

Stephen Anderson

# LANGUAGES

A Very Short Introduction

OXFORD

# **Languages: A Very Short Introduction**



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

© Stephen R. Anderson 2012

The moral rights of the author have been asserted

First Edition published 2012

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

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloging in Publication Data

Data available

ISBN 978-0-19-959059-9

Printed in Great Britain on a acid-free paper by Ashford Colour Press Ltd,  
Gosport, Hampshire

**Very Short Introductions available now:**

ADVERTISING • Winston Fletcher

AFRICAN HISTORY • John Parker and Richard Rathbone

AGNOSTICISM • Robin Le Poidevin

AMERICAN POLITICAL PARTIES AND ELECTIONS • L. Sandy Maisel

THE AMERICAN PRESIDENCY • Charles O. Jones

ANARCHISM • Colin Ward

ANCIENT EGYPT • Ian Shaw

ANCIENT PHILOSOPHY • Julia Annas

ANCIENT WARFARE • Harry Sidebottom

ANGLICANISM • Mark Chapman

THE ANGLO-SAXON AGE • John Blair

ANIMAL RIGHTS • David DeGrazia

ANTISEMITISM • Steven Beller

THE APOCRYPHAL GOSPELS • Paul Foster

ARCHAEOLOGY • Paul Bahn

ARCHITECTURE • Andrew Ballantyne

ARISTOCRACY • William Doyle

ARISTOTLE • Jonathan Barnes

ART HISTORY • Dana Arnold

ART THEORY • Cynthia Freeland

ATHEISM • Julian Baggini

AUGUSTINE • Henry Chadwick

AUTISM • Uta Frith

BARTHES • Jonathan Culler

BESTSELLERS • John Sutherland

THE BIBLE • John Riches

BIBLICAL ARCHEOLOGY • Eric H. Cline

BIOGRAPHY • Hermione Lee

THE BOOK OF MORMON • Terryl Givens

THE BRAIN • Michael O'Shea

BRITISH POLITICS • Anthony Wright  
BUDDHA • Michael Carrithers  
BUDDHISM • Damien Keown  
BUDDHIST ETHICS • Damien Keown  
CAPITALISM • James Fulcher  
CATHOLICISM • Gerald O'Collins  
THE CELTS • Barry Cunliffe  
CHAOS • Leonard Smith  
CHOICE THEORY • Michael Allingham  
CHRISTIAN ART • Beth Williamson  
CHRISTIAN ETHICS • D. Stephen Long  
CHRISTIANITY • Linda Woodhead  
CITIZENSHIP • Richard Bellamy  
CLASSICAL MYTHOLOGY • Helen Morales  
CLASSICS • Mary Beard and John Henderson  
CLAUSEWITZ • Michael Howard  
THE COLD WAR • Robert McMahon  
COMMUNISM • Leslie Holmes  
CONSCIOUSNESS • Susan Blackmore  
CONTEMPORARY ART • Julian Stallabrass  
CONTINENTAL PHILOSOPHY • Simon Critchley  
COSMOLOGY • Peter Coles  
THE CRUSADES • Christopher Tyerman  
CRYPTOGRAPHY • Fred Piper and Sean Murphy  
DADA AND SURREALISM • David Hopkins  
DARWIN • Jonathan Howard  
THE DEAD SEA SCROLLS • Timothy Lim  
DEMOCRACY • Bernard Crick  
DESCARTES • Tom Sorell  
DESERTS • Nick Middleton  
DESIGN • John Heskett  
DINOSAURS • David Norman  
DIPLOMACY • Joseph M. Siracusa

DOCUMENTARY FILM • Patricia Aufderheide  
DREAMING • J. Allan Hobson  
DRUGS • Leslie Iversen  
DRUIDS • Barry Cunliffe  
THE EARTH • Martin Redfern  
ECONOMICS • Partha Dasgupta  
EGYPTIAN MYTH • Geraldine Pinch  
EIGHTEENTH-CENTURY BRITAIN • Paul Langford  
THE ELEMENTS • Philip Ball  
EMOTION • Dylan Evans  
EMPIRE • Stephen Howe  
ENGELS • Terrell Carver  
ENGLISH LITERATURE • Jonathan Bate  
EPIDEMIOLOGY • Roldolfo Saracci  
ETHICS • Simon Blackburn  
THE EUROPEAN UNION • John Pinder and Simon Usherwood  
EVOLUTION • Brian and Deborah Charlesworth  
EXISTENTIALISM • Thomas Flynn  
FASCISM • Kevin Passmore  
FASHION • Rebecca Arnold  
FEMINISM • Margaret Walters  
FILM MUSIC • Kathryn Kalinak  
THE FIRST WORLD WAR • Michael Howard  
FORENSIC PSYCHOLOGY • David Canter  
FORENSIC SCIENCE • Jim Fraser  
FOSSILS • Keith Thomson  
FOUCAULT • Gary Gutting  
FREE SPEECH • Nigel Warburton  
FREE WILL • Thomas Pink  
FRENCH LITERATURE • John D. Lyons  
THE FRENCH REVOLUTION • William Doyle  
FREUD • Anthony Storr  
FUNDAMENTALISM • Malise Ruthven

GALAXIES • John Gribbin  
GALILEO • Stillman Drake  
GAME THEORY • Ken Binmore  
GANDHI • Bhikhu Parekh  
GEOGRAPHY • John Matthews and David Herbert  
GEOPOLITICS • Klaus Dodds  
GERMAN LITERATURE • Nicholas Boyle  
GERMAN PHILOSOPHY • Andrew Bowie  
GLOBAL CATASTROPHES • Bill McGuire  
GLOBAL WARMING • Mark Maslin  
GLOBALIZATION • Manfred Steger  
THE GREAT DEPRESSION AND THE NEW DEAL • Eric Rauchway  
HABERMAS • James Gordon Finlayson  
HEGEL • Peter Singer  
HEIDEGGER • Michael Inwood  
HIEROGLYPHS • Penelope Wilson  
HINDUISM • Kim Knott  
HISTORY • John H. Arnold  
THE HISTORY OF ASTRONOMY • Michael Hoskin  
THE HISTORY OF LIFE • Michael Benton  
THE HISTORY OF MEDICINE • William Bynum  
THE HISTORY OF TIME • Leofranc Holford-Strevens  
HIV/AIDS • Alan Whiteside  
HOBBS • Richard Tuck  
HUMAN EVOLUTION • Bernard Wood  
HUMAN RIGHTS • Andrew Clapham  
HUME • A. J. Ayer  
IDEOLOGY • Michael Freeden  
INDIAN PHILOSOPHY • Sue Hamilton  
INFORMATION • Luciano Floridi  
INNOVATION • Mark Dodgson and David Gann  
INTELLIGENCE • Ian J. Deary  
INTERNATIONAL MIGRATION • Khalid Koser

INTERNATIONAL RELATIONS • Paul Wilkinson  
ISLAM • Malise Ruthven  
ISLAMIC HISTORY • Adam Silverstein  
JOURNALISM • Ian Hargreaves  
JUDAISM • Norman Solomon  
JUNG • Anthony Stevens  
KABBALAH • Joseph Dan  
KAFKA • Ritchie Robertson  
KANT • Roger Scruton  
KEYNES • Robert Skidelsky  
KIERKEGAARD • Patrick Gardiner  
THE KORAN • Michael Cook  
LANDSCAPES AND GEOMORPHOLOGY • Andrew Goudie and  
Heather Viles  
LAW • Raymond Wacks  
THE LAWS OF THERMODYNAMICS • Peter Atkins  
LEADERSHIP • Keth Grint  
LINCOLN • Allen C. Guelzo  
LINGUISTICS • Peter Matthews  
LITERARY THEORY • Jonathan Culler  
LOCKE • John Dunn  
LOGIC • Graham Priest  
MACHIAVELLI • Quentin Skinner  
MARTIN LUTHER • Scott H. Hendrix  
THE MARQUIS DE SADE • John Phillips  
MARX • Peter Singer  
MATHEMATICS • Timothy Gowers  
THE MEANING OF LIFE • Terry Eagleton  
MEDICAL ETHICS • Tony Hope  
MEDIEVAL BRITAIN • John Gillingham and Ralph A. Griffiths  
MEMORY • Jonathan K. Foster  
MICHAEL FARADAY • Frank A. J. L. James  
MODERN ART • David Cottington



MODERN CHINA • Rana Mitter  
MODERN IRELAND • Senia Paseta  
MODERN JAPAN • Christopher Goto-Jones  
MODERNISM • Christopher Butler  
MOLECULES • Philip Ball  
MORMONISM • Richard Lyman Bushman  
MUSIC • Nicholas Cook  
MYTH • Robert A. Segal  
NATIONALISM • Steven Grosby  
NELSON MANDELA • Elleke Boehmer  
NEOLIBERALISM • Manfred Steger and Ravi Roy  
THE NEW TESTAMENT • Luke Timothy Johnson  
THE NEW TESTAMENT AS LITERATURE • Kyle Keefer  
NEWTON • Robert Iliffe  
NIETZSCHE • Michael Tanner  
NINETEENTH-CENTURY BRITAIN • Christopher Harvie and H. C. G.  
Matthew  
THE NORMAN CONQUEST • George Garnett  
NORTHERN IRELAND • Marc Mulholland  
NOTHING • Frank Close  
NUCLEAR WEAPONS • Joseph M. Siracusa  
THE OLD TESTAMENT • Michael D. Coogan  
PARTICLE PHYSICS • Frank Close  
PAUL • E. P. Sanders  
PENTECOSTALISM • William K. Kay  
PHILOSOPHY • Edward Craig  
PHILOSOPHY OF LAW • Raymond Wacks  
PHILOSOPHY OF SCIENCE • Samir Okasha  
PHOTOGRAPHY • Steve Edwards  
PLANETS • David A. Rothery  
PLATO • Julia Annas  
POLITICAL PHILOSOPHY • David Miller  
POLITICS • Kenneth Minogue

POSTCOLONIALISM • Robert Young  
POSTMODERNISM • Christopher Butler  
POSTSTRUCTURALISM • Catherine Belsey  
PREHISTORY • Chris Gosden  
PRESOCRATIC PHILOSOPHY • Catherine Osborne  
PRIVACY • Raymond Wacks  
PROGRESSIVISM • Walter Nugent  
PSYCHIATRY • Tom Burns  
PSYCHOLOGY • Gillian Butler and Freda McManus  
PURITANISM • Francis J. Bremer  
THE QUAKERS • Pink Dandelion  
QUANTUM THEORY • John Polkinghorne  
RACISM • Ali Rattansi  
THE REAGAN REVOLUTION • Gil Troy  
THE REFORMATION • Peter Marshall  
RELATIVITY • Russell Stannard  
RELIGION IN AMERICA • Timothy Beal  
THE RENAISSANCE • Jerry Brotton  
RENAISSANCE ART • Geraldine A. Johnson  
ROMAN BRITAIN • Peter Salway  
THE ROMAN EMPIRE • Christopher Kelly  
ROMANTICISM • Michael Ferber  
ROUSSEAU • Robert Wokler  
RUSSELL • A. C. Grayling  
RUSSIAN LITERATURE • Catriona Kelly  
THE RUSSIAN REVOLUTION • S. A. Smith  
SCHIZOPHRENIA • Chris Frith and Eve Johnstone  
SCHOPENHAUER • Christopher Janaway  
SCIENCE AND RELIGION • Thomas Dixon  
SCOTLAND • Rab Houston  
SEXUALITY • Véronique Mottier  
SHAKESPEARE • Germaine Greer  
SIKHISM • Eleanor Nesbitt

SOCIAL AND CULTURAL ANTHROPOLOGY • John Monaghan and  
Peter Just

SOCIALISM • Michael Newman

SOCIOLOGY • Steve Bruce

SOCRATES • C. C. W. Taylor

THE SOVIET UNION • Stephen Lovell

THE SPANISH CIVIL WAR • Helen Graham

SPANISH LITERATURE • Jo Labanyi

SPINOZA • Roger Scruton

STATISTICS • David J. Hand

STUART BRITAIN • John Morrill

SUPERCONDUCTIVITY • Stephen Blundell

TERRORISM • Charles Townshend

THEOLOGY • David F. Ford

THOMAS AQUINAS • Fergus Kerr

TOCQUEVILLE • Harvey C. Mansfield

TRAGEDY • Adrian Poole

THE TUDORS • John Guy

TWENTIETH-CENTURY BRITAIN • Kenneth O. Morgan

THE UNITED NATIONS • Jussi M. Hanhimäki

THE U.S. CONGRESS • Donald A. Ritchie

UTOPIANISM • Lyman Tower Sargent

THE VIKINGS • Julian Richards

WITCHCRAFT • Malcolm Gaskill

WITTGENSTEIN • A. C. Grayling

WORLD MUSIC • Philip Bohlman

THE WORLD TRADE ORGANIZATION • Amrita Narlikar

WRITING AND SCRIPT • Andrew Robinson

**AVAILABLE SOON:**

LATE ANTIQUITY • Gillian Clark

MUHAMMAD • Jonathan A. Brown

GENIUS • Andrew Robinson

NUMBERS • Peter M. Higgins  
ORGANIZATIONS • Mary Jo Hatch

## **VERY SHORT INTRODUCTIONS**

*VERY SHORT INTRODUCTIONS are for anyone wanting a stimulating and accessible way in to a new subject. They are written by experts, and have been published in more than 25 languages worldwide.*

*The series began in 1995, and now represents a wide variety of topics in history, philosophy, religion, science, and the humanities. The VSI Library now contains over 200 volumes—a Very Short Introduction to everything from ancient Egypt and Indian philosophy to conceptual art and cosmology—and will continue to grow to a library of around 300 titles.*

### [VERY SHORT INTRODUCTIONS AVAILABLE NOW](#)

For more information visit our web site  
[www.oup.co.uk/general/vsi/](http://www.oup.co.uk/general/vsi/)

**Stephen R. Anderson**

# LANGUAGES

**A Very Short Introduction**

OXFORD  
UNIVERSITY PRESS

*Dedicated to the memory of Ken Hale, whose passion for the world's  
linguistic diversity was unequalled and inspiring*

# Contents

[Acknowledgements](#)

[List of illustrations](#)

[List of tables](#)

[1 Introduction: dimensions of linguistic diversity](#)

[2 How many languages are there in the world?](#)

[3 Phylogenetic linguistics: establishing linguistic relationships](#)

[4 The future of languages](#)

[5 Some problems in the counting of languages](#)

[6 The genotypes of languages](#)

[7 The diversity of signed languages](#)

[8 Conclusion: the unity of human language](#)

[References](#)

[Further reading](#)

[Index](#)



## Acknowledgements

Many others have contributed to this work, some indirectly in that I have drawn on their research, and some much more directly in supplying some of my content. Among the latter are K. David Harrison, who helped greatly with the discussion of language endangerment in [Chapter 4](#); Laurence Horn, who provided much of the material in [Chapter 5](#); Raffaella Zanuttini, who furnished an initial form of the discussion in [Chapter 6](#); and David Lightfoot, who contributed to [Chapter 8](#). These people all participated in a symposium organized by me at the Annual Meeting of the American Association for the Advancement of Science in Seattle in February 2004, entitled ‘How Many Languages Are There in the World?’ which constituted the initial impetus that has led to this book. The contents of that symposium were presented in summary form in Anderson (2004b; see the References). I would also like to thank the students in my Freshman Seminar at Yale in the autumn of 2010 devoted to this material, along with Erich Round, who co-taught that seminar with me. In addition to those mentioned above, I have received very helpful comments on preliminary versions of all or parts of the manuscript from Mark Aronoff, Andrew Carstairs-McCarthy, Norbert Hornstein, Victor Mair, Carol Padden, Sally Thomason, and two readers for Oxford University Press. As usual in such a context (but no less seriously intended for that), I stress that none of these people are responsible for the use I have made of their work, or for my failure to follow their useful suggestions.

I am also grateful to John Davey at Oxford University Press, whose idea it was that this subject would make an appropriate book for the *Very Short Introduction* series.

## List of illustrations

[1 Map showing speakers of Indo-European languages  
Wikipedia](#)

[2 Map showing the languages of New Guinea  
Wikipedia](#)

[3 Map showing the languages of part of Papua New Guinea  
© Ethnologue SIL International](#)

[4 Cultural specifics: map of France and its regional cheeses](#)

[5 Map showing the languages of France  
Wikipedia Commons](#)

[6 Map showing the languages of the northwestern USA  
© Ethnologue SIL International](#)

[7 Northern Spotted Owl  
© dannyacres-fotolia.com](#)

[8 Map showing the Sinitic languages  
Wikipedia Commons](#)

[9 Varieties of a salamander species in California](#)

[10 The sign 'tree' in three different signed languages  
from Klima and Bellugi \(1979\): pg. 21 Copyright Ursula Bellugi, The  
Salk Institute](#)

## List of tables

[1 Language families](#)

[2 Developments and similarities in words between languages](#)

[3 Sanskrit and Old English word forms compared](#)

[4 Indigenous languages of North America, with numbers of speakers](#)

[5 Threatened biological species](#)

[6 Common words compared across different Sinitic languages](#)

# Chapter 1

## Introduction: dimensions of linguistic diversity

The object of inquiry in linguistics is human language, and in particular, the extent and limits of diversity in the world's languages. It might be supposed, therefore, that linguists would have a clear and reasonably precise notion of how many languages there are in the world. As we explore various ways of approaching this question in the course of the present book, though, we will find that there is no such definite count – or at least, no such count that has any status as a scientific finding of modern linguistics. The reason for this gap is not (just) that parts of the world such as highland New Guinea or the forests of the Amazon have not been explored in sufficient detail to ascertain the range of people who live there and the languages they speak. The problem is rather that the very notion of enumerating languages is a lot more complicated than it might seem.

A recent discussion on the blog 'Language Log' puzzled over an article in a Canadian newspaper:

India concluded its national census this week, having tallied up some 1.2 billion souls, and the last night of counting focused on homeless people – of whom there are an estimated 150,000 in Delhi alone. Getting them into the count was just one in an array of staggering challenges: how to enumerate in the dozen areas under control of various armed rebel movements, and in the 572 tiny islands that make up Andaman and Nicobar; how to train 2.5 million enumerators and handle answers in 6,661 languages.

