean and is more than zero over its entire real line (here: all languages). The variation of languages by number of speakers follows this pattern. The graph tells us that the median of 50 per cent is a bit less than 10,000, that is, the number of languages with up to 10,000 speakers account for slightly more than 50 per cent of all languages. Beyond that size the number of languages diminishes, the real titans with more than 100 million speakers being just 8 or 0.1 per cent of all languages.

Another interesting characteristic of the distribution of the world's languages is about continents. It again bears witness to population movements in the distant past and the nature of the habitats where migrants settled. As we have seen in Figure 1.1, the languages are dis-

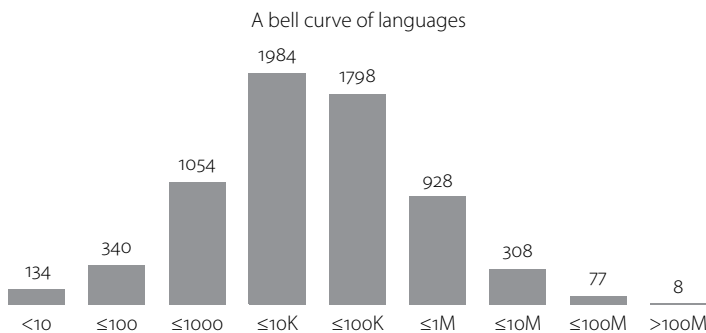


Figure 1.2 Gaussian distribution of languages by number of native speakers.

persed very unevenly across continents. The disparities are even bigger when we compare continents by percentage of world population and percentage of languages. A simple division of the latter by the former gives us a linguistic diversity index (Table 1.3).

The linguistic diversity index (LDI) of a given territory, in the event continents, is the quotient, a/b , of its share of the world's languages (a) and its share of the world population (b). On the basis of this calculation, Oceania is the most linguistically diverse continent by a large measure, since it is home to 18 per cent of all languages, but of only half a per cent of the world population. Considering the natural environment, this is not surprising. While in the period of earliest human migration the other continents were connected to each other by land bridges, Oceania has been completely separated. Many thousands of islands were first inhabited by small groups of seafaring migrants who then lived there for many generations with little or no contact with the outside world. Archaeological evidence suggests that the earliest inhabitants of Australia and New Guinea similarly lived a secluded life in small groups (O'Connell and Allen 2003). Inhabitation of Australia and New Guinea is thought to go back to the oldest migration out of Africa, maybe some 50,000 years ago. The great time depth of some 1600 generations and the fractured landscape, provided the environment for a high number of languages to develop, many hundreds in Australia and as many as 1,000 in New Guinea.

The natural environment was not the only force that favoured linguistic fragmentation—you can also call it language richness—but it did play an important role. As Daniel Nettle (1998: 357) quoting Breton (1991) and Nichols (1992) has noted, language diversity is greatest in

Table 1.3 Linguistic diversity index calculated by dividing percentage of languages by percentage of world population per continent.

Continent	Percentage of world population	Percentage of languages	Linguistic diversity index	Rank
Asia	60	32	0.53	4
Africa	15.5	30	1.94	2
Americas	14.2	15	1.1	3
Europe	10.4	5	0.5	5
Oceania	0.5	18	36	1

Source: following Wunderlich 2015: 39, with adjustments.

tropical regions near the equator where the diversity of natural species is also the greatest. Language diversity is often thought of as being analogous to natural species richness. The idea that this might not be fortuitous goes back to Charles Darwin.

1.2 Family affairs

The languages of the world differ in many ways exhibiting an astounding range of variance; but they also resemble each other in non-arbitrary ways. English resembles Dutch more than both resemble Cantonese, and Tamil has more in common with Yerukala than with Polish. English and Dutch, and Tamil and Yerukala are genetically related to each other and are therefore customarily said to belong to the same language family, the former two to the Indo-European family and the latter two to the Dravidian family. This metaphor carries a long way. It allows us to bring some order into the confusing array of languages and their inexhaustible variety.