*Toronto Globe and Mail*, Wednesday, 2 March 2011

Blog comments focused on the somewhat improbable-sounding number of languages reported in India: 6,661. Suggestions included the idea that perhaps the same language had been identified by more than one name, or perhaps that ‘recognized dialects’ were included in the count, and so on. However you look at it, though, this seems a remarkably large range of languages, even for a country as ethnically diverse as modern India. But what is it we mean when we say that India has 6,661 languages, or 438 (the figure cited by *Ethnologue*, about which we will say more in [Chapter 2](#)), or any other particular figure?

## **What does it mean to ask ‘how many languages are there in the world?’?**

As a matter of basic fact, we must recognize that in the absence of severe pathology, all humans use language to express themselves and to communicate with others, but the language they use is not similarly uniform across the species. Language varies from person to person, of course, as a purely individual matter, but the modes of speaking of some are quite similar, while those of others are quite different. When people speak in much the same way, we can identify the cluster of similar systems as a particular language, and may give it a name like ‘English’ or ‘Hindi’. When we cite a specific number of languages for a country like India, or for the world as a whole, what we appear to be referring to is the number of such identifiable clusters of similar ways of speaking (or signing, as we will point out in [Chapter 7](#)). But the assumption that there *is* such a number, if only we could determine it, begs several key questions. How distinct (and discrete) are these clusters? Does every way of speaking belong to one (and only one) of them? How do we identify and individuate them, in the sea of linguistic diversity we can observe anywhere on Earth we choose to look?

In fact, the task of identifying (and then counting) the world’s languages is strikingly similar to that of identifying (and counting) the world’s biological species. Every society has names for a variety of animals and plants with which they are familiar, at a more or less comparable degree of ‘granularity’ of distinction. The ways in which

these are identified and distinguished may vary in detail, but they correspond roughly to the notion of species that biologists have tried to make precise. The effort to arrive at a satisfactory notion of ‘species’ that would be precise while also corresponding more or less to intuitive usage, though, has proven remarkably difficult.

Much of this categorization of the natural world, both naïve and scientific, depends on readily observable properties of an organism, but much does not. We clearly have to go beyond the obvious, for instance, to justify identifying both chihuahuas and Bernese mountain dogs as ‘dogs’, members of the same species (*Canis lupus familiaris*), while leopards (*Panthera pardus*), snow leopards (*Uncia uncia*), and clouded leopards (*Neofelis nebulosa*) are members not only of distinct species but on some accounts distinct genera.

Charles Darwin’s best-known book was of course *On the Origin of Species*, but that work actually has very little to contribute to the question of what species are and how they originate. The problem of speciation is one that biologists, and naturalists before them, have been arguing about since well before Darwin’s time, and one to which there is still really no agreed upon solution. Even ignoring the problem of organisms that reproduce asexually, there are a number of different ways in which biologists have proposed to define species. The notions discussed in general works on evolutionary theory such as Freeman and Herron (2007) include variations on the following:

## **The morphological species concept**

In ordinary usage across cultures, species are identified by some set of observable characteristics including their form, their behaviour, possibly the places where they are characteristically found, and so on.

## **The phylogenetic species concept**

Biological systematics provides an alternative, in classifying organisms in terms of their genealogical relation to one another. When

some group of organisms includes all of the descendants of a single common ancestral type, the group is called *monophyletic*. The smallest monophyletic groups in a genealogical tree for which there are no unique derived characters that distinguish sub-populations can then be identified as individual species.

## **The biological species concept**

Populations of organisms that can potentially provide viable, fertile offspring when mated with one another are regarded on this (widely accepted) approach as belonging to the same biological species, while populations that are reproductively isolated from one another belong to different species.

## **The genetic species concept**

A group of individuals with a characteristic set of genetic markers, such that there are few or no intermediate forms linking them to other populations, can be considered to form a species in genetic terms. It is assumed here that we can identify polymorphisms of a single gene within a species, and distinguish these from genetic differences separating it from other species.

Each of these approaches to the definition of ‘species’ has its advocates, and each has its problems. The most basic problem, though, is not that any of them is internally inconsistent, or impossible to apply, or otherwise obviously ‘wrong’: it is, rather, that while they are all reasonable approaches to what a species is, they do not always give the same results, and when taken seriously, all are deficient in one way or another in reconstructing our pre-systematic notion of what is (and is not) an identifiable species.

When we ask how we ought to individuate the world’s languages, we find rather interestingly that each of these approaches from biology has fairly direct analogues in the linguistic domain. Interestingly, many of the same problems that make the notion of species problematic for

biologists recur when we consider related issues in the definition and enumeration of languages.

Languages are not, of course, organisms, however much we tend to describe them in terms of their ‘birth’, ‘growth’, ‘descent’, ‘genetic relationship’, ‘death’, and so on. These are all metaphors, and sometimes misleading ones into the bargain. Nonetheless, the logical problem of linguistic speciation is uncannily parallel to the corresponding puzzle that arises in biology. This similarity was noted by Darwin in *The Descent of Man* when he observed that ‘[t]he formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are curiously the same’. Darwin overstates the similarity somewhat, but the parallels are still worth pursuing.

## **Approaches to a ‘census’ of languages**

In the chapters that follow, we survey a variety of perspectives from which to approach an estimation of the world’s linguistic diversity. First, [Chapter 2](#) provides a rapid sketch of some basic facts about the world’s major families of spoken languages and their distribution around the globe. The notion of ‘family’ comes into play in considering the historical relations among languages, and gives rise to a linguistic analogue of the phylogenetic approach to speciation mentioned just above. Some discussion of the basis for linguistic classification based on such historical factors is thus the goal of [Chapter 3](#).

Whatever the current degree of diversity among the world’s languages, and the number of these that we might recognize, linguists are in quite general agreement that that diversity is declining at a precipitous rate. [Chapter 4](#) discusses the reasons we can be quite sure that an astonishing proportion of the world’s languages are on an inexorable path to extinction; and also why we should care about this, in terms of the loss of human cognitive and conceptual diversity, as well as concrete traditional knowledge. The problem of language endangerment constitutes another striking parallel with biological species, for which we are quite accustomed by now to deploring the loss of diversity.



Given the remarkable number of languages that are threatened or clearly moribund in comparison with the corresponding facts for any class of biological species, the language-related version of the problem ought also to be seen as rather alarming.

One factor that is widely seen as important in the decline of particular languages is the notion that when they come into contact with one another, that necessarily leads to mortal combat in the competition for the allegiance of speakers. [Chapter 4](#) discusses some of the modes of interaction among languages in such cases, and concludes that there is no reason to believe that language contact is a sort of zero-sum game that inevitably needs to lead to the elimination of one or the other of the languages involved.

Returning to the question of how to distinguish and label the languages of the world, [Chapter 5](#) raises the question of how we identify a language, including the vexed issue of what differentiates a ‘language’ from a ‘dialect’. It turns out that much of what most people think of as criterial for identifying languages corresponds to political and social factors, and not to anything characteristic of the languages themselves. Among other things, this highlights the most common notion of how to tell when we are dealing with two languages rather than one: the criterion of mutual intelligibility. This notion is quite parallel to the one on which the biological species concept introduced above is based, and shares with it a range of problems and indeterminacies that ultimately reduce its utility and general applicability.

Without denying the importance of a political or social notion of linguistic identity, we have to conclude that it is only imperfectly related to anything in the inherent character of languages. We might, therefore, ask the language scientist to improve on this approach. The natural response is that a language is defined by its grammar (and somewhat more transiently, its vocabulary). We might then say that linguistic systems that are characterized by different grammars themselves are *ipso facto* different languages, and ask about how these grammars can be observed to differ – an approach somewhat similar to that of the genetic concept of biological species. When we do that, however, the

result is somewhat surprising: there turn out to be a great many detailed ways in which one grammar can differ from another. Since a significant number of these are logically independent, the number of possible human languages as a function of differences in grammar is probably huge, and surprisingly many of these possibilities actually turn out to be instantiated. [Chapter 6](#) discusses ways of formulating differences among languages in terms of independent dimensions, or *parameters*, of grammatical variation, and the consequences of taking such an approach to linguistic speciation.

Most of the discussion in this book concerns the kind of language that nearly all readers will be most familiar with, *spoken* languages. What else is there, you ask? But in communities of hearing-impaired individuals, there often arise languages of a superficially quite different sort: *signed*, or manual, languages, produced by the systematic use of gestures (including not only the hands but the face) and perceived visually. Although signed languages are often misconstrued as rudimentary attempts to mimic objects and actions in the world, or as an alternative way to indicate words in a spoken (or written) language such as English, research has made it abundantly clear that neither of these notions is at all accurate. Signed languages are fully expressive, natural systems of communication, and they can be deployed for all of the purposes (with all of the precision and communicative content) of spoken languages. Furthermore, signed languages turn out to instantiate many of the same general properties familiar from the study of spoken languages. It is clear that speech and signing are simply two different modalities in which the human capacity for language can be expressed.

Individual signed languages are quite independent of the spoken languages of the surrounding communities, although of course there are contact phenomena to be studied here, as in the case of spoken languages in contact. Signed languages are also independent of one another in the same way spoken languages are (even more so, perhaps, since signed languages tend to arise in complete isolation from one another), so they are just as much in need of individuation as spoken languages, and should be included in any assessment of the linguistic diversity of humankind. These matters are the topic of [Chapter 7](#).

The issue of general or universal properties of human language arises in connection with signed languages, highlighted by the initially surprising result that both signed and spoken languages can be shown to fall within the same general class of systems. When we compare the properties of languages of either type with other communication systems found in nature, in fact, we cannot fail to be impressed with how very similar they all are in the broader scheme of ways organisms can communicate. It has been suggested that, to an anthropologist from another planet studying communication and expression on Earth, the similarities among human languages would be so much more striking than the differences as to suggest that there is really only one language in use by our species, with minor local variants. The nature of linguistic universals and the fundamental differences between all human languages, on the one hand, and anything else in nature, on the other, will form the content of [Chapter 8](#).

By itself, the question ‘How many languages are there in the world?’ seems quite a trivial one, the sort to which there ought to be a quite simple answer without much intrinsic interest. When we try to ask how we might arrive at such an answer, though, the various paths that present themselves are rather more complex and interesting. In the course of exploring the bases and degree of human linguistic diversity, we can learn a great deal about the nature of language itself – one of the most distinctive characteristics of one biological species, our own.

## Chapter 2

# How many languages are there in the world?

When people are asked the question of this chapter's title, the answers vary quite a bit. One random sampling of New Yorkers, for instance, resulted in answers like 'probably at least several hundred.' However we choose to count them, though, this is not even close.

When we look at reference works, we find estimates that have escalated over time. The 11th edition of the *Encyclopedia Britannica* (1911), for example, implies a figure somewhere around 1,000, a number that climbs steadily in subsequent editions over the course of the next century. That is not due to any increase in the number of languages in the ensuing years, but rather, in large part, to our increased understanding of how many languages are actually spoken in areas that had previously been underdescribed.

## Some numbers

Much pioneering work in documenting the languages of the world has been done by missionary organizations (such as the Summer Institute of Linguistics, now known as SIL International) with an interest in translating the Christian Bible. As of 2009, at least a portion of the Bible had been translated into at least 2,508 different languages, still a long way short of full coverage. SIL International also publishes the most extensive catalogue of the world's languages, *Ethnologue*, a work that is generally taken to be as authoritative as any and whose detailed classified list included 6,909 distinct languages in its latest (2009) edition.

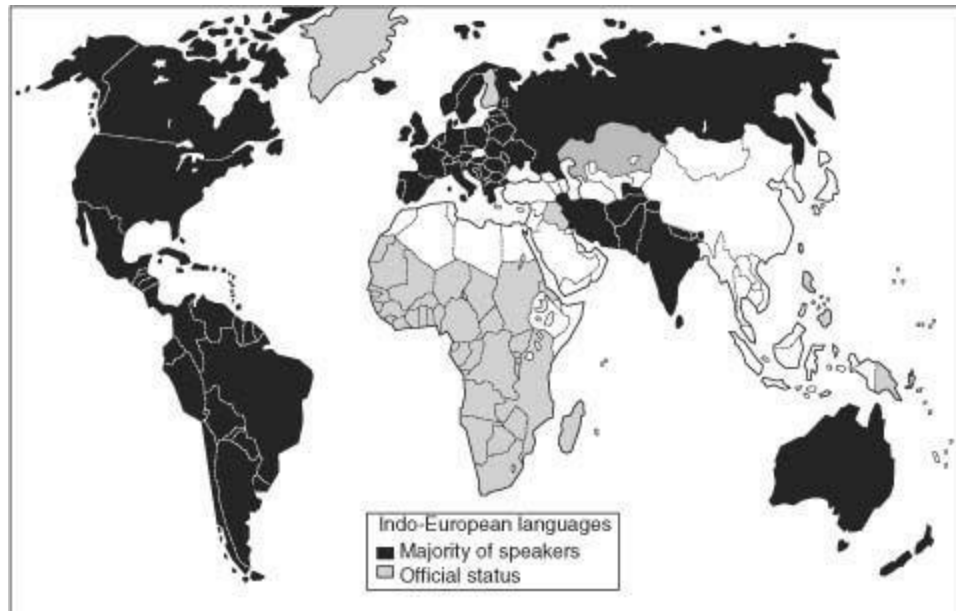
A family is a group of languages that can be shown to be genetically

related to one another; the method by which this kind of relationship is demonstrated will be discussed in [Chapter 3](#). The languages that are likely to be best known to readers of the present book, including the one in which it is written (English), are primarily members of the large Indo-European family. Considering how widely the Indo-European languages are distributed geographically, and their influence in world affairs, one might assume that a good proportion of the world's languages belong to this family. This is not the case, however: there are only about 400 Indo-European languages, although many of these are spoken over large territories (see [Figure 1](#)).

In fact, even ignoring the many cases in which a language's genetic affiliation cannot be clearly determined, there are probably more independent *families* of languages in the world (at least several hundred) than there are members of the Indo-European family.

Languages are not at all uniformly distributed around the world. Just as some places are more diverse than others in terms of plant and animal species, the same goes for the distribution of languages. Out of *Ethnologue's* 6,909, for instance, only 234 are spoken in Europe, while 2,322 are spoken in Asia. Another 2,110 are spoken in Africa, 993 in the Americas, and 1,250 in the Pacific.

One area of particularly high linguistic diversity is Papua New Guinea, where there are an estimated 832 languages spoken by a population of around 3.9 million. That makes the average number of speakers of each language around 4,500, possibly the lowest of any area of the world. These languages belong to between 40 and 50 distinct families: many of these are classified as part of the large (and not uncontroversial) 'Trans-New Guinea Phylum' (see [Figure 2](#)). Of course, the number of families and their grouping into larger taxonomic units may change as scholarship improves, but there is little reason to believe that these figures are radically off the mark.



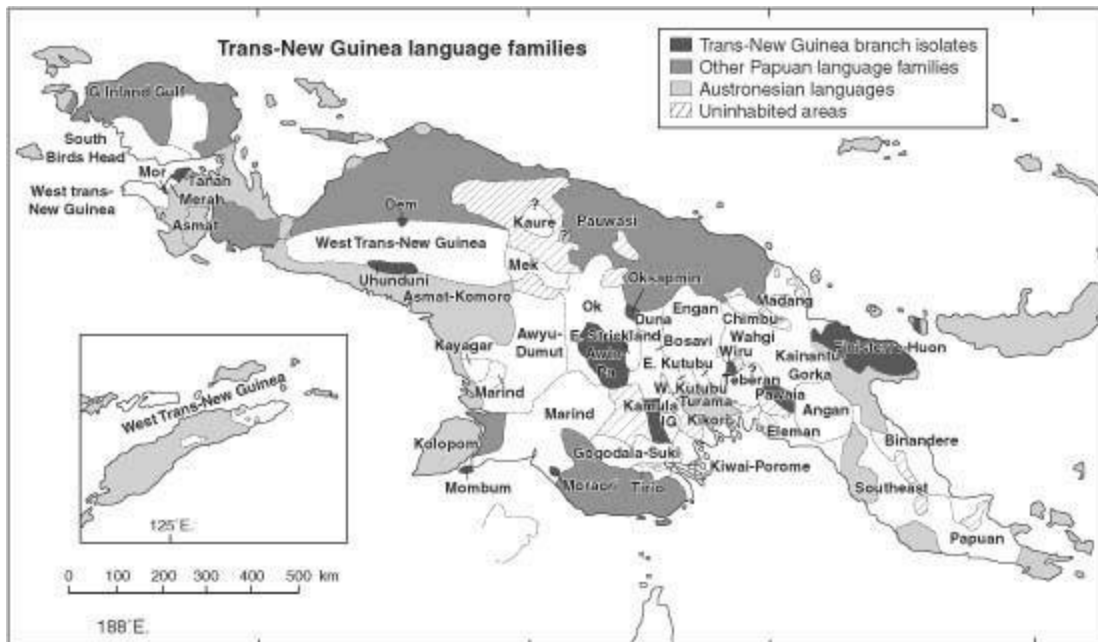
**1. Countries with a majority of speakers of Indo-European languages (dark grey) or where an Indo-European minority language has official status (light grey)**

The density of language diversity in this area is quite remarkable. In a small section of the north coast of the island of Papua New Guinea, shown in more detail in [Figure 3](#), there are some 91 distinct languages spoken, from 8 different families plus some isolates – nearly one-quarter as many as in the entire Indo-European family.

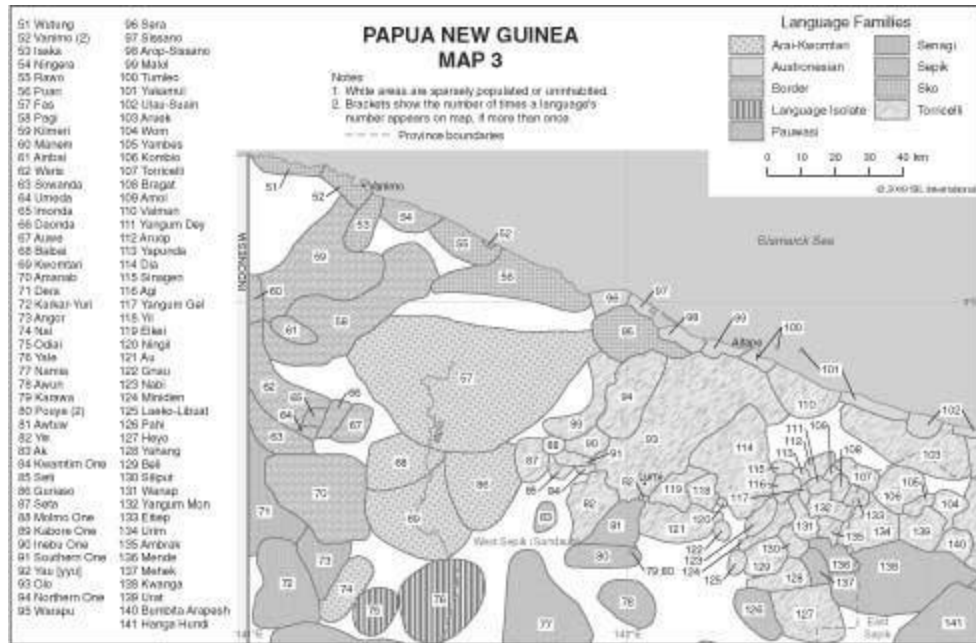
We do not find linguistic diversity only in out of the way places. Centuries of French governments have striven to make that country linguistically uniform, but with less success than we might think. Just as France is quite diverse in many famous aspects of its culture ([Figure 4](#)), it is also diverse in its languages ([Figure 5](#)). Even disregarding Breton (a Celtic language), the Germanic language spoken in Alsace, and the language isolate Basque, the 1999 edition of *Ethnologue* identified at least 10 distinct Romance languages spoken in France, including Picard, Gascon, Provençal, and several others in addition to ‘French’. Despite a change in the 2009 *Ethnologue* that shows just ‘French’ for most of the country, I want to stress that these are not just local ‘accents’. They are fully distinct speech forms that are as different from ‘French’ in at least some cases as, for instance, Spanish is from Portuguese.

The reader may well be wondering at this point just what I mean when I say these are different languages, and where these numbers come from. In fact, the point of this book is to discuss the complexity of those questions.

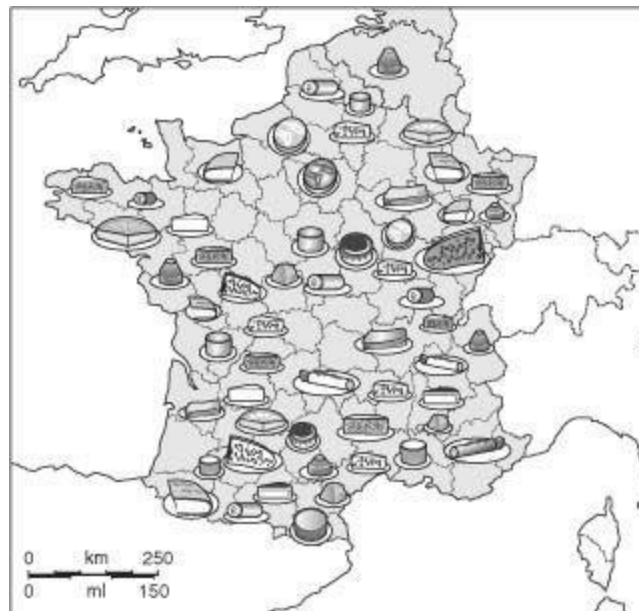
Multilingualism in North America is usually discussed (apart from the status of French in Canada) in terms of English versus Spanish or the languages of immigrant populations such as Cantonese or Khmer, but we should remember that the Americas made up a region with many languages well before modern Europeans or Asians arrived. In pre-contact times, over 300 languages were spoken in North America. Of these, about half have died out completely, and all we know of them comes from early word lists or limited grammatical and textual records. But that still leaves about 175 of North America's indigenous languages that are spoken at least to some extent today.



## 2. Languages of New Guinea, highlighting those of the broad Trans-New Guinea Phylum



### 3. Languages of a small part of Papua New Guinea

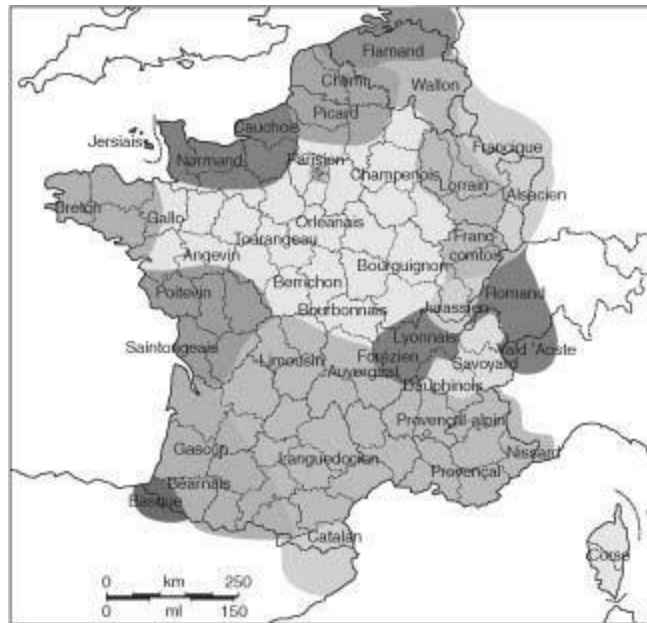


### 4. 'How is it possible to govern a country with 246 cheeses?' – Charles de Gaulle

As with places like Papua New Guinea, some areas of North America display much more diversity than others. The northwest coast, including the Seattle area, is probably the richest, including members of at least 9

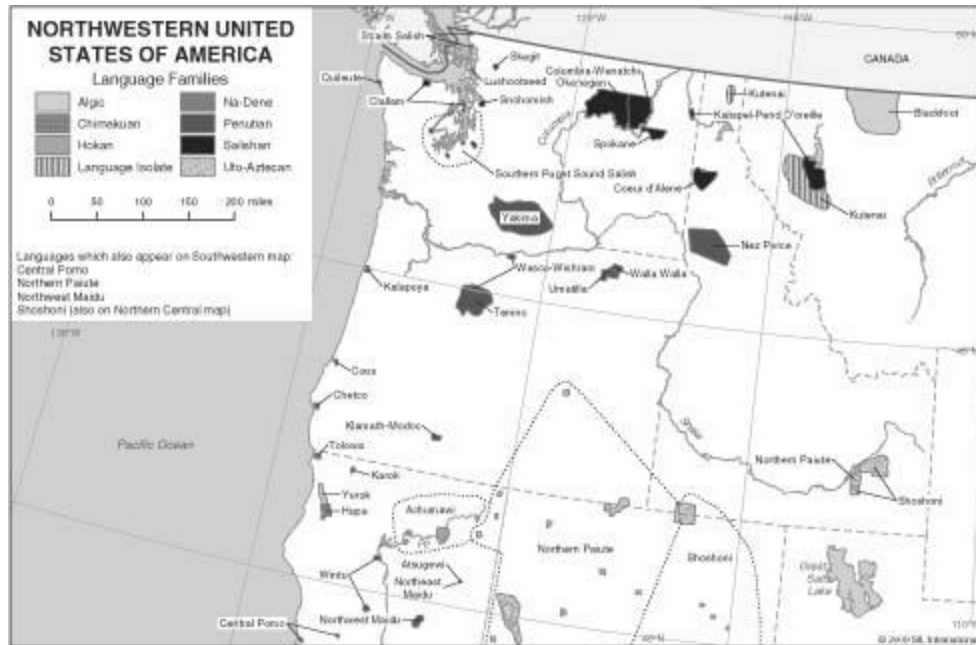


distinct families as well as a few isolates ([Figure 6](#)). Other areas of the Americas with substantial numbers of languages include the southwest USA and much of Central and South America. Guatemala, for example, has more than 50 indigenous languages that are still spoken in addition to Spanish, and Brazil more than 175.



## 5. Languages of France

Of the native languages still spoken in the USA, however, few can be said to be secure. Only 8, for instance, have as many as 10,000 speakers, and even among those, not all are still being actively learned by young children as living languages. Another 25, with between 1,000 and 10,000 speakers, are even worse off, and the remaining 142, with between a small handful of elderly people and about 1,000 speakers, are well on the way to extinction. As we will discuss in [Chapter 4](#), linguistic moribundity can be equated rather precisely with the situation of a language's no longer being learned by children.



## 6. Languages of the northwestern USA

In North America, even the most robust of native American languages (Navajo, with somewhat fewer than 150,000 speakers, and being increasingly replaced by English) is dwarfed by the number of native speakers of other non-indigenous languages. Hungarian, for instance, with something over 400,000 native speakers, Tagalog (over 375,000), and Danish (nearly 200,000) all benefit from their connections with larger linguistic communities elsewhere, a factor that contributes to their long-term stability. The native languages of the Americas lack this support, and in part, as a result, whatever the original diversity of the aboriginal languages of the region may have been, it is now decreasing at a remarkable rate.

## Other ways of quantifying languages

The differences in language distribution are not only geographical, of course. A comparatively small number of languages (389, or 6%) with over a million speakers account for most (94%) of the world's speakers, and in fact a tiny number with over 100,000,000 account for the bulk of that figure. The largest language cited by *Ethnologue* is Chinese, with

1,213,000,000 speakers; as we will see in [Chapter 5](#), the notion of ‘Chinese’ as a single language is quite misleading, but even confining attention to Mandarin (itself a somewhat diverse notion), the figure of 845,000,000 speakers puts this language in first place. A similar problem exists for ‘Arabic’, whose 221,000,000 reported speakers actually speak a diverse range of related languages (known somewhat misleadingly as Arabic ‘dialects’). Spanish, English, Hindi, Bengali, Portuguese, Russian, and Japanese round out the ‘over 100 million’ club, and together with Chinese and Arabic constitute the great majority of speakers overall. The 6,520 languages with under a million speakers (94%) are collectively spoken by only about 6% of the world’s population.

**Table 1. Families with over 200 languages (data from *Ethnologue*, 2009)**

Family	Languages	Speakers
Niger-Congo	1,510	382,257,169
Austronesian	1,231	353,585,905
Trans-New Guinea	475	3,334,267
Sino-Tibetan	445	1,259,227,250
Indo-European	426	2,721,969,619
Afro-Asiatic	353	359,495,289
<i>Totals:</i>	4,440	5,079,869,499

Stepping up one level of abstraction, a small number of families (six) shown in [Table 1](#) account for about two-thirds of the languages in *Ethnologue*’s catalogue, and the languages of over 85% of the world’s population. Languages outside this set, roughly 2,500, belong to one or another of several hundred families, or are not known to be related to any other languages in the historical terms to be described in [Chapter 3](#).

It is often quite hard to be precise about these relationships, because most of the relevant languages have not been described and studied in sufficient detail for us to be sure about their ‘genetic’ connections, but it is safe to say that the tools available for establishing such relationships on a scientific basis are unlikely to produce major reductions in the hundreds of apparently independent families. Of course, many would *like* to believe in more connections than can be demonstrated by existing methods, and attempts to promote such ‘long-distance comparisons’ are

common. If one accepts, as many of those interested in the evolution of the human language faculty do, that (spoken) language was invented only once in human history, then all currently spoken languages must be related, but that by itself does not mean we can recover the specifics of the relationships involved. The only methods that actually demonstrate with some certainty that languages are related have a necessarily shallow horizon of at most around 10,000 years, and since the hypothesized common ancestor must have diverged into numerous descendants much longer ago than that, it is unlikely that we will ever be able to establish such a unity as a genuinely scientific fact.

The discussion thus far has been confined to spoken languages, but we now know that these by no means exhaust the variety of ways in which the human language faculty manifests itself. Signed languages (such as American Sign Language, or ASL) have all of the properties of the spoken ones except for modality (manual/visual versus oral/aural), although they are completely independent of these apart from limited forms of lexical borrowing. Reliable estimates of the number of signed languages in the world are even harder to get than corresponding figures for spoken languages, but it is safe to say their number is far from insignificant, and new ones arise constantly as communities of the deaf in a given part of the world reach a certain critical mass. An appreciation of the spectrum of the world's languages must thus take these into account as well, as we will discuss in [Chapter 7](#).

Whatever the degree of the world's linguistic diversity, one thing that is fairly certain is that a surprising proportion of the world's languages are in fact disappearing – even as we speak. We will address this in [Chapter 4](#).

## **Chapter 3**

# **Phylogenetic linguistics: establishing linguistic relationships**

One way of defining a species, as mentioned in [Chapter 1](#), is as the smallest group of populations of organisms that share a common ancestor, such that there is no unique derived character that distinguishes one sub-group from another. This kind of approach is fundamentally historical, in that the sharing of a common historical antecedent and the absence of genetic innovations affecting only some of those involved constitute the essential properties that define a group of living organisms as a single species.

The same sort of approach applied to languages might take a ‘language’ to be the set of historical descendants of a single ancestral linguistic system, to the extent no distinctive changes have taken place that would distinguish one sub-group of speakers from another. This historical approach has certain advantages: for example, we might want to say that ‘English’ has been spoken in England continuously for a thousand years and more, but the speakers of today’s ‘English’ in, say, London could certainly not understand (or be understood by) a speaker of ‘English’ from Chaucer’s time, or perhaps even Shakespeare’s. The identity of ‘English’ has an historical dimension that goes beyond other notions of what constitutes the ‘same’ language.

The next step up in abstraction, corresponding to the grouping of species into genera, orders, phyla, and so on (more generally, in current usage, into higher-order clades) is to extend that historical relationship to groups of individual languages that do share a common ancestor, but which have undergone changes that make them distinct from one another. This leads to the grouping of languages into families on the

basis of their histories. Darwin's *Descent of Man* once more anticipates this analysis: 'Languages, like organic beings, can be classed in groups under groups; and they can be classed either naturally according to descent, or artificially by other characters.' This was a plausible enough position for a time (1871) when historical analyses of the facts of individual languages and their relations to one another were achieving remarkable results and were generally accepted as the only genuinely scientific approach to language, and it is still worth reflecting on today.

## Linguistic change

The topic of the present chapter is the basis on which we establish such relations. In that connection, a central notion is that of linguistic change as an inevitable aspect of history. To get a sense of this, consider the following thought experiment. When we consider the world's linguistic diversity, regardless of whether *Ethnologue's* 6,909 is the right number – or even a meaningful one – we know that there are lots of different languages in the world. Suppose, as some would desire, that that variability were to be magically eliminated in an instant. Imagine, that is, that by a stroke of linguistic enchantment, everyone in the world suddenly spoke exactly the same form of some language, say, Esperanto. Ignore the question of whether this would be a good or a bad thing, whether there would be knowledge lost, and so on. Just imagine the situation.

What do you think would happen? In fact, it is pretty certain that within, say, 20 or 30 years, local differences would have emerged, so that at minimum there were local accents. Given that the world changes, and that it changes in different ways in different cultures, there would be new words and expressions limited to particular places, but there would also be somewhat different local forms of speech. Within, let us say, 50 years, there would be clear local dialects, and quite certainly by the end of a century some of these would be so different from one another that speakers from some places would have considerable difficulty in understanding the speech of those from some other places. Depending on how those differences lined up with social and political realities, it would soon become common to speak of them as different languages.

Carrying out this scenario far enough, we arrive inexorably at a point where the differences among these languages are at least as great as those among modern Romance languages such as Spanish, Portuguese, French, Italian, Romanian, and others. But all of those languages would nonetheless represent linear historical developments from a single original form of speech, Esperanto in our thought experiment. And this is what linguists mean when they speak of ‘genetic’ relations among languages: that the languages concerned all represent divergent developments of a single original common ‘ancestor’.

Notice that this sense of genetic relationship has nothing to do with biology, in the literal sense. For instance, there is some controversy about the origin of Hungarian. On strictly linguistic grounds, Hungarian is a Uralic language, related to Finnish and Estonian. But genetic studies have shown that there is little or no distinctive biological connection between Finns and Hungarians, and so a variety of other origins have been proposed. What seems plausibly to have happened is that a Ugric tribe emigrated from central Asia into the area of modern Hungary, and conquered the local population, who subsequently adopted the language of their conquerors. Linguistically, then, Hungarian is ‘genetically’ related to Finnish and Estonian, although the biological affiliations of most of its speakers are somewhat different.

A variation on this situation arose when a small Viking (or Scandinavian) population conquered parts of modern France, and then adopted the local form of Old French, giving rise to ‘Norman’ French. This, in turn, was then imported into England, and greatly affected the development of English, although not to the extent of being completely adopted. Perhaps ironically, the Germanic speech of the residents of England was thus altered by the (Romance, not Germanic) linguistic practice of the invaders, whose biological relation to the English was in many cases somewhat closer than the linguistic one.

Returning to the point that is at stake, linguistic unity inevitably gives rise to diversity as soon as there are any dividing lines within the population that can serve to limit its homogeneity. The most obvious sources of such internal boundaries derive from geographical isolation:

if rivers, mountains, deserts, and so on separate some speakers from others, subsequent changes that take place in one region may not be reflected in others and *vice versa*. And this is quite enough to result in the establishment of distinct forms of speech, and eventually distinct languages.

But why should languages change? Why shouldn't the speakers in our hypothetical scenario just go on speaking the sort of Esperanto that they were endowed with in their magic moment? The fact is that, even ignoring things like vocabulary change due to technological and social development, language is constantly changing. There is a variety of reasons for this, and the search for a comprehensive account is the business of historical linguistics.

Two major sources of change are: (a) the tendency of speakers to alter their pronunciation in ways that simplify articulation; and (b) the possibility that listeners will interpret what they hear in a different way from what was intended, and alter their own linguistic practices accordingly.

Many instances of the first sort fall under the general category of assimilation. For instance, in Latin *octo* the two consonants are articulated at different places: the first is velar, with the tongue contacting the back of the hard palate, and the second is dental, with the tongue contacting the upper teeth. That means the tongue has to do two different things in rapid succession, a situation that could be simplified if both were produced at the same point – as in Italian *otto*, where the velar consonant is replaced by a somewhat longer dental occlusion. If enough speakers do this often enough, some learners will presume that that is the way the word really should be pronounced, and so *octo* will be replaced by *otto*, as of course happened in the history of Italian. Similarly, Swedish *drikka* 'to drink' and *takka* 'to thank' represent assimilation of a nasal consonant followed immediately by a stop (as we see in the *nk* of English *drink*, *thank*) to produce just a longer stop consonant with no nasal part.