The idea that a family of languages may be more than just a plausible analogy was first put forth by Charles Darwin in his famous treatise *On the Origin of Species* which exerted a huge influence on many sciences in the nineteenth century. As its subtitle explains, it is about ‘the preservation of favoured races in the struggle for life’. Darwin proposed a classification of the natural system in which different groups are ranked ‘under different so-called genera, sub-families, families, sections, orders, and classes’. And he went on to illustrate this view of classification ‘by taking the case of languages’ which he saw as directly linked to ‘the genealogical arrangement of the races of man’. He explained that ‘the various degrees of difference between the languages of the same stock, would have to be expressed by groups subordinate to groups; but the proper or even the only possible arrangement would still be genealogical’ (Darwin 1859: 406).

That different languages share certain features of phonology, syntax, and lexicon is apparent to anyone who has studied a foreign language. The degree of similarity depends on the amount of shared features. For instance, Mandarin, Cantonese, Vietnamese, Thai, and many other languages spoken in China and in Southeast Asia use pitch to distinguish lexical meaning. These languages are therefore known as tone (or tonal) languages. However, tone languages are also found in Africa and in North and South America. Pitch is one of many features that humans

can use to make distinctions in their speech. Similarly, noun class systems, a grammatical organization principle to categorize nouns, is a characteristic feature of Niger-Congo languages. However, Japanese and Korean also have elaborate noun class systems. Or take vowel harmony, an assimilatory process of vowels which is a distinctive feature of Turkic languages, but also of Igbo, a language spoken in south-eastern Nigeria, and Telugu of South India, among others. The occurrence of a single feature is thus not enough to establish any genealogical relationship between languages, but a clustering of features in combination with shared vocabulary does.

In the nineteenth century, Darwin's suggestion of a substantial relationship between 'human races' and languages fell on fertile ground in linguistics where the historical-comparative method made great progress putting 'the various degrees of difference' on a solid empirical footing that measures the degree of variance and distance between different members of a group of languages. Lexicostatistical methods determine the distance between languages on the basis of shared vocabulary, while linguistic typology classifies languages according to structural features, such as the use of tone, word order, morphology, etc. The set of the features that distinguish languages has been likened to a gene pool in population genetics. Progress in and standardization of language description worldwide have made reasoned classifications of languages possible. They are now commonly presented as family trees in the manner of the simplified tree of Sino-Tibetan languages in Figure 1.3. The whole family is much larger, comprising some 450 languages.

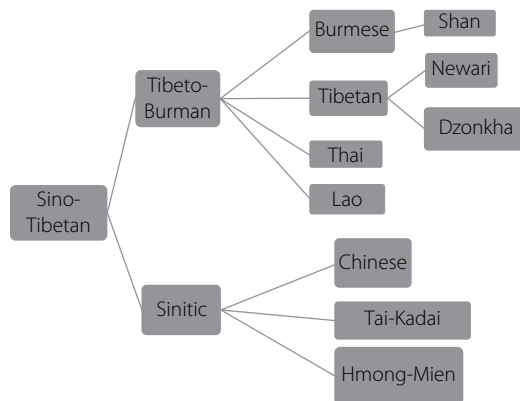


Figure 1.3 Family tree of Sino-Tibetan languages, greatly simplified.

Physical anthropologists, archaeologists, and linguists have joined forces to put Darwin's idea into practice quite beyond the metaphorical level, demonstrating that the genetic structure of population groups is significantly correlated with linguistic affiliations (e.g. for Europe: Sokal et al. 1989; for Africa: Scheinfeldt et al. 2010). There is still much controversy about particular aspects of language classification, and research about the connection between linguistic and genetic classifications continues. However, in the light of findings so far, language families must be seen as a reality that incorporates much information about the dispersal of modern humans in prehistoric times. It also invites conclusions, or at least hypotheses, about social patterns because it suggests that groups carry their languages with them and tend to stick to them. The time depth of early population movement is not definitely known, and from the number of extant languages we cannot conclude that there always were that many groups, or when and where they formed.

The following list is how *Omniglot*, an online encyclopaedia of languages, sorts languages into 45 families:

Afroasiatic, Algonqian, Altaic, Arawakan, Austroasiatic, Australian, Austronesian, Aymaran, Barbacoan, Cariban, Cahuapanan, Caucasian, Chibchan, Dravidian, Eskimo-Aleut, Guaicuruan, Hmong, Indo-European, Iroquoian, Japonic, Jivaroan, Khoisan, Mayan, Mirndi, Misumalpan, Muskogean, Na-Dené, Niger-Congo, Pama-Nyungan, Panoan, Peba-Yaguan, Oto-Manguean, Nilo-Saharan, Quechuan, Salishan, Sino-Tibetan, Siouan, Tai-Kadai, Tucanoan, Tupi-Guaraní, Uralic, Uto-Aztecan, Yenisei, Zaparoan.