We can observe assimilation happening before our ears in modern English when someone says [fʊbbɔl] for *football*. The square brackets



here and elsewhere set off a phonetic representation, similar to the pronunciation guides in dictionaries, as opposed to the usual orthography. The values of the special symbols here should be apparent: [U] is just the vowel in *foot* and [ɔ] that in *ball*.

There are many other conceptually similar sub-types of change that are motivated by articulatory considerations, but other changes result from misperceptions. One such type is that of mishearings due to limited acoustic distinctiveness. For instance, nasal consonants (like [m, n]) before a stop consonant (like [p, t]) are harder to distinguish there than they are in some other positions, such as at the beginning of a word. A hearer may thus misinterpret the sound produced by another speaker, hearing for instance [grænpa] as [græmpa] (with [æ] representing the vowel in the author's pronunciation of *grand*). Of course, in this case, the speaker may actually have assimilated the nasal to (the labial articulation of) the stop, but even if this has not happened, the hearer has very little evidence, apart from prior knowledge of the word, that the nasal consonant in the word was really an [n] and not an [m].

Another source of misperception is genuine ambiguity in the speech signal. Thus, in some languages, a vowel next to a voiced aspirated consonant [bh, dh, gh] has a special 'breathy' quality, regardless of whether the [bh] etc. precedes or follows within the syllable; and in fact this breathiness may be the main acoustic evidence for the aspirated nature of the consonant. Thus, a 'breathy' [a] between [b] and [d] could signal any of *bhad*, *badh*, or *bhadh*. If speakers tend to assume as a default that the breathiness comes from the preceding consonant, for instance, then original *badh* will be interpreted as (and perhaps eventually replaced by) *bhad*.

There is always a certain amount of variation in the pronunciation of nearly every word, and listeners must abstract away from this variation to arrive at a constant identity for a word. When they abstract in the wrong way, this can result in a change. For instance, the sequence of consonants in the word *athlete* is moderately difficult, and some speakers sometimes pronounce the word with a short vowel between them, as *athelete*. A listener not knowing the word, and hearing

sometimes *athlete* and sometimes *athelete* might conclude that the word was really *athelete*, and that the pronunciations without the vowel in the middle resulted from dropping it. Such short vowels in the middle of words do often go unpronounced, as when we say *famly* for *family*, so infact their occasional presence gives little clear evidence for whether they are in fact an organic part of the word. Either inference is defensible – although making the wrong one in a particular case may lead to a change.

Restructurings and misinterpretations of these sorts are happening all the time all around us, and the important point is that any particular change is not inevitable or completely predictable. That is, we can say a certain amount about what sort of change *might* happen, but this is not in general limited to a single, unique alteration of the original; and more importantly, we cannot say that a given possible change *will* happen. As a result, the changes that affect the same basic system in different communities are likely to be different, and to lead to differentiation of the speech of those communities. The changes that do get incorporated in the language, though, are institutionalized with each new generation. That is, children learning the language do so on the basis of the data available to them, and their interpretations of those data characterize the new base state of the language, a state that can itself undergo further changes, and so on. As these changes cumulate, the languages of different communities diverge more and more, as when the original spoken Latin of Rome developed into the local Romance languages from a more or less common starting point.

## **The regularity of linguistic change**

The kinds of change just discussed seem like developments that ought to affect individual words one by one, but in the late 19th century, it was observed that this is not the usual case. Generally, that is, when we find that a change affects a given sound in a given position in some words, it has the same effect everywhere. When some final voiced stops (like [b, d, g]) were replaced by voiceless ones ([p, t, k]) at an early stage in the development of German (perhaps because of the difficulty of maintaining vibration of the vocal cords at the end of the word), this

affected not just a few words, but every word where the relevant conditions were met. A group of historical linguists in Germany in the 1860s noticed this, and proposed that it was a general principle: the regularity of sound change. What changes, on this view, is not basically the individual word but the very sound *type*. Another group, the ‘Neolinguisti’, pursued the alternative view that ‘every word has its own history’, but this approach soon bogged down in empty description and was largely abandoned.

A particularly important role was played in this theoretical development by a set of changes in the history of the Germanic languages known as Grimm’s Law (after one of the brothers Grimm, who was an historical linguist as well as a collector of scary folk tales). Consider the comparisons in [Table 2](#).

**Table 2. Developments and similarities in words between languages**

Greek	Latin	Sanskrit	Gothic	Old English	
patēr	pater	pitā	fadar	fieder	'father'
treīs	trēs	trayas	þreis	þrī	'three'
(he-)katón	[k]entum	śatam	hund	hund	'hundred'
déka	decem	daśa	taihun	tēon	'ten'
geúomai	gustus	jōs	kīusan	cēosan	'taste, - test, choose'
phérō	ferō	bharāmi	baira	beoru	'I carry'
(e-)thē-ka	fē-c-i	(a-)dhā-m	(ga-)dē-þ-s	dæd	'put/do; deed'
kheúō	fu-n-d-ō	ho-tar	giutan	gēotan	'pour'

There are many things going on in these forms, but the point to notice is that the first three words here are presumed originally to have had initial voiceless stops [p, t, k], the next two voiced stops [d, g], and the last three voiced aspirates [bh, dh, gh]. We make that assumption on the basis of the preponderance of the evidence from all of the languages taken together, not from the presumed ‘ancient’ status of one or another individual language. What happened in the history of Germanic seems to have been a combination of the developments:

- voiceless stops became voiceless fricatives (e.g., [p]>[f]);
- voiced stops became voiceless (e.g. [b]>[p]); and
- voiced aspirates became simple voiced stops (e.g. [bh]>[b]).

This set of developments, known as Grimm’s Law, appeared to be a

very general principle, and accounted for a wide variety of words in the way their form in Germanic differs from that in the other Indo-European languages. A number of apparent exceptions were promptly incorporated into this story by refinements of the law. But some others remained recalcitrant, even in forms of the same word, as illustrated in the Old English forms in Table (3), where the corresponding Sanskrit forms show that the sound realized either as *þ* or as *d* was originally a *t*.

**Table 3. Sanskrit and Old English word forms compared**

Sanskrit	Old English	
vártate	weorþan	'turn, become (present)'
vavárta	wearþ	(past singular)
vavrtúr	wurdon	(past plural)
vrtaná-	(ge)worden	(past participle)

Why should original *t* turn into *þ* as expected in some forms of this word, but into *d* in others? It looks as if this sound change is not regular, in the sense of applying to all words, but rather applies to some words and not others.

The linguist Karl Verner provided an explanation: the development of *t* into *d* instead of *þ* takes place exactly when all of the following conditions are met: (a) the *t* is not initial; (b) it is surrounded by voiced sounds (including vowels, nasals, *r*, *l*); and (c) the original accent was not on the preceding syllable. The reference to the original accent was the key to Verner's solution; its location in the words in [Table 3](#) is another thing that is shown by the Sanskrit forms.

Verner's resolution of this apparent problem with Grimm's Law sounds like a pretty minor detail, but it was enormously important at the time. In eliminating the apparent capriciousness of the changes in Grimm's Law, it held out the promise that sound change would *always* turn out to be regular, with apparent exceptions due to the operation of different, independent changes.

This principle of the exceptionless nature of sound changes became the foundation of the historical study of languages. It was reinforced in later years, when the American linguist Leonard Bloomfield provided a dramatic demonstration that it was applicable not only in well-

established languages with literary histories like those of the Indo-European family, but also in unwritten languages like those of the Algonquian family. More recently, much the same thing has been shown for the Aboriginal languages that constitute the large Pama-Nyungan family in Australia, which had been claimed to have histories that made application of the comparative technique illustrated here impossible.

The whole notion of exceptionless sound laws that have operated at some time in the past in the development of modern languages seems somewhat mystical, but more substance was given to it in the 1970s when William Labov studied variation among speakers in American cities. Basically, Labov discovered considerable variation that correlated with a number of social variables, and not just with geography. He also found variation that correlated with age, in clear and elegant ways. What he demonstrated in essence was that it is possible to observe sound change in progress, by tracking a single variable of pronunciation across generations. Taken together with a greater appreciation of the phonetic and perceptual factors underlying sound change, this made the overall picture more plausible.

## **Language change and linguistic relationship**

The relevance of all this to our concerns is two-fold. On the one hand, the occurrence of sound changes, regular in retrospect but unpredictable in their incidence, is one of the factors (along with change in other areas of language, of course, such as word formation, syntax, semantics, and vocabulary) that produces linguistic diversity out of original unity. On the other hand, though, it also provides us with a way of assessing the likelihood that a given set of diverse languages are actually historically related, in the sense of having developed out of an originally unitary state.

The argument here goes as follows: if a set of diverse languages all constitute divergent developments from a common ancestor, it ought to be the case that each of them represents, at least in substantial part, that ancestor as it was affected by some sequence of regular changes. This is similar to the way in which we can say that distinct species descended

from a common ancestor are characterized by sequences of genetic mutations that occurred in the histories of some but not the others. Of course, some of the words of the original language may have been lost, or replaced by others through borrowing, but to the extent we can control for these factors (or at least bear them in mind), the regular nature of linguistic change allows us to treat each observed language as a systematic deformation of the original state, resulting from a specific set of systematic processes that have produced it. And that in turn means that each of the different languages derived from the same ancestor would represent a different transformation of that original core.

On that basis, the difference (*modulo* changes in areas other than sound) between corresponding words in a set of different languages derived from the same ancestor ought to be systematic too: the difference between the outcomes of one set of systematic transformations and another. And that is the basis of the fundamental technique in language comparison, the *comparative method*.

Basically, when you want to compare a pair (or larger set) of languages to see if they are related, what you do is collect large samples of words from each and look for systematic similarities. When we compare Latin and English, for example, we find a fair number of words that differ primarily in that the Latin word has *p*, *t*, *c* (where *c* = [k]), while the English word has *f*, *th*, *h*. On that basis, we can propose that these words all descend from earlier forms with the same sound in that position, [p, t] or [k]. Note that we do that not because Latin is a more ‘ancient’ language than English, but because the evidence from a variety of languages both ancient and modern (such as Greek, Sanskrit, Russian) points collectively in the direction of this interpretation of the sounds involved.

We thus posit something like a [p] as the initial sound in the ancestral word for ‘father’, for instance, and write this reconstructed sound (of whose exact phonetic value we cannot of course be certain) as \*p. In historical linguistics, an asterisk preceding a sound or a form indicates that this is reconstructed from comparative evidence, rather than something we can observe directly. In other areas of linguistics, the

asterisk is often used in a different way, to indicate that a given sequence of words is not a grammatical sentence (or phrase, etc.) in the language under discussion. Asterisks will appear in that function in the syntactic discussions in [Chapter 6](#), for example.

We can now say that Latin preserves \*p as [p] (in *pater*), while English has changed \*p into [f] in the corresponding word *father*. The more we can explain the forms of a substantial number of words in the observed languages on the basis of proposed original forms together with distinct sets of changes in each, the more evidence we have in favour of the proposition that they are in fact related.

The logic of this argument is based on what Ferdinand de Saussure and subsequent linguists have called the ‘arbitrariness of the sign’. For all words except a small set of onomatopoeic items (words for the noises made by animals, for example), the form of a word is quite independent of its meaning. A ‘tree’ is called a *tree* in English, but *puu* in Finnish, without any necessary difference in the trees these words refer to. When we find that different languages have quite different words for more or less the same thing, that is what we expect; but in the opposite case, where we find that words with the same (or very similar) meanings are systematically related in form, that calls for an explanation.

One possibility, of course, is that one of the languages has simply borrowed the word from the other (or that both have borrowed it from a common third language). The primary alternative, and the possibility that interests us here, is that both languages are descended from a common ancestor, and that individual words in that ancestral language have followed distinct paths of change in the development of that ancestor into the two languages being compared. The more words for which we can exclude the possibility of borrowing, and which exhibit the same regular relations between the languages, the more plausible it is to claim that the resemblances are due to an historical relationship. And in fact, this is the *only* scientific way to demonstrate such a relationship between languages whose external history is no longer available for inspection.

Essential to the integrity of this argument is the systematic nature of

the correspondences demonstrated. What gives plausibility to the claim that languages A and B are derived from a common ancestor is the evidence that substantial parts of each can be derived from the same forms, forms attributed to their common ancestor, by a series of presumptively regular changes. We are unlikely ever to achieve absolute certainty in these matters, since there are other things that intervene to obscure the situation, but we can often obtain overwhelming support. That is the case with most of the well-known families whose existence is now generally presumed: Indo-European and its sub-families like Germanic, Italic, Slavic, and so on; Austronesian; Uralic; and a number of others. Some remain controversial, precisely because the amount of hard evidence of this kind is small. This is the case for the proposed 'Altaic' family, for instance, a grouping that would unite the languages of the Turkic, Mongolian, and Tungusic families (this last spoken in Eastern Siberia and Manchuria). Each of these individual groupings is clearly a unit, but their relation to each other is much less clear, and their derivation from a single ancestral language cannot be regarded as firmly established.

I want to stress again that the application of the comparative method is really the only scientific tool we have for establishing relatedness among languages. Genetics in the directly biological sense would not suffice, even if it provided the right sort of resolution, as cases like those of Hungarian and Norman French demonstrate. Other sorts of comparison, such as looking for similar rules of syntax or word form, also do not provide secure evidence, because these can arise easily in diverse languages even in the absence of a common ancestor. The existence of an 'Altaic' family, for instance, cannot be established on the basis of observations such as that all three of its constituent families primarily use suffixes rather than prefixes, and that each makes use of vowel harmony (a principle requiring agreement in certain properties among all of the vowels in a word). These broad structural parallels really do not contribute much to establishing a common origin, as opposed to systematic relations in the (otherwise arbitrary) forms of words.

While the comparative method is a powerful tool for demonstrating



historically based linguistic relationships, there are many ways to go wrong in attempting to apply it. Scholars must make sure they are dealing with genuinely inherited words, for instance, and not with borrowings. Words must be excluded whose form is not really arbitrary: nursery words, onomatopoeic forms, and a few other types tend to be similar across languages for reasons having nothing to do with common inheritance.

A particularly common class of words of this sort is the set of terms for ‘mother’ and ‘father’, words that arise very early in child speech and whose form can be viewed as determined in part by the limited range of the child’s very early phonetic competence. While such words as *mama* and *daddy* may recur in many unrelated languages, they are not completely uniform. A small subset of languages with no historical connection among them amusingly have the meanings of some words of this set interchanged. In Georgian, for instance, *mama* means ‘father’ and *deda* means ‘mother’. Regardless of this, the discovery that two languages being compared both have words like *mama* that mean ‘mother’ (or ‘father’!) does not really contribute to a demonstration that they are related.

It may seem obvious that it is necessary to control the languages in question well enough to know what to compare. Many erroneous proposals, however, can be shown to be misconceived because the scholar who suggested them was misled into comparing apples with oranges, as it were, through lack of sufficient understanding of the systems being examined.

It is also important to be on the lookout for accidental similarities: thus, in the (now extinct) Australian language Mbabaram, the word [dɔg] meant ‘dog’, for reasons having nothing to do with any proposed common ancestry with English or even borrowing. Mbabaram [dɔg] was simply the result of regular sound changes in the language as they affected an original form *\*gudaga*.

Even when everything is done right, comparison may not yield a determinate answer. Words are lost, or replaced, or rebuilt, in the history of a language, and over time the amount of common vocabulary shared

(in altered form) by two descendants of the same language will inevitably decline. Current estimates are that any languages that diverged from a common ancestor more than about 8,000 years ago will have changed so much that any systematic similarities that existed will have been lost in the noise of chance. The exact time limit is hard to establish, but it is clear that this is much nearer the present than it is to the origin of language some 100,000 or more years ago. Thus even if all modern languages do derive from a single ‘proto-world’, we will probably never be able to establish that as a scientific fact.

Language identity has an historical dimension, not only in terms of the relationship among languages, but also in the continuity of what counts as a particular language over time. Exploring the extent and depth of linguistic families tells us something about the world’s linguistic diversity, even if it does not seem very helpful in the enterprise of quantifying it. Whatever the histories of many languages may be, however, the conditions of the modern world suggest that these are about to come to a somewhat abrupt end. That impending mass extinction is the topic of [Chapter 4](#).

## **Chapter 4**

### **The future of languages**

Whatever the world's linguistic diversity at the present, it is steadily declining, as local forms of speech increasingly retreat before the advance of the major languages of world civilization. This is an issue which, when properly understood, is surprisingly similar to that posed by the extinction of biological species.

### **Endangered diversity**

Practically everyone in North America has heard of the Northern Spotted Owl (*Strix occidentalis caurina*), largely because of the massive controversy surrounding efforts to preserve this species and the resulting conflict with the timber industry in the northwest. The owl's traditional range is old growth forests from southwestern British Columbia in Canada through Washington and Oregon south to Marin County in coastal northern California. It is threatened by the progressive loss of its habitat in this region to lumbering, and also by competition with the Barred Owl. It has been officially listed as a 'Threatened' species since 1990 in the US, and in Canada as 'Endangered' since 2002. There are estimated to be around 3,000 breeding pairs remaining in the US, and fewer than 100 in Canada. There is little doubt that the Northern Spotted Owl is seriously in decline, and that this decline serves as an indicator of the threat to the entire ecosystem represented by the old growth forests of the northwest. As such, measures to resist the threat to this species, and to preserve its habitat, are undoubtedly warranted.



## 7. Northern Spotted Owl (*Strix occidentalis caurina*)

Given the concern that has been expressed about the Northern Spotted Owl and the extent to which the general public has become aware of its plight, it is worthwhile to compare this situation with another sort of endangerment. Within the historical range of the Northern Spotted Owl, a few more than two dozen of the indigenous languages of North America are still currently spoken, representing a half dozen or so distinct families and some isolates. A list of these, with figures on the number of current speakers drawn primarily from *Ethnologue*, is given in [Table 4](#).

As will be seen, the total number of remaining speakers of these languages is perhaps a little over 1,600 (and these figures are undoubtedly optimistic) – all of them combined represented by a population of speakers equal to just about one quarter of the number of remaining Northern Spotted Owls.

The northwest coast of North America is historically a region of great linguistic (and cultural) diversity. A number of languages once spoken in the same region as those in [Table 4](#) are not listed, because there are no remaining speakers; they are preserved only in the records, texts, and descriptions of earlier generations of explorers and anthropologists. And indeed, there is little reason to doubt that this diversity will be much further reduced in the very near future, given the small number of speakers of many of the remaining languages and the fact that, in many cases, all of these are quite elderly.

**Table 4. Indigenous languages still spoken within the historical range of the Northern Spotted Owl**

Language	Speakers	Language	Speakers
Achumawi	16	Chetco	5
Clallam	10	Coos	1
Halkomelem	225	Hupa	8
Kalapuya	1	Karok	10
Klamath-Modoc	1	Lillooet	200
Lushootseed	60	Makah	0 <sup>a</sup>
Pomo (Central)	4	Quileut	10
Sechelt	40	Skagit	100
Snohomish	10	Southern Puget Sound Salish	110
Squamish	15	Straits Salish	20
Tenino	200	Thompson	500
Tolowa	4	Wasco-Wishram	69
Wintu	5	Yurok	12

<sup>a</sup> The last fully fluent first-language speaker of Makah died in 2002, but some tribal members still have some command of the language and a tribal programme is trying to revive it as a second language.

The plight of the languages in this part of the world would thus seem rather more extreme than that of the Northern Spotted Owl, and yet the extent of public awareness and concern is vastly less. If indeed languages are in some sense analogous to biological species, and linguistic diversity comparable to biological diversity, the rate at which the world's languages are disappearing presents itself clearly as a serious issue that deserves greater attention than it usually attracts.

Just what do we mean when we describe a language or a species as 'endangered'? In the biological case, the continuation of a species depends on the existence of a sufficient pool of individuals to maintain successful reproduction into the future. We might add that this population must contain sufficient genetic diversity to avoid the destructive consequences of excessive inbreeding due (primarily) to the spread of deleterious recessive mutations, but in fact, it all comes down to a population sufficient to allow for continued reproduction and to avoid extinction.

The linguistic case is quite comparable, once we recognize the way in which languages 'reproduce'. This is by the transmission of the language from its current speakers to new generations, and the criterial point for a language's survival is thus the extent to which new generations are learning it. Since full competence in a language is best (or, some might say, only) acquired in childhood prior to adolescence,

what matters for the future is the extent to which young children are acquiring the language. Even if the entire population of all of the retirement communities in North America spoke fluent Potawatomi, if that were the only situation in which the language was used, we could confidently predict that it would be extinct within a few decades. A language whose only speakers are middle-aged or older, no matter how many of these there may be, is on the path to extinction unless revived among younger generations. Nearly all of the languages in [Table 4](#) are in that condition at present.

Some would say that the death of a language is much less conclusive than that of a species. After all, when a group abandons its native language, it is generally for another that is more economically advantageous to them: why should we question the wisdom of that choice? And in any case, are there not instances of languages that died and were reborn, like Hebrew? We will deal with the question of whether the death of a language really matters in the next section. But we must recognize that languages really do die when they cease to be spoken. The case of Hebrew is quite anomalous, since the language was not in fact abandoned over the many years when it was no longer the principal language of the Jewish people. During this time, it remained an object of intense study and analysis by scholars. And there are few, if any, comparable cases to support the notion that language death is reversible.

The situation in North America is not at all atypical. As noted in [Chapter 2](#), out of about 165 indigenous languages, only 8 are spoken by as many as 10,000 people. About 75 are spoken by only a handful of older people, and can be assumed to be on their way to extinction. While we might think this is an unusual fact about North America, due to the overwhelming pressure of European settlement over the past 500 years, it is actually close to the norm. Around one quarter of the world's languages have fewer than 1,000 remaining speakers. Linguists generally agree in estimating that within the next century, the extinction of at least several thousand of the 6,909 languages listed by *Ethnologue*, or nearly half, is virtually guaranteed under present circumstances.

Precise statistics on the number of endangered languages in the world are difficult to provide, in part because the degree of endangerment is not always easy to assess. Circumstances vary considerably from one to another and the likelihood of a language being lost depends heavily on the details of its speaker community. A number of categorizations are used to describe languages, usually ranging from something like ‘stable’ through ‘vulnerable’, ‘threatened’, and various degrees of ‘endangered’ through ‘extinct’. In addition to what can be learned about the transmission of a language across generations, other factors contributing to its vitality can include the absolute number of its speakers and their proportion within the community, the domains in which it is used, explicit government policies, and the attitude of community members towards its use.

Taking all of these factors into account, the United Nations Educational, Scientific, and Cultural Organization (UNESCO) maintains a continuously updated interactive atlas of endangered languages, which the reader can consult online. As of April 2011, this list identified 2,473 languages in the categories from ‘vulnerable’ (601) through ‘extinct’ (230). Assuming that the total against which this should be compared is something like *Ethnologue*’s 6,909, that means that at least 35% of the world’s languages are in serious trouble. And that still does not take into account the languages that are not sufficiently known to be assigned to one of UNESCO’s categories. These are surely the most conservative of figures: estimates among linguists of the proportion of languages currently spoken that will have disappeared a century from now range from around 50% to as high as 90%.

To put this in perspective, we can compare these numbers with a list of endangered biological species compiled by the International Union for the Conservation of Nature (IUCN). [Table 5](#) gives numbers from this organization’s ‘Red List’, which is widely considered reliable, indicating for several sorts of species the proportion found to be ‘vulnerable’, ‘endangered,’ or ‘extinct’.

**Table 5. Threatened biological species (2010)**

	Number of evaluated species	Number of threatened species	Proportion of evaluated threatened
Mammals	5,491	1,209	22%
Birds	10,027	1,376	14%
Reptiles	2,806	615	22%
Amphibians	6,296	1,937	31%
Bony fishes	7,777	1,731	22%
Insects	3,269	794	24%
Crustaceans	2,152	608	28%

In some instances (reptiles, insects, crustaceans, fish), the number of species that have been evaluated is only a fraction of the total that have been described, and presumably the process of evaluation has focused on species that are likely to be in danger. In such cases, the degree of endangerment reported is likely to be at least somewhat overstated.

The number of threatened species, both biological and linguistic, clearly represents a significant challenge to the diversity of our world. In the biological case, this has of course become a matter of major concern around the world. The proportion of the world's languages that are threatened with extinction is at least as great, however, and public attention is only beginning to be drawn to this dimension of the overall biodiversity problem faced by the modern world.

## Endangered knowledge

One might well ask (as many have) whether the prospect of the extinction of so many of the world's languages is something to be concerned about. After all, it could be said that if the Chickasaw people, for instance, abandon their use of that language and all come to speak English, the result is that they will be better enabled to participate fully in the culture of the surrounding community in the US, and better positioned for economic and political success, than they would be if they spoke only Chickasaw.

There are certainly good reasons for people everywhere to become competent in one or another of the major world languages of power and influence (or perhaps a less general but locally important language), but this way of seeing the situation represents a false dichotomy. As we will see later in this chapter, it is perfectly possible to maintain the language



of one's culture while also becoming competent in another language that provides access to wider opportunities. And the abandonment of a language can have real and unfortunate effects.

A posting to the Language Log blog in December 2009 began with the observation that '[i]t seems you can't swing a dead cat in a bookstore these days without hitting a recent book on language endangerment and language death'. Most of those books, including those mentioned in the Further reading section at the end of this book as well as others, present detailed studies of what is lost when specific languages disappear, and the reader whose attention is engaged by these issues is encouraged to consult that literature. The bases for being concerned about language loss, and for taking action to resist it, fall into several general categories, which can be indicated schematically.

## **Practical reasons**

Languages spoken by small and historically isolated groups may preserve knowledge of the natural world that can be of immense value. Traditional cultures that have been based in a particular location over a long period of time have generally developed an intimate acquaintance with resources of the region that are available to be exploited for particular purposes.

The most obvious category of this sort is familiarity with the curative and medicinal uses of various local plants and other substances that form part of the ethnobotanical knowledge of many cultures. Many of these plants are unknown to science outside the region, along with their potential pharmacological value. The loss of a traditional language generally entails loss of access to this knowledge base: when another language such as English, Spanish, or Mandarin, comes to be the only one spoken, there is literally no way to talk about the species uniquely identified in the local language.

A particularly extreme example in this regard is the Kallawaya language of Bolivia, whose situation is described briefly by K. David Harrison. The grammar of this language is essentially a form of

Quechua, but most of its vocabulary is either from unknown sources or from a language of a now otherwise extinct language family called Pukinan, languages once spoken in the region but now abandoned in favour of Quechua, Aymara, or Spanish. Kallawaya is the language of a group of traditional healers, going back at least to the retinue of Inka rulers in the 15th century. The language is used primarily for ritual purposes, but is able to serve a full range of conversational purposes among those who know it. Passed on from generation to generation among males through a sort of initiation, it is largely secret, and its use is confined to a small group. An immense portion of the local knowledge of healing practices and the resources (plant, animal, and others) that are essential to these is encoded in Kallawaya, and the loss of this language would entail the loss of that knowledge base.

Apart from preserving the specifics of largely practical matters such as the healing value of various botanicals, traditional languages often encode basic information about biodiversity that would be of considerable interest to science. Scientific description and classification has uncovered only a fraction of the species that there are in our world, and new ones are being reported constantly. In many cases, species that are 'new' to Western science are in fact familiarly distinguished by the people who live in the area where they are found, and identified as distinct in their languages. The loss of those languages would eliminate a significant resource that can aid our efforts to understand the full range of life on Earth.

## **Cultural and ethical reasons**

It is apparent that many languages spoken in areas that are especially rich in plant and animal species, and other natural resources, encode information about these aspects of the environment that can be of considerable utility and importance to a wider audience. Perhaps less obviously, these languages and many others also encode information that can be essential to the identity of the community of speakers itself, and to the sense of connection those speakers have to their background.

The majority of the world's languages have no standard written form.

In the case of many of those that do have such a system, very little of the literature, folklore, and other local tradition has actually been recorded in a way that would make this material accessible to members of the community. So long as their language is maintained, oral transmission can serve this purpose as it has for centuries, but as soon as the language is lost, an essential link to the group's history and cultural patrimony is cut.

All around the world, groups of people are reasserting their association with their traditional cultures. There is nothing mutually exclusive between this and their participation in a wider society: of course, some groups really do wish to withdraw into their past, but in most cases, they simply want to preserve their own sense of themselves while also connecting to a larger national or social identity. As long as their heritage is preserved in the form of their language and the resources it makes available to them, they retain at least the option of doing this. Once the language dies, access to everything it embodies dies with it, and future generations will no longer have this choice.

## **Scientific reasons**

Another class of reasons to be concerned about the imminent disappearance of a large fraction of the world's languages can be seen in the consequences of this trend for our potential understanding of the nature of language itself. Virtually every new language that is described holds some surprises for the linguist. If the Big Nambas language of Vanuatu (with about 3,350 current speakers, according to *Ethnologue*) had disappeared before linguists investigated it, we would still be unaware that languages can make use of sounds involving the articulation of the tongue against the upper lip. Indeed, prior to the description of this language, some linguists (including the present author) had asserted that a principle of the theory of sound structure had the consequence that such sounds could never be distinctive in any language. Our appreciation of the structural parameters of sound systems has been revised and improved in a way that would have been impossible without the evidence of this language.

Without the evidence of a small set of languages like Wichita (a Caddoan language of the American great plains, with only a single native speaker remaining at the time of writing, down from 5 in 2006), we would not know that it is possible for the verb of a sentence to agree not with its subject but with the possessor of the subject: thus, in Wichita one says the equivalent of *My horse am running, your horse are running*. In Movima, an isolated language of the Bolivian Amazon with about 1,400 speakers, verbs have no marking for tense, but nouns do, and the marking on a noun can determine the interpretation of a sentence so that the equivalent of *she kisses her late husband* is a way to say *she kissed her husband* (where the husband is still alive, although the act of kissing is past). A wealth of unexpected phenomena turn up when we look at new languages, phenomena that expand our conception of just what the grammars of languages can be like. It is unfortunately likely that many such illuminating phenomena will disappear with the languages that present them before linguists have a chance to document these possibilities.

Linguists are not the only investigators with an interest in preserving the full range of the world's languages. Traditionally, anthropologists have seen the language of a culture as a crucial key into it, important for understanding the limits of cultural diversity (always assuming that there still are scholars with this interest, in an age when social/cultural anthropology denies its identity as 'science' and seems more interested in problems of post-modernist criticism than in direct exploration of 'exotic' cultures). One aspect of this central role of language is the fact that most traditional knowledge, folklore, mythology, and so on is directly and uniquely represented in oral language.

For this reason, it was once a basic prerequisite for research in the field that a student learn something about linguistic 'field methods': the procedures by which one develops a basic command of a new language in the course of working directly with one or more (possibly monolingual) consultants in the community. Much of the activity of cultural investigation can consist of the elicitation, recording, and analysis of texts in the local language. Obviously, if that language disappears, the possibility of access to the culture it encoded is greatly

reduced, if not eliminated altogether.

It is also the case that the language may itself directly represent data of great interest to the cultural anthropologist. A familiar instance is the terminology used for various kinship relations. In English, for example, we use the same word *uncle* for a male sibling of either our father or our mother (and *aunt* for a corresponding female sibling of a parent of either sex). In the Australian language Kayardild, on the other hand, the word *kanthathu* refers to either my father or my father's brother, but not my mother's brother: he is my *kakuju*. This does not reflect any confusion on the part of the Kayardild as to who their fathers are, but rather the fact that, in this culture, one's father and his brother(s) would share the responsibility for conveying certain skills, a responsibility not shared by one's mother's brother(s). This is all based on a kinship system which is derived from categories called *moieties* that have no direct correspondence in the system of English speakers.

Kayardild itself has only a single fully fluent speaker at the time of writing, but the system underlying its kinship vocabulary is familiar to anthropologists from other societies. We do not need to go further into the nature of this system to see that the categorization by a language of the relatives that are considered in some way 'equivalent' can provide essential clues to the way the society as a whole organizes the lives of its members. When the language dies, however, these clues are no longer as available, and in fact without the support of language, the speakers themselves may lose track of distinctions that were once central to their community and its social life.

If we claim that it is important to preserve as many of the world's languages as possible so that science will have an opportunity to explore the range of linguistic and cultural features they embody, it might be objected that this is a rather self-centred way of looking at the problem. Why, after all, should people retain their language in order to provide a steady supply of dissertation topics and academic jobs for those who study them? If this were the only motivation for language preservation, and if, furthermore, the preservation of these languages necessarily entailed an exorbitant cost for those called on to continue speaking

them, this might be enough reason to abandon the attempt to preserve endangered languages.

The professional concerns of academics are not, however, quite as insulated as they might appear. Their goal, after all, is to understand the full range of possibilities instantiated by one particular species that is of special concern to us: *Homo sapiens*. If we are willing to argue that endangered species of animals and plants should be preserved so as to maintain the world's marvellous overall biodiversity, is it not just as worthwhile for a linguist or anthropologist to be concerned with the extent of the diversity of human language and culture as for an ornithologist to be concerned with the diversity of bird life?

As for the cost to speakers, when properly understood, the balance falls out quite differently. First of all, as suggested above, the preservation of people's essential links to their past, their identity, and their cultural patrimony relies on their retention of the linguistic tools that provide them with access to those things. And secondly, as I will argue in the next section, the retention of a culture's traditional language is not at all mutually exclusive with obtaining the benefits another language may provide in the way of enhanced social and economic opportunity. Having one's cake and eating it too is perfectly natural, when it comes to language diversity.

## **Language contact and the fates of languages**

Linguistic diversity all over the world is in imminent danger of major decline, and at least part of the reason for that is competition between 'local' languages and the major languages of wealth and power. But is it inevitable that all such confrontations will be resolved by the triumphing of the world's dominant languages over others? Is it, in fact, appropriate to picture situations of contact between languages as inherently confrontational, with the only possible outcome the 'victory' of one over another? My purpose here is to suggest that there are many varieties of language contact, and that in some cases the results are rather more harmonious than this picture would suggest.

The usual picture of what happens when two languages are spoken within the same region or population suggests the image of animal species competing for the same resource. For example the North American grey squirrel (*Sciurus carolinensis*) has largely displaced the native red squirrel (*Sciurus vulgaris*) throughout most of the woodlands in Great Britain, as a result of competition for food between the two. Fire ants (*Solenopsis invicta*) and the Argentine ant (*Linepithema humile*) have had similarly devastating effects on other ant species with which they compete. The linguistic analogue is the extinction of many local languages in the face of the onward march of global languages such as English, Spanish, and Chinese, as discussed earlier in this chapter.

Another possible outcome of overlap, at least between closely related but distinct species, is that some hybridization may take place, with the result of a limited transfer of genetic material from one to the other. For example, the African Wild Dog (*Lycaon pictus*) is endangered in part because of gene flow from domesticated dogs (*Canis lupus familiaris*) that interbreed with the wild species. Mallard ducks (*Anas platyrhynchos*) interbreed rather freely with other members of the genus *Anas*, and this has resulted in the virtual extinction of some rare wild species.

Of course, a corresponding phenomenon is much commoner when we are dealing with languages, where it corresponds simply to borrowing. Even where there is virtually no mutual comprehensibility between two languages (corresponding to complete reproductive isolation between species), it is possible for one to borrow isolated words from the other, and if there are enough individuals who speak both, even structural features of phonology or grammatical structure can be borrowed.

Cases of complete blending of the characteristics of two species through hybridization are common in plants, but not in animals. Something rather like this seems to have happened with languages, surprisingly, in a small set of cases. Michif, for example, is a language spoken by Métis in Saskatchewan, Manitoba, and North Dakota. Michif combines the grammar and lexicon of noun phrases in French with the

grammar and lexicon of verb phrases from a variety of Plains Cree. Articles and adjectives also come from French, but demonstratives from Cree. Similarly, the nearly extinct Mednyj (or Copper Island) Aleut language spoken by a mixed-blood Russian and Aleut population combines those two languages, as described by Nikolaj Vakhtin.

Roughly speaking, the majority of verbal stems and many noun stems are of Aleut origin, as is the derivational morphology [non-inflectional word structure – *sra*]. Most auxiliaries and adverbs are of Russian origin, and all of the verbal morphology [verb inflection] is Russian as well.

When two (or more) languages compete for the attention of a population of speakers, though, there is a possible resolution which has no obvious analogue in the case of biological species: bilingualism. In fact, it is probably true that *most* people in the world are bilingual or multilingual. In some instances, this is just a way-station between their use of a traditional, local language and their eventual adoption of the language of a dominant community, but often this state is quite stable, not just for individual speakers but for the larger society of which they are a part.