Other indexes are more fine-grained, recognizing up to 300 language families. Yet others are more parsimonious making do with just 22 families (Wichmann and Grant 2012). These discrepancies suggest not only that there is still work to be done. Classifications are systems that help us to organize the world. If they are good they are derived from the world of objects and at the same time add something to what is directly perceptible. They are never quite independent of the researcher's point of view and therefore not hewn in stone. Disagreements about detail notwithstanding, there is virtually no dispute that languages can be classified genetically and that degrees of similarity and distance bear witness to the length of time since groups split up—however difficult it may be to draw a realistic timeline into the ancient past—and that

conclusions about migration flows can be drawn from the geographical distribution of languages.

Language classification systems often contain the caveat that the transplantation of European languages to other continents during the past 500 years is left out of consideration, and for good reasons, because this half millennium has seen more drastic population shifts than ever before. Yet, this self-imposed limitation is a bit ironic, for the first major discovery of genealogical relationships between languages was a by-product of the European expansion. In 1786, William Jones, a British judge stationed in Calcutta who took a serious interest in Indian languages, observed that Sanskrit had more commonalities with Greek and Latin than could reasonably be thought to be coincidental. Greece and Bengal are separated by a distance of 6,000 kilometres and a sea lane half way round the world, how could this be? The most convincing explanation: migration. Darwin's point was to relate linguistic diversity to what he called the 'arrangement of the races of man', and that is all about migration.

1.3 Richness of languages and the wealth of nations

The first migrants out of Africa went east along the coast to India, reaching Southeast Asia and Australia. They moved in small groups staying in moderate climate zones. Nomads ventured to northern latitudes of Asia and Europe only later, eventually crossing the land bridge that connected Asia with the Americas. After all continents had been colonized, the human journey continued, and it still does, changing the linguistic face of the planet; but the geographic distribution of languages still testifies to early migration flows. Today, the ten countries with the most languages are without exception in the tropics: Papua New Guinea, Indonesia, Nigeria, India, Mexico, Cameroon, Democratic Republic of the Congo, Brazil, Chad, and Myanmar, listed in Table 1.4 in descending order of number of languages.

Since countries differ widely in size and population, the number of languages per country is not a meaningful measure of linguistic diversity. The linguistic diversity index (LDI) in column F of Table 1.4, which matches percentage of world population with percentage of languages, offers a more expressive assessment. It reveals at a glance that Papua New Guinea is the real outlier, housing just 0.11 per cent of the world population but 11.81 per cent of all languages. It also shows that despite

Table 1.4 Diversity index of the ten countries with the most languages, C/E.

A	B	C	D	E	F
Country	Number of languages	% languages	Population (millions)	% of world population	C/E
PNG	802	11.81	8	0.11	107
Indonesia	742	9.95	258	3.54	2.8
Nigeria	427	7.41	186	2.56	2.8
India	405	6.39	1284	17.6	0.36
Mexico	243	4.07	122	1.67	2.4
Cameroon	275	3.96	23	0.33	12
Dem. Congo	219	0.92	85	0.65	1.4
Brazil	209	3.22	205	2.82	1.14
Chad	126	1.84	14	0.2	9.2
Myanmar	105	1.65	54	74	3.01

its many languages, the degree of India's linguistic diversity is not very high, because of its huge population.

The ten countries in Table 1.4 are clustered around the equator, and except Brazil they are poor as measured in per capita income. Taking this observation a little further, we can try to find out whether this is coincidental or indicates an interesting correlation. In a seminal contribution to charting the languages of the world, American linguist Joseph Greenberg proposed, quite in the spirit of Darwin, that by comparing disparate geographic areas, it will be possible 'to correlate varying degrees of linguistic diversity with political, economic, geographic, historic, and other non-linguistic factors' (Greenberg 1956: 109). As quoted in section 1.2, Darwin was concerned with 'the preservation of favoured races in the struggle for life'. Again, at the risk of sounding rather anthropomorphic, this notion can be applied to languages. There is a great disproportion between the number of languages of language families and their demographic strength. The six biggest families, Indo-European, Sino-Tibetan, Niger-Congo, Austronesian, Afro-Asiatic (formerly called Hamito-Semitic), and Dravidian, account for close to 90 per cent of the world population, but comprise only 65 per cent of all languages. With roughly 21 per cent or 1,524 languages the Niger-Congo family is the largest, followed by 1,221 Austronesian languages (17 per cent). The Indo-European family comprises about 5.5 per cent of all languages, but almost

twice that share of the world population. Evidently, some of these languages were 'favoured in the struggle for life'. So have Austronesian languages been, for now. They have been passed on through the generations, although their speech communities were invariably very small.