There is more than one variety of stable bilingualism. In some instances, local languages have large populations of speakers who use their language for most daily interactions, but who also maintain considerable fluency in another language with greater international currency. Most university-educated people in India know English – in many instances, at least as well as the Indian language of their community (and the national language Hindi, if this is not their own). Speakers of Dutch in the Netherlands, and of Danish in (at least urban) Denmark also quite generally speak excellent English with ease. None of the major languages of India, or Dutch, or Danish, or a variety of other languages in similar situations, seem likely to disappear in the face of the regular use of English by many of their speakers. The roles of the different languages involved are distinct in the lives of these speakers, and neither aspect of life is significantly diminished by the existence of the other.



In some cases, bi- or multilingualism is actually a national policy. The most famous example, perhaps, is that of Paraguay, where both Guaraní and Spanish are official languages. Both are quite generally known and used, although there is reason to believe that after many years of effective parity, Guaraní is currently losing ground relative to Spanish.

Sometimes, though, the national policy is not as representative of the actual state of affairs as it might seem. Switzerland is notable for having four national languages (German, French, Italian, and Rumantsch – another distinct Romance language), but the Swiss are certainly not uniformly quadrilingual in the sense of speaking all of these. First of all, the German that is official is High German, not the native Swiss German which is actually spoken at home and on the street in the German parts of the country. People in these areas are generally (at least) bilingual, but the (first) two languages involved are High German and (some local form of) Swiss German.

In some areas of the country, the situation is more elaborate. In the Poschiavo valley, in the canton of Graubünden, children grow up speaking the local Romance language, Pus'ciavin (or Poschiavense). They may also acquire some command of Lombardo, a *lingua franca* useful across a substantial part of the northeast of Italy and adjacent parts of Switzerland; and then when they go to school, they learn standard Italian. These young adults are already trilingual, without thereby having gone beyond what most outsiders would think of as 'Italian'!

Very few Swiss outside of the cantons of Ticino and Graubünden, however, speak more than the most basic Italian. In all parts of the country, students learn at least one other of the national languages (though in some areas, this requirement is being replaced by English), but the extent to which they actually speak this language after leaving school is highly variable. Those from the French part often cannot hold a conversation in German (almost never in Swiss German, and only rarely in Italian), and those from the German part are only marginally better in French. And hardly anyone other than native Rumantsch speakers from Graubünden understands any Rumantsch at all. Swiss multilingualism is

official policy, but that policy corresponds somewhat imperfectly to the reality on the ground.

A rather striking form of systematic multilingualism in some indigenous communities is linguistic exogamy: a rule or tendency preferring that one marry someone speaking a different language from that of one's home. The inhabitants of the Vaupés region in northwest Amazonia, for example, practise exogamy among groups speaking various languages of the Arawakan and Eastern Tukanoan families (such as Tariana, Tukano, Desano, Barasana). Families live with the father's group, but marriage is with someone speaking a different language: 'We don't marry our sisters.' Both parents are at least passively bilingual, while the children speak both of the languages of their parents actively, and perhaps others as well. Multilingualism is quite general, with little language mixing, limited lexical borrowing, but substantial diffusion of grammatical features throughout the region.

Lyle Campbell and Verónica Grondona describe a related pattern in Misión La Paz, Argentina.

Three indigenous languages, Chorote, Nivaclé, and Wichí, are spoken here, but interlocutors in conversations usually do not speak the same language to one another. There is extensive linguistic exogamy, and husbands and wives typically speak different languages to one another. Individuals identify with one language, speak it to all others, and claim only to understand but not to speak the other languages spoken to them. Children in the same family very often identify with and thus speak different languages from one another.

Another region in which this system was traditionally found is the Cape York peninsula in Australia, where multilingualism was standard among Aboriginal people in the service of being able to communicate with others across a range of dialect and language areas, and often involving linguistic exogamy. In modern times, speakers from a number of communities have been brought together at a single mission, and their languages have been largely subordinated to a single form (Guugu Yimidhirr), in addition to English.

Sometimes control of multiple languages within a population plays a somewhat different role, corresponding to other differences within the society. A famous example is the village of Kupwar in Maharashtra state in central India. The state language is Marathi (an Indic language); the village is near the border with the state of Karnataka, where the state language is Kannaḍa (Dravidian). Five different social groups speak four different languages. In addition to Kannaḍa (spoken by Jain landowners and Lingayat craftsmen) and Marathi (spoken by landless labourers, the Untouchables), Urdu is spoken by a group of Muslim landowners and another Dravidian language, Telugu, by ropemakers. Most members of the community control all of these languages to at least some extent, and interact regularly with one another. Each group speaks its own language at home and others in public when dealing with members of the other groups. The different languages are associated in a stable way with different segments of the community, and the resultant multilingualism seems destined to persist as a function of the multi-ethnic nature of the society.

In many parts of the world – including India, the Amazon, New Guinea, Aboriginal Australia, much of Africa – knowledge of a number of languages is an essential part of a life where one naturally comes into contact with members of other groups. From this evidence, it seems reasonable to consider that multilingualism is a perfectly natural condition of humankind. Nonetheless, various conditions can lead to the reduction or elimination of such diversity.

An example of the loss of a language in a multilingual situation is provided by Votic, a Fenno-Ugric language spoken in Russia near Estonia. The few remaining speakers use Votic, Ingrian (another Fenno-Ugric language, also called Ižora), Russian, Estonian, and Finnish. As a consequence of forced movements of population, and the fact that all of these other languages (including the closely related Ingrian) have rather higher prestige than Votic, speakers tend to use them or to import words and grammatical features from them into Votic, with the result that Votic has essentially disappeared as a distinct language.

Sometimes such language shift has been the effect of political policy,

as in the case of numerous North American and Australian native languages beaten out of the children forced into English-speaking residential schools in the US, Canada, and Australia. Often it has been the result of the expansion of major languages of Europe into other parts of the world, as in most of the former colonies of European powers. And not only European languages have had such effects: other instances range from the replacement of Egyptian by Arabic in Egypt to the replacement of Eyak (a language related to the Athabaskan family) by Tlingit along the northwest coast of America.

The dynamics of such linguistic replacements vary quite a bit. In some instances, of course, they are mandated or at least strongly encouraged by an expanding or occupying group as a function of its effective control over another. In other cases, such 'linguistic imperialism' may not be what drives the change: when speakers of the Cushitic language Yaaku in Kenya shifted to the Nilotic language Maasai, it was for a variety of social and economic reasons, to obtain the prestige associated with a higher-status culture rather than because the Maasai themselves encouraged it.

This sort of motivation is perhaps the dominant force driving the languages of small communities to extinction in the world today. The perception is that in order to avoid marginalization and to obtain access to economic and social wellbeing, it is necessary for such groups to adopt the dominant culture, and as a path to that, the language of that culture. There is little doubt that becoming competent in the language of a larger surrounding community is an essential step in obtaining access to a variety of benefits associated with that community. But it is false to present the learning of such a language as necessarily entailing the abandonment of the group's original language.

The economic argument by itself does not require that speakers of a 'small' and perhaps unwritten language must give up that language simply because they also need to learn a widely used language such as English, Arabic, or Mandarin Chinese. The alternative, of course, is to maintain both languages, and, as we have seen, that is a common and stable way of life in many places. Where there is no one dominant local

language, and groups with diverse linguistic heritages come into regular contact with one another, multilingualism is a perfectly natural resolution of the various influences on speakers.

And as we have also seen, there is good reason to make the effort to preserve an original language alongside a new one, even when the new one is central to economic security and social integration into the surrounding society. When a language dies, a world dies with it, in the sense that a community's connection with its past, its traditions, and its base of specific knowledge are all typically lost when the vehicle linking people to that knowledge is abandoned. This is not a necessary step, however, in order for them to become participants in a larger economic or political order. Like others elsewhere, they could choose to have the best of both worlds by retaining their traditional language for some purposes and adopting another as well for those spheres of activity in which it is essential.

## Chapter 5

### Some problems in the counting of languages

In [Chapter 2](#), the world's languages were compared in terms of the numbers of their speakers, and families and geographical regions in terms of the number of their languages. Both discussions were based on the most authoritative figures currently available, those of *Ethnologue*, and both presume that it makes sense to count these things, and that we have reasonably clear ways of performing such counts. But those presumptions are not self-evidently true.

### Counting the speakers

In the case of numbers of speakers, it is notable that the figures provided for each language in *Ethnologue* represent 'the total number of people who use those languages as their first language, no matter where in the world they may live'. In some instances, especially for rather 'small' languages, this may well give a fair estimation of a language's prominence among modern humans. In others, it does not: the figure of 328 million for English excludes more than 167 million speakers of English as a second (or third, etc.) language. The figure of 182 million for Hindi omits the fact that there are roughly as many non-native speakers who use this language regularly as there are native speakers.

In some instances, these complications are recognized in official figures. The Swiss census of 2000 asked two questions about language: which language do you have best command of? and which language do you speak regularly? For Rumantsch, the country's fourth official language, the first question yielded a figure of about 35,000, and this is the figure that appears in *Ethnologue*. Counting those who described Rumantsch as a language of regular use (at home, in school, at work),

however, this grows to more than 60,000, and the census reports both realities.

The situation for Arabic is particularly confusing in this regard: *Ethnologue* gives a total of 221 million for the various varieties of colloquial Arabic, but Modern Standard (or Literary) Arabic does not appear on the list of the ‘biggest’ languages, despite its status as a second language used for ‘education, official purposes, written materials, and formal speeches’ throughout the Arab world and beyond, and neither does the Classical Arabic of Islamic religious life, despite the number of people who have varying degrees of command of it. It is clearly difficult to quantify the precise extent of many languages within the world’s population, because the way one counts depends largely on these political and social factors.

## Counting the flags

With respect to the enumeration of languages, *Ethnologue*’s total of 6,909 is surely not to be taken as a precise count, but we have assumed that in principle we know how to count the world’s languages. It might seem that the remaining imprecision is similar to what we might find in any other census-like operation: perhaps some of the languages were not home when the *Ethnologue* counter came calling, or perhaps some of them have similar names that make it hard to know when we are dealing with one language and when with several; but these are problems that could be solved, in principle, and the fuzziness of our numbers ought thus to be quite small. But in fact, what makes languages distinct from one another also turns out to be much more a social and political issue than a linguistic one, and most of the cited numbers are matters of opinion rather than science.

These points are certainly not unknown to the compilers of *Ethnologue*, who discuss them to some extent in the Introduction to that work. With respect to the individuation of languages, for example, they note that:

[h]ow one chooses to define a language depends on the purposes one has in identifying that language as distinct from another. Some base their definition on purely linguistic grounds. Others recognize that social, cultural, or political factors must also be taken into account. In addition, speakers themselves often have their own perspectives on what makes a particular language uniquely theirs. Those are frequently related to issues of heritage and identity much more than to the linguistic features of the language(s) in question.

Nonetheless, they have to make some choices in order to have data to present, and the ones they make are not unreasonable. It is important to understand, however, that neither are they purely objective or definitive. ‘6,909’ is one way of looking at the world’s linguistic diversity, but there are others as well. It is the goal of this chapter to consider some of the complications that make this such a difficult problem.

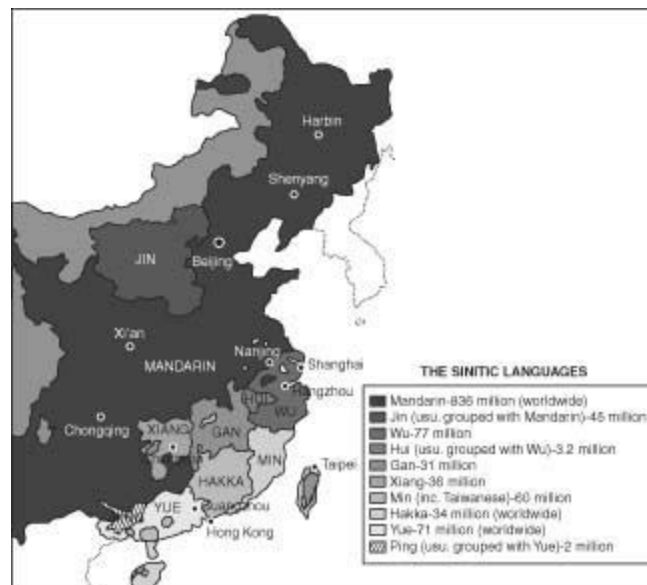
The late Max Weinreich used to say that *‘a shprakh iz a diyalekt mit an armey un a flot’* (‘A language is a dialect with an army and a navy’). He was talking about the status of Yiddish, long considered a ‘dialect’ because it was not identified with any politically significant entity. One might reformulate Weinreich’s witticism as ‘a language is a dialect with a flag’: languages are spoken by nations, while dialects are spoken by tribes, or towns, or other less important groups. The distinction is still often implicit in talk about European ‘languages’ versus African ‘dialects’. What counts as a language rather than a ‘mere’ dialect typically involves issues of statehood, economics, literary traditions, and writing systems, and other trappings of power, authority, and culture – with purely linguistic considerations playing a less significant role.

For instance, Chinese ‘dialects’ such as Cantonese, Hakka, Shanghainese, and so on constitute the family of Sinitic languages, and are just as different from one another (and from the dominant Mandarin) as are Romance languages like French, Spanish, Italian, and Romanian. They are not mutually intelligible, but their status as ‘dialects’ derives from their association with a single nation and a shared writing system, as well as from explicit government policy. Like the Romance languages, they can be sub-grouped into several quite distinct families,



of which the major ones are distributed within China as shown in [Figure 8](#). This map only shows the highest-level family divisions; ‘Mandarin’, in particular, covers a number of substantially separate sub-groups.

The distinctness of the Sinitic languages involves features ranging across all the major areas of linguistic structure: at least phonetics, phonology, word structure, syntax, and lexicon. To some extent, these differences are concealed by the fact that Chinese writing is based on characters that correspond as a whole to words (or parts of words), and not directly to sounds, which allows words in different languages that are very different in sound to be written with the same character. There are in fact some characters that only correspond to words in some of the languages, not all, so the generality of this writing system is not quite as great as it is sometimes portrayed, but it is still much more encompassing than any other in the world in the diversity of forms of expression that are all brought under it.



## 8. The Sinitic languages

While the lexicon of a language is only one aspect of its identity, it is one that is relatively easy to present. To illustrate how much the Sinitic languages differ among themselves, some lexical differences in common words across a few of them are given in [Table 6](#). Word tone, which is distinctive in all Sinitic languages, is not marked in [Table 6](#), and the

transcriptions of the different languages are not completely equivalent, but this should still give some idea of the extent to which these languages are distinct.

The putative unity of ‘Chinese’ (as a construct encompassing the Sinitic languages) thus rests on facts such as the largely shared writing system, the existence of a standardized form of Mandarin (Putonghua) which is widely used as a sort of koine, and especially the political unity of the modern Chinese state: the individual Sinitic languages do not have separate flags, armies, and navies.

In contrast, Hindi and Urdu are essentially the same system (referred to in earlier times as ‘Hindustani’), but associated with different countries (India and Pakistan), different writing systems, and different religious orientations. Although varieties in use in India and Pakistan by well-educated speakers are somewhat more distinct than the local vernaculars, the differences are still minimal — far less significant than those separating Mandarin from Cantonese, for example.

**Table 6. Common words compared across different Sinitic languages**

	Mandarin	(Wu) Shanghai	(Xiang) Changsha	(Ga) Nanchang	Hakka	(Yue) Cantonese	(Min) Amoy
I	wo	ngu	ngo	ngo	ngai	ngo	goa
you	ni	nong	n	n	n	nei	li
man	ren	nə	lan	lan	ngin	jan	lang
fish	yu	hng	yu	gnie	ngie	ju	hi
meat	rou	gno?	zou	gniuk	ngiuk	joek	bah
ear	erduo	gni	ə	o	ngi kung	ji	hī
drink	he	xa?	qia	chiak	siit	jam	lim
say	shuo	kā	kan	ua	sot	gong	kong
see	kan	khe	uan	mong	khon	tai	khoā
lie down	tang	khuen	than	khun	thong	fan	to

For an extreme example of this phenomenon, consider the language previously known as Serbo-Croatian, spoken over much of the territory of the former Yugoslavia and generally (up until the early 1990s) considered a single language with different local dialects and writing systems. Within this territory, Serbs (who are largely Orthodox) use a Cyrillic alphabet, while Croats (largely Roman Catholic) use the Latin alphabet. Within a period of only a few years after the breakup of Yugoslavia as a political entity, at least four new ‘languages’ (Serbian,

Croatian, Bosnian, and, most recently, Montenegrin) have emerged: the actual *linguistic* facts have hardly changed at all, but the number of armies and navies has grown substantially.

One can now find separate dictionaries for all four languages, although the linguistic diversity these document is not materially different from that which existed within the former ‘Serbo-Croatian’. When the Serb Slobodan Milošević was on trial for war crimes in the Hague, the indictment was read out, and as required by the procedures of the tribunal, he was asked to acknowledge that the charges against him had been read to him in his language. He acknowledged that he had heard and understood the charges in the indictment, but denied that they had been read in his own language: apparently, the person who performed this task had a Bosnian accent.

## Languages and dialects

The answer to the question ‘How many languages are there in the world?’ (if this makes sense, and there *is* an answer) probably lies somewhere between *Ethnologue*’s 6,909 and many billions – at least one for every human being, since everyone’s ability to speak and understand can be seen as a little bit different from anyone else’s, and many people control more than one system. Making sense of what is really involved, though, requires us to confront the difference between ‘languages’ and ‘dialects’.

This is not a mere quibble – the distinction is certainly seen by many as having real-world consequences. This was clear in the controversy over the resolution by the Oakland California School Board in 1996 concerning the language of instruction for African-American students in its system. Among other things, the resolution (in its 1997, slightly amended, form) affirmed that ‘African-American students as a part of their culture and history as African people possess and utilize a language described in various scholarly approaches as “Ebonics” (literally “Black sounds”) or “Pan African Communication Behaviors” or “African Language Systems”; and [...] African Language Systems *have origins in West and Niger-Congo languages* and are not *merely* dialects of

English,” [emphasis in the original], going on to clarify in a policy statement their understanding that ‘[t]he [...] linguistic evidence is that African-Americans (1) have retained a West and Niger-Congo African linguistic structure in the substratum of their speech and (2) by this criteria are not native speakers of black dialect or any other dialect of English.’