This was possible because for a long time, the region was left alone, bypassed by history, as it were. Like other peoples in remote areas, Pacific islanders were left behind in the tropics which the arrival of the white man turned into the *Tristes Tropiques*. This phrase was coined by French anthropologist Claude Lévi-Strauss in his famous melancholic memoir of his encounter with a people he studied in the rain forest of Brazil in the 1930s, and whose way of life he knew had no future. That tropical countries are rich in languages and poor in material wealth may be two sides of the same coin. For longer periods than other parts of the world they were deprived of, or protected against, the development that comes with the upheavals of war, colonization, and the ever faster rolling wheels of progress. In Oceania this is perhaps most obvious, but the general tendency can be observed in many other areas too.

Ferdinand de Saussure ([1916] 1985: 281), pioneer of structural linguistics, recognized two influences in the history of the world's languages, provincialism [*esprit de clocher*] and intercourse [*force d'intercourse*], which he thought counteract each other, but in modern times were skewed towards the side of intercourse. Isolation fosters idiosyncrasy and continuation of heredity, while intercourse propels adaptation and change. Adaptation and heredity are the key ingredients of Darwin's concept of natural selection. Evolution theory explains (predicts) that those individuals survive who are best equipped to adapt to their environment. Genetic alterations (mutations) may help organisms to adapt more quickly to their environments or, by contrast, lead to extinction. The merits of applying evolution theory to the world of languages are anything but clear, however, although languages are not living organisms but tools used by living organisms (humans), it is customary in present-day discourse to speak of the extinction of languages (Romaine and Nettle 2000). Many linguistic traditions are discontinued, as parents no longer hand down their language to their children or children refuse to use them. No mono-causal explanation can do justice to this complex social process, but on a very general level it can be said that relatively sudden exposure to modern life left the communities concerned with not enough time to adapt their languages to a drastically changed environment and they were therefore abandoned in

favour of others better adapted to and more useful for the functions of modern-day life.

This is another way of saying that every language is a system of signs governed by interrelated rules that follow their own principles, but are not invulnerable to non-linguistic factors and deliberate intervention. To every generation of speakers, their language is both a ready-made structure they acquire in early childhood, and what they make of it by adding to, transgressing, and bending the rules they were offered by their elders. Is one's mother tongue destiny? Yes, in the sense that we cannot choose the first words we hear; no, in the sense that we can decide to affiliate ourselves with a different language later in life. To answer this question more theoretically, we can apply the conceptual framework of 'structure and agency' used in the social sciences to explain social reproduction. From this point of view, human behaviour is the result of a complex interplay of objective factors—structure—and subjective factors—agency. Languages cannot evolve but as a mode of human behaviour. In many ways, they reflect the social existence of their speakers—for example kinship terminology—and their interaction with the ecosystem in which they live. Speakers can change their language, for instance by introducing generic *she* or singular *they*; they can change functional domain allocations of languages, for instance by opening the school to immigrant languages; and they can opt out of a language in favour of another. Whether they do one or the other is strongly influenced by the socioeconomic conditions in which they live.

For languages adjusted to the communication needs of hunter-gatherers, the conditions to be handed down to future generations are nowadays very unfavourable. The small size of many speech communities is an important factor, because the marketization of ever more spheres of life does not stop before languages. To linguists, the size of a language's community of speakers does not matter, to economists it does. Linguists are fascinated by the inventiveness of the human mind and the endless variation of structural options the many languages embody. Economists, by contrast, think in terms of efficiency and possibilities of reducing transaction costs. Economies of scale have produced the big companies that now dominate many markets, and similar forces favour big over small languages, and few languages over many languages. As large-scale global surveys have found, 'bilateral merchandise trade flows are higher between pairs of countries that share a common language' (Helliwell 1999: 5; see also Ginsburgh and Weber 2011: 60).