There are many things going on here, some of them matters of fact (such as the relation of the speech of contemporary African-Americans to specific African languages) which need not concern us. The basic point, though, is that the linguistic usage of the students in question constitutes a system of its own, distinct in significant ways from what is generally thought to be ‘standard’ English. This is a conclusion that had been well established in the linguistic literature during the 1970s and later, on the basis of research in which the phonology, morphology, syntax, and lexicon of ‘Black (Vernacular) English’ or ‘African-American Vernacular English’ were explored in some detail. What is at stake in the Oakland School Board’s resolution, though, is the assertion that such systems are not ‘mere dialects of English’, but a *language* in its own right, and thus entitled to the same treatment in the educational system as the first languages of other non-English-speaking students, such as Spanish, Japanese, or Hmong. This was buttressed on the one hand by associating this language with a unique cultural identity, and on the other by a distinctive name: ‘Ebonics’, a designation that makes no reference to ‘English’.

Recognition of a linguistic system as a *language* and not a *dialect* (of some other language), then, has an importance in social and political terms that is quite independent of the linguistic facts separating one such system from others. Converting the status of a ‘dialect’ into that of a ‘language’ by associating it with a distinctive political or social entity, giving it a name that minimizes its association with those of other languages, and so on are seen as legitimizing the status and rights of its speakers. This is apparent in the replacement of ‘Irish English’ or ‘Hiberno-English’ by ‘Anglo-Irish’ in describing speakers of Irish heritage, in ‘Afrikaans’ as the historical descendant of the Dutch spoken in South Africa, and elsewhere.

But in enumerating the languages of the world, should we count ‘Ebonics’ and ‘Anglo-Irish’ separately from ‘English’ because they are associated with social or political entities in this way, but treat the speech of, say, Down-East Maine, the Appalachians, Dorset, and Yorkshire as just part of ‘English’? If we do, we are certainly not describing a properly *linguistic* reality, but rather a matter that is well outside the domain of the science of language. From a strictly linguistic point of view, the distinction between different languages and different dialects is just a fuzzy matter of degree, with no systematic status. As should be clear by now, the separation between one linguistic system and another can be of varying degrees, and our understanding of that separation is not improved by treating some systems as ‘languages’ and others as ‘dialects’.

## **Differentiating languages from one another**

One common-sense notion of when we are dealing with different languages, as opposed to different forms of the same language, is the criterion of mutual intelligibility: if the speakers of *A* can understand the speakers of *B* without difficulty, *A* and *B* must be the same language. This is the basic principle that *Ethnologue* uses to distinguish languages, although they are not completely consistent in applying it. For instance, the language listing for Switzerland includes ‘Swiss German’, distinct from standard German and treated as one language. Nonetheless, they note that ‘[e]ach canton has [a] separate variety, many mutually unintelligible’. In fact, there are many more varieties of Swiss German than there are cantons, and some of these are indeed mutually unintelligible to a significant extent, and so ought properly to have been treated as separate languages if this criterion were taken literally. Apparently, one has to stop the fragmentation somewhere, but the result is somewhat problematic.

In fact, the notion of separating languages on the basis of mutual intelligibility is very similar to the ‘biological species’ concept mentioned in [Chapter 1](#), on which organisms are assigned to different species to the extent they cannot reproduce with one another. While widely accepted, that notion also has its problems. Some of these are

unique to the biological context: for instance, horses and donkeys are assigned to separate species, but in fact they can reproduce with one another, yielding either mules (if the father is a donkey and the mother a horse) or hinnies (in the reverse case). Mules and hinnies, on the other hand, are generally infertile and cannot reproduce with horses, donkeys, or each other. To what species do these animals belong? And is it relevant that a small number of female mules actually *have* produced offspring with male horses or donkeys? Fortunately, we do not have to resolve comparable conundrums in the case of languages.

Just as the biological species concept does not yield a completely clean division of organisms into species, so the notion of mutual intelligibility also fails in practice to cut the world up into clearly distinct language units. The problems are not precisely parallel, of course. Mutual intelligibility among languages is a somewhat gradient phenomenon (Catalan speakers and Spanish speakers without a background in the others' language cannot completely understand one another, but either can understand more of the other language than they could understand from a monolingual Japanese tourist), while reproductive isolation is much more nearly categorical. But there are some interesting similarities.

In some instances, speakers of *A* can understand *B*, but not *vice versa*, or at least speakers of *B* will insist that they cannot understand *A*. In some instances, these asymmetries are probably quite real. Among the Scandinavian languages, native speakers of Danish generally report being able to understand a great deal of what is said in (Riksmål) Norwegian, which is not surprising, given that this language has its origin in the Danish spoken in Norway as a standard under Danish rule of the country from the 16th through the early 19th century. On the other hand, intelligibility of modern Danish for native speakers of Norwegian is considerably lower, probably because Danish has undergone a series of sound changes that have rendered its surface phonetic form quite unusual.

Apart from what it shows us about the messiness of mutual intelligibility as a criterion of language separation, the Scandinavian

situation brings up another similarity in the speciation problem in biology and in linguistics. Notice that although Danish and Norwegian are closer to one another historically than either is to Swedish, in intelligibility terms, Swedish and Norwegian are more nearly similar and Danish is the odd one out. The biological concept of speciation is similarly ahistorical. Suppose we have a set of populations whose relation to one another is as diagrammed below.



Now suppose that although characteristic traits of these populations have been innovated at the various branching points in the diagram, these have not been sufficient to disrupt the reproductive compatibility of the members of A, B, C, and D. But now suppose that D innovates a trait that does disrupt this compatibility, such that while A, B, and C can still reproduce with one another, D is reproductively isolated. The result is that in terms of the biological species concept, the set of A, B, and C constitute one species, while D is a separate species. The resulting grouping is quite contrary to the historical relations, in terms of which C and D are more closely related to each other than either is to A or B. This is more or less comparable to the problem just noted for the Scandinavian languages.

In other instances, failures of mutual intelligibility surely have no basis in objective facts about the languages involved, but rather are rooted in social and cultural attitudes. Bulgarians, for instance, consider Macedonian a dialect of Bulgarian, but Macedonians insist that it is a distinct language. When Macedonia's president Gligorov visited Bulgaria's president Zhelev in 1995, he brought an interpreter, although Zhelev claimed he could understand everything Gligorov said. The signing of a protocol broke down when Gligorov insisted on a statement that it had been written 'in the Macedonian language'.

Somewhat less fancifully, Kalabari and Nembe are two linguistic varieties spoken in Nigeria, both treated as forms of Eastern I some

scholars. The Nembe claim to be able to understand Kalabari with no difficulty, but the rather more prosperous Kalabari regard the Nembe as poor country cousins whose speech is unintelligible.

Another reason why the criterion of mutual intelligibility fails to tell us how many distinct languages there are in the world is the existence of *dialect continua*. Large parts of Germany, Switzerland, and the Benelux countries, for instance, make up a broad area in which West Germanic languages are spoken that vary only slightly from one locality to the next, but considerably between regions that are some distance apart. To illustrate, suppose you were to start from Amsterdam and walk the 300 or so miles to Frankfurt am Main, covering about 10 miles every day. You can be sure that the people who provided your breakfast each morning could understand (and be understood by) the people who served you supper the same evening. Nonetheless, the Dutch speakers at the beginning of your trip and the German speakers at its end would have somewhat more trouble, and certainly think of themselves as speaking two quite distinct (if related) languages. Part of that sense surely comes from the fact that you would have crossed a boundary between one flag and another. The day's walk during which that occurred (let us say, between Maastricht and Aachen), though, would probably not represent a greater difference between local colloquial speech (as opposed, say, to the road signs) than other days.

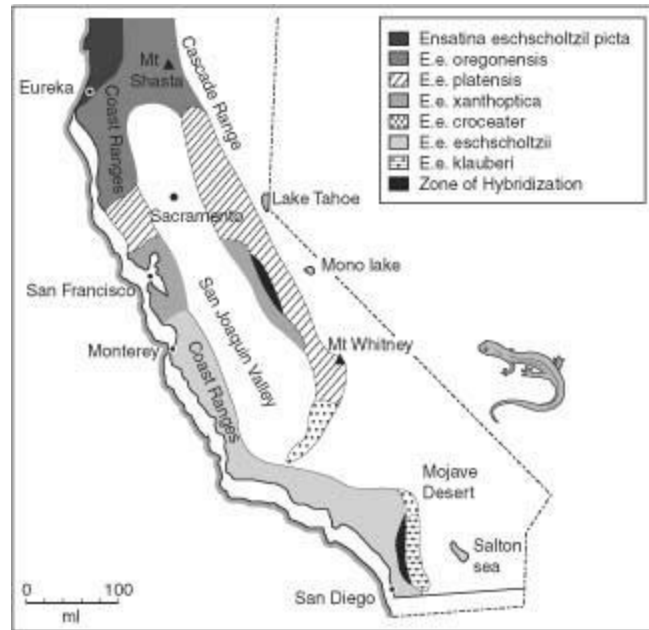
In some parts of the world, such as the Western Desert in Australia, such a continuum can stretch well over a thousand miles, with the speakers in each local region able to understand one another while the ends of the continuum are clearly not mutually intelligible at all. How many languages are represented in such a case?

A very similar issue arises in the biology of species, where the analogue of such dialect continua is the existence of collections of organisms that constitute a *Rassenkreis*, or 'ring species' as it was called by Ernst Mayr. The classic example of this is a collection of subspecies of salamander (*Ensatina eschscholtzii* var.) found in Oregon and California. This species apparently originated in the north, and spread southward along the two sides of the California central valley, as shown



in [Figure 9](#). As they spread, they diversified in form, but a certain amount of gene flow among populations seems to have been maintained. At the extremes of their southern range, however, the coastal and inland varieties overlap, and in these areas the two do not interbreed.

Another famous example of a ring species are a set of varieties of gulls (*Larus spp.*) found all around the northern sub-polar region.



## 9. Varieties of the salamander species *Ensatina eschscholtzii* in California

The number of distinct species to be recognized is controversial, but it is reasonably clear that gulls from Siberia interbreed with ones from America, and these with herring gulls in Great Britain. In the other direction, Siberian gulls interbreed with European varieties westward to the lesser black-backed gull, and there is thus gene flow all around the circle; but where the herring gulls and the lesser black-backed gulls varieties overlap in range, they do not interbreed.

Striking from a linguist's point of view is the ring species represented by a collection of populations of the greenish warbler (*Phylloscopus trochiloides*), a songbird found in Asian forests all around the Tibetan plateau. Again, we find that there is interbreeding between adjacent

populations, but where the two ends of the ring meet, the distinct varieties do not interbreed. One major contributor to the failure of interbreeding in the overlap zone is particularly interesting to us in this case. The birds do not interbreed at least in part because they apparently do not recognize one another's song as that of a conspecific. Reproductive isolation is here ensured by lack of mutual intelligibility.

Returning to the role of such intelligibility in differentiating human languages, a different problem was already noted above in [Chapter 3](#): the non-identity of what we call the 'same' language over time. We refer to the language of, say, Chaucer (1400), Shakespeare (1600), Thomas Jefferson (1800), and George W. Bush (2000) all as 'English', but it is safe to say these are not all mutually intelligible. Shakespeare might have been able, with some difficulty, to converse with Chaucer or with Jefferson, but Jefferson (and certainly Bush) would need an interpreter for Chaucer. Languages change gradually over time, maintaining intelligibility across adjacent generations, but eventually yielding very different systems. The resolution of this difficulty through the recognition of a 'phylogenetic' dimension to language identity as suggested in [Chapter 3](#) is quite similar to the way biologists confront the variation in the members of a single species over evolutionary time.

The notion of distinctness among languages, then, is much harder to resolve than it seems at first sight. Political and social considerations trump purely linguistic reality, and the criterion of mutual intelligibility is ultimately inadequate.

## Chapter 6

# The genotypes of languages

[Chapter 5](#) has shown us that commonly assumed ways of determining the number of distinct languages in the world fail to provide a satisfying approach to the problem. National identification is obviously significant for social identity, but that should be kept apart from the question of when we are (or are not) dealing with distinct languages. The apparently obvious criterion of mutual intelligibility turns out to be much more problematic than it seems, and does not really give a satisfying answer either. And the whole question of when we are dealing with different languages and when with different dialects of a single language seems quite impossible to turn into a rigorous distinction.

Does the science of linguistics provide a better basis for measuring the number of different languages spoken in the world? When we address the question of just when forms of speech differ systematically from a linguistic point of view, the answers that we get are potentially crisp and clear, but rather surprising.

The basis of this approach to enumerating languages is to appeal to a basic distinction made by many linguists, that between a language as something ‘out there’ in the world, and language as something in the minds of speakers. That is, we can think of a language as a collection of sounds, words, phrases, sentences, etc. as these are realized in speech or signing (or derivatively, in writing), on the one hand, or else as a system of knowledge speakers have and on the basis of which they are able to produce and interpret those objects, on the other.

The latter conception, which I will refer to as an *I-language* (for ‘internal’) refers to that aspect of a speaker’s cognitive organization by virtue of which we say that she ‘speaks’ or ‘knows’ the language in

question. A person with a particular I-language is then able to produce and comprehend particular utterances (objects that are part of *E-language*, for ‘external’). This takes place on the basis of an interaction of I-language knowledge with facts of the environment (the presence of food in the mouth or alcohol in the bloodstream, for example), other aspects of cognition (memory, what one knows about the world, etc.), and a variety of other factors.

## Differences in languages as differences in grammars

We can draw an (imperfect) analogy between the I-language/E-language distinction and that between the *genotype* and the *phenotype* in biology. One is the abstract pattern of a language or an organism, and the other is the way that pattern is realized in concrete circumstances. The parallel is not exact, of course, but it is at least suggestive. The genotype is a sort of ‘recipe’ for building an organism, while the phenotype is the result of implementing that recipe in a specific way in a particular context. Similarly, an I-language provides a general recipe for the structure of linguistic objects, which plays a role in conjunction with various other effects and systems in the actual production and understanding of E-language words, phrases, sentences, and so on.

Of course, the I-language of an individual is a component of that person’s (biological) phenotype and not his genotype, a point which will be made again in [Chapter 8](#). The parallel being suggested here has to do with a similar relation of abstraction, not with biological implementation.

In biology, the attempt to differentiate organisms on the basis of differences in the phenotype leads to the morphological species concept mentioned in [Chapter 1](#). Biologists have generally been unhappy with this approach, since it tends to exaggerate the value of incidental and circumstantial differences, and have preferred to look for similarities and differences in the genetic makeup of organisms as the basis for identifying species (at least since the development of an understanding

of genetics and mechanisms for identifying genetic material). This leads to the genetic species concept, and the analogous move in the study of language would be to identify differences among I-languages and treat as separate any two of these that were significantly distinct.

When we try to distinguish languages from one another ‘phenotypically’, in terms of their words and the patterns we can observe in E-language sentences, problems arise. Very different languages can share words (through borrowing), while different speakers of the ‘same’ language may vary widely in their vocabulary due to factors of education or speaking style. Different languages may display the same sentence patterns, while a single language may display a great variety of patterns. In general, linguists have found that the analysis of the external facts of language use, however interesting in its own right, gives us at best a slippery object of study for the specific purpose of individuating the world’s languages. Rather more coherent, it seems, is the study of the abstract knowledge speakers have which allows them to produce and understand what they say or hear or read: their internalized knowledge of the *grammar* of their language, or I-language.

As an example of speakers’ knowledge, let us consider the past-tense forms of verbs in English. A number of verbs, of course, are ‘irregular’, like *eat/ate*, *bring/brought*, *find/found*, and especially *be/is/was*, and these have to be learned one by one. For most verbs, however, the past tense consists of the same form as the present, with an ending. The ending for the past is spelled uniformly as *-ed*, but it is pronounced (phonetically) either as [əd] (e.g. in *waited* [wejtəd]; [ə] is ‘schwa’, the short, reduced vowel that appears in the second syllable of *appetite*), as [t] (e.g. in *missed* [mɪst]), or as [d] (e.g. in *rowed* [rowd]). These various phonetic shapes are not at all randomly distributed: *-ed* is pronounced as [əd] when the verb ends in *t* or *d*; as [t] otherwise when the verb ends in a voiceless consonant like *p*, *t*, *k*, *f*, *s*, or *ch*; and as [d] elsewhere (after a vowel or a voiced consonant).

The distribution of the three phonetic forms of the *-ed* ending might be just a fact about the vocabulary of English, learned word by word like

the irregular past-tense forms *ate*, *brought*, *found*, etc. Things cannot be that simple, though. When English speakers are presented with a supposed new verb, they can produce a past-tense form of it without hesitation, and the forms that they give will follow the same principle. If shown a picture in which they are told that ‘The man is *dending/plipping/sprimming* his cat’, and then asked ‘what happened?’, the responses will be that the man [dendəd]/[plɪpt]/[sprɪmd] the cat. That is, speakers not only know the words of their language, they also know a principle (or ‘rule’) that governs the formation of words, and are able to make use of this rule to produce novel words that follow the same pattern as other familiar ones. And on that basis, we can say that the I-language of speakers of English is characterized in part by the fact that it includes this rule.

Let us now compare these facts with what we might be able to learn about the I-language of a speaker of Scots. Scots is a language of lowland Scotland which is closely related to English, but which separated from it historically some time during the Middle English period, roughly by the end of the 15th century. We might call it a very divergent dialect of English, or a language in its own right, but as we saw in [Chapter 5](#), such a difference is not a fact about the language itself, as opposed to the social and political opinions of those who speak it and their neighbours.

Like its close relative (standard) English, Scots has a variety of ways of making past-tense forms of verbs. As in English, a good number of verbs have irregular forms that must be learned: *bite/bate* ‘bite/bit’; *rive/ruive* ‘tear/tore’; *begin/begoud* ‘begin/began’, etc. But for most verbs (the ‘regular’ ones), there are three phonetic forms that are used, again depending on the final sound of the verb. There is some variation from one locality to another in the principle involved (Scots does not have nearly as strong influence from a unitary standard as standard English has, and so local variation is not subject to normalization to the same extent), but one pattern is as follows. Verbs ending in a stop consonant ([p,t,k,b,d,g]) add *-it* (e.g., *drap/drappit*, ‘drop/dropped’; *want/wantit*, ‘want/wanted’; *keek/keekit*, ‘peep/peeped’; *sab/sabbitt*, ‘sob/sobbed’; *mynd/myndit*, ‘remember/remembered’; *big/biggit*,

‘build/built’). Verbs ending either in a voiceless obstruent that is not a stop, like *s*, *sh*, or in a sonorant consonant (*l*, *m*, *n*, *r*, *ng*) add *t*: *loss/lost*, ‘lose/lost’; *fash/fasht*, ‘bother/bothered’; *dirl/dirlt*, ‘vibrate/vibrated’; *ken/kent*, ‘know/knew’; *soum/soumt*, ‘swim/swam’, etc. Finally, verbs ending in a voiced obstruent consonant other than *d* like *v*, *z*, *dʒ* (the first and last sounds of *judge*) etc. or in a vowel add *d*: *deave* [di:v] ‘deafen’, past tense *deaved* [di:vd]; *lowse* [lowz] ‘loosen’, past tense *lowsed* [lowzd]; *wadge* [wadʒ] ‘wedge’, past tense *wadged* [wadʒd]; *lue* ‘love’, past tense *lued*; *pey* ‘pay’, past tense *peyd*.

The rule of past tense formation just illustrated is just as much a part of the I-language of our Scots speaker as the other one is for a speaker of standard English. Confronted with novel verbs like *dend*, *plip*, and *sprim*, the Scots speaker will make past-tense forms *dendit*, *plippit*, and *sprimt*, and the English speaker *dended*, *plipped* ([plɪpt]), and *sprimmed* ([sprɪmd]). We can say that the ‘recipes’ for constructing past tense forms differ between the two systems – that is, their ‘linguistic genotypes’ differ in a way that expresses itself in the ‘phenotypic’ difference between these two sets of possible novel creations.

The internal knowledge, the I-language, or simply put, the grammar of the two differs (at least) in the presence of one rule as opposed to the other, and so we can say that there are two distinct systems, or languages, that are represented. We might then propose that instead of counting languages in terms of their external forms, we could instead count the range of distinct grammars in the world. This may or may not turn out to correspond to the ordinary understanding of what we mean when we say we are dealing with different languages, but at least it would have the virtue of describing an objective linguistic reality rather than a subjective matter of social or political opinion.

## **Constants and variables in grammars**

This naturally leads to the question of what can potentially differentiate one I-language system from another. In fact, it is probably the case that some aspects of a speaker’s knowledge are not subject to variation. For example, consider these sentences.

(1) a. Doris expected to appoint *her* as the new chair.

b. I wonder who Doris expected to appoint *her* as the new chair.

If we ask who the pronoun *her* might refer to in these sentences, it should be clear that in the first example, *her* might potentially refer to any female in the world – with one exception. *Her* in this case cannot refer to the same person as *Doris*. This is quite a consistent judgement across speakers, and follows not from the meaning of the sentence (Doris might quite possibly be in a position to determine who should be the new chair, and to appoint herself, but we could not express her intention to do so in this way), but only from its grammatical form. On the other hand, in the second example, which contains the same sequence of words, it is perfectly possible (perhaps even preferred, out of context) to interpret *her* as referring to Doris.

Similar facts obtain in the sentences in (2), this time with the pronoun (*she* or *her*) coming earlier in the sentence than *Doris*.

(2) a. *She* intended to appoint Doris as the new chair.

b. *Her* mother intended to appoint Doris as the new chair.

Again, in the first sentence the pronoun (*she*) can refer to any female individual *except* the one referred to by *Doris* in the same sentence, while in the second sentence there is no such limitation on the interpretation of *her*. The facts are quite clear, and again our judgement cannot be due to the meaning of the sentences but only to their grammatical form.

The sentences in (3) provide further examples, in sentences of a slightly more complex form.

(3) a. Doris was amazed at how quickly *she* recovered.

b. How quickly Doris recovered amazed *her*.

c. *She* was amazed at how quickly Doris recovered.

d. How quickly *she* recovered amazed Doris.

In these examples, we see that the pronoun (*she* or *her*) and *Doris* can



refer to the same individual in all but one case: (3c), where *she* must refer to someone other than Doris.

There must therefore be some principle of grammar, part of the I-language knowledge of speakers, from which it follows that pronouns can sometimes be interpreted as referring to the same individual as some other expression within the same sentence and sometimes cannot be interpreted in that way. The principle involved can be formulated roughly as in (4):

- (4) **Co-reference:** A pronoun cannot refer to the same individual as another expression in the same sentence (its *antecedent*) if it both (a) precedes the antecedent, and (b) is ‘higher’ in the structure.

Discussion of this matter in the technical literature of linguistics has been devoted to making the relevant notion of being ‘higher’ in grammatical structure precise, but the condition involved is fairly well understood.

It is interesting to ask where our knowledge of a principle like (4) might have come from. Introspection should make it clear that no one has ever explicitly taught us this: indeed, as far as we know, no one had ever even noticed the phenomenon prior to its independent discovery by two linguists, John Ross and Ronald Langacker, around 1967. Even more surprisingly, it turns out that when the details of (4) are clarified, the same principle applies as a part of I-language knowledge in every language in which it has been seriously investigated!

## Differences in grammars: parameters

It seems, then, that some aspects of grammatical knowledge, like those governing the interpretation of pronouns, may be common across languages. On the other hand, the fact that adjectives precede their nouns in English (we say *a red balloon*, not *a balloon red*) is a fact about English, since the opposite is true, for instance, in French. If we had a complete inventory of the set of *parameters* of variation that can

serve in this way, we could then say that each particular collection of values for those parameters that we could identify in the knowledge of some set of speakers should count as a distinct language.

But let us see what happens when we apply this approach to a single linguistic area, say northern Italy. Consider the facts of negative sentences, for example. Standard Italian uses a negative marker which precedes the verb (*Maria non mangia la carne*, ‘Maria not eats the meat’), while the language spoken in Piemonte (Piedmontese) uses a negative marker that follows the verb (*Maria a mangia nen la carn*, ‘Maria she eats not the meat’). Here we have a difference in grammar between standard Italian and Piedmontese, and these thus constitute two distinct I-languages.

There are other differences in the grammar of negation between the two systems as well. Standard Italian cannot have a negative with a second person singular imperative verb, but uses the infinitive instead: ‘Don’t eat!’ is *Non mangiare!*, ‘nottoeat!’, and not *\*Non mangia!*, ‘not eat!’ (note that in this chapter, an asterisk preceding a sequence of words indicates that the sequence is not a grammatical sentence in the language under consideration, at least not with the intended meaning), while Piedmontese allows negative imperatives: here, ‘Don’t eat!’ is *Mangia nen!*, ‘eat not!’. Standard Italian requires a ‘double negative’ in sentences like *Non ho visto nessuno*, ‘not have I seen nobody’, while Piedmontese does not use the extra negative marker, and so on. The functioning of negation here establishes at least three dimensions of difference between these (and potentially other) grammars.

This is only the beginning, though. When we look more closely at the grammatical systems found in various areas in northern Italy, we find a number of other dimensions of variation, potentially distinguishing many I-languages from one another within this area. In principle, each of these parameters could vary from place to place in ways that are independent of all of the others. Still staying within northern Italy, let us suppose that there are, say, ten such dimensions on which one I-language can differ from another. This is actually quite a conservative estimate, in light of the variation that has actually been found there:

many more than ten such differences have been documented. But if each of the ten can vary independently of the others, collectively they define a set of two to the tenth, or 1,024 distinct grammars, and indeed scholars have estimated that somewhere between 300 and 500 of these distinct possibilities are actually instantiated in the region!

This is so even after taking into account a potential simplification of the situation. In the discussion above, three aspects of the grammar of negation were noted in terms of which standard Italian differs from Piedmontese, and the implication was that in another I-language, each of these could potentially vary independently of the others. But in fact, when we examine a range of northern Italian systems, that is not what we find. Instead, we find that the standard Italian pattern, with preverbal negatives, the infinitive instead of another form in negative imperatives, and double negation is completely replicated in the otherwise quite distinct systems characteristic of Genoa, Venice, Trento, and Trieste. Similarly, the Piedmontese pattern of post-verbal negation, true negative imperatives, and no double negation is replicated in the languages of Turin, Milan, Aosta, and Pavia. It seems, that is, that instead of three independent dimensions of contrast, there is a single potential parameter of difference, and the choice of a value for this parameter on one dimension (for instance, pre-verbal versus post-verbal negative markers) brings along with it choices on the other two.

A serious discussion of these matters must be based on an explicit account of the structure of grammars, so that the differences in the behaviour of negation can be made systematic. We can hope that it will be possible to go beyond simply listing a number of properties of grammars that go together. What we want is to provide an analysis of grammatical structure such that a change one way or another at a single point will have all of these properties as automatic, logically implied consequences.

A serious effort in that direction, even for the single area of negation in the languages of northern Italy, would take us well beyond the scope of the present book. But even without fully fleshing out the picture in such a way, the general point at stake here can be appreciated. When

possible, we should try to unite a number of individually minimal differences between grammars under a single parametric choice. This is clearly a desirable move, to the extent it can be carried out, because the number of specific ways in which I-language systems can be shown to differ across the world's languages is truly vast, and connecting multiple such differences and uniting them under a single more abstract choice brings at least some order to the overall space of possible languages.

## The 'null subject' parameter

A particularly well-known set of individual differences among languages that seem susceptible to unification under a single parameter starts from the observation that some languages require the subjects of sentences to be overtly represented even when it is possible to recover their content from the situation, while others allow such subjects to go unpronounced. In (standard) Italian, the sentence *Ha già telefonato* is perfectly well formed, and is interpreted as either *she has already telephoned* or *he has already telephoned*, depending on the context. It is quite impossible to omit the subject in English, however: *has already telephoned* by itself is incomplete and not acceptable without the addition of a subject.

The obligatory nature of subjects in English is true even when the subject has no real content, and whatever appears there does not refer to anything. We say *it is raining/snowing/hailing/thundering*, even though *it* in such sentences has no referent. In contrast, Italian says simply *piove/nevica/grandina/ tuona*, with the same meaning but with no overt subject. Not only is no subject necessary here: in fact, there is no element of the language that *could* serve as the subject. English uses elements like *it* and *there* as meaningless (or 'expletive') subjects in such cases, while Italian does not have (or need) them, being content to leave the subject unexpressed.

Another difference between the two languages is that Italian freely allows subjects to follow the verb. *Ha telefonato Gianni* is fine in this language, while the exact English equivalent *\*Has telephoned John* is not, and must be expressed as *John has telephoned*, with the subject in

its usual place before the verb.

A somewhat subtler difference, again involving subjects, has to do with the formation of questions. In both languages, information questions (as opposed to yes/no questions) involve replacing some part of the sentence with a question word, and placing this at the front. Thus, we say *Who did you say that John met?* or in Italian, *Chi hai detto che Gianni ha incontrato?* If we want to question the subject of an embedded clause in English, though, things are not quite so straightforward. We cannot say \* *Who do you think that will come?* (where the expected answer might be something like *I think that John will come*); rather, we have to omit the *that*, and say *Who do you think will come?* In Italian, however, no such change is necessary: the corresponding question is simply *Chi credi che verrà?* where *che* corresponds directly to English *that*.

Here we see four differences between English and Italian with regard to subjects: (a) the possibility of omitting an understood subject; (b) the need for an ‘expletive’ subject with weather verbs and the like; (c) the ability of the subject to follow the verb in declarative sentences; and (d) the possibility of questioning the subject of an embedded clause by simply putting the question word at the front, without other changes. These do not obviously have anything to do with one another, but just as with the various aspects of negation considered earlier, it might turn out that they co-vary with one another in the grammars of languages.

When linguists looked at other languages with this in mind, they hoped to find that in general they would all pattern in each of the respects just described either with Italian, as is the case for Spanish, Romanian, and Greek; or else with English, as in French. Rumantsch, German, and the Nigerian language Edo. Within some families, especially Romance, this binary division appears to be correct. In a particularly elegant example from a group of languages unrelated to Romance, it turned out that in some forms of Levantine Arabic, all of the properties associated with empty subjects follow the Italian pattern, while another closely related language, the Arabic of Bani Hassan Bedouins in Jordan, patterns like English in these respects.

However, flies began to appear in the ointment before long. For one thing, languages like Chinese and Japanese allow pronouns to be omitted quite freely, with no special connection to subject position and no connection with the other phenomena characteristic of Italian. In some languages, subjects can be omitted under some circumstances (e.g., first and second person, as opposed to third person, in Finnish; with past tenses as opposed to the present in Hebrew) with no concomitant correlation with the other grammatical phenomena discussed.

These cases are apparently independent of the proposed generalization about omitted subjects that seems to be suggested by the comparison of English and Italian: rather than directly falsifying the proposed correlation, they appear simply to illustrate other circumstances under which subject pronouns can be left as ‘understood’. Other languages were soon uncovered, though, that seem closer to the heart of the matter and that suggest that the various differences between the grammars of English and Italian are not in fact systematically related. Icelandic, for example, is like English or German in that it generally does not allow the omission of subjects or their free placement in post-verbal position. Icelandic also has expletive elements that are required as the subjects of weather verbs. On the other hand, Icelandic allows subjects to be questioned in complement clauses without otherwise modifying the structure of the clause:

- (5) Hver heldur þú að framið glæpinn? who believe you that has committed the-crime Who do you believe has committed the crime?

A distinct problem for the proposed correlation is presented by the African (Bantoid) language Denya. This language is generally like Italian in all of the relevant respects except one: inversion of the subject to a position following the verb is not allowed. Close examination across a variety of languages has demonstrated that contrary to first impressions, the properties of grammars that are in question do not in fact co-vary in the consistent pattern we might have hoped to find.

Unification of the differences in subject behaviour between English

and Italian thus seems to fail the test of general applicability, and this particular example of the attempt to unite diverse differences between grammars under a single parameter has largely been abandoned. Battle continues to rage, however, between those who see the search for such large-scale connections ('macro-parameters') as the central goal in studying the range of the world's languages and those who suggest either that no such connections exist, or that our knowledge of the grammatical structure of human language is too fragmentary at present for us to identify them, and who thus prefer to focus on identifying small, discrete differences between grammars ('micro-parameters').

The resolution of the differences between these two programmes is surely important for a general theory of linguistic structure, but for our present purposes, it is not the central point to focus on. Regardless of whether linguistic variation eventually turns out to be characterizable in terms of a comparatively small number of macro-parameters, or only in terms of a much larger number of micro-parameters, or something in between, we can still say that the placement of a given language in terms of its character with respect to each possible dimension of grammar allows us to distinguish I-languages from one another. This is a way of individuating (and enumerating) the languages of the world that is firmly grounded in the nature of the languages themselves, and not in external considerations of social and political opinion. As suggested at the beginning of this chapter, we can regard the linguistic identity of an I-language system, expressed in terms of the full range of structural parameters revealed by linguistic research, as comparable in some ways to its genotype. By reducing the question of differences among languages to that of differences in their grammars, we provide them with an identity that does not share the problems of other approaches.

Even if we were to manage to reduce the number of *independent* dimensions of linguistic variation to the minimum number possible, though, we would surely still be left with a very substantial number of apparently irreducible parameters – almost certainly many more than ten or however many vary just within the languages of northern Italy. Of course, the implications of this result for the world as a whole must be based on a thorough study of the range and limits of possible

grammatical variation. All of those languages that a visitor from London might think of as forms of 'Italian' have a great deal in common, and there are many ways in which they are all distinct as a group from many other languages in many other parts of the world.

The number of possible grammatical systems expands exponentially as the number of parameters grows, and so even if we arrived at a set of only about 25 or 30, the number of possible languages that could be described becomes huge: well over a billion, on the assumption of 30 distinct parameters; 40 binary parameters yields over a trillion; and 100 would result in unimaginable numbers of possible languages. One of the advantages of thinking about I-languages in these terms is that it is obviously out of the question to imagine the language learner comparing all of these candidates for what is to be acquired. On the other hand, if what needs to be done in the acquisition of any particular system is simply to determine the values for a fixed, limited set of parameters, the task becomes a realistic one.

Obviously, not all of the possibilities provided by a fully adequate set of dimensions of variability will be actualized, but if the space of possible grammars is covered uniformly to anything like the extent we appear to find in northern Italy for the limited set of parameters in play there, the number of languages in the world on this understanding of the nature of the question must be much greater than the *Ethnologue's* 6,909.



## Chapter 7

# The diversity of signed languages

If you consult the *Ethnologue*'s index of language families, you will find that (as of the 2009 edition) it includes a family of 'Deaf Sign Languages', and that family represents some 130 of the total of 6,909 catalogued languages. This immediately raises a number of questions. Are the sign languages of the deaf really 'languages' and not simply collections of gestures supporting a minimal level of communication? Or are they simply ways of representing in another medium (manual/visual) some spoken language, such as English? Isn't sign language the same everywhere? If they really are languages, and different from one another, how much of the world's linguistic diversity do they account for? And is this 'family' a set of languages related in the way discussed in [Chapter 3](#), where we considered the phylogenetic approach to language identity? The answers to these and related questions may be surprising to those who have never had a signing deaf friend or otherwise been motivated to consider the matter seriously.

It is almost certainly the case that humans have communicated with one another by means of manual gestures for at least as long as language has been in existence. Indeed, one influential theory of the origin of language sees signing as the original medium, with vocalizations coming to accompany the gestures and eventually supplanting them entirely. This is not an uncontroversial view, of course: little about the origins of language can be said to be definitely established, since most of the relevant facts are of the sort that leave no trace in the fossil record and for which there are no interesting homologues in our primate (or other animal) relatives. But whether a link between signing and speech goes back to the beginnings of human language or not, the two modalities of linguistic expression are more closely linked than one

might have anticipated. There have surely been deaf people throughout human history, and there is every reason to expect that they have always communicated in the manual/visual modality available to them.

## **Spontaneously originated signed language**

Manual expression comes quite naturally to humans under circumstances when it is not effectively pre-empted by spoken language. This is shown by the behaviour of deaf babies born into households of hearing people, as most are: the spoken language of their parents and others around them is inaccessible to these infants, and in the absence of signers as models, they are cut off from the experience of language that is essential to their normal development. In this situation, quite remarkably, children do not just remain incommunicative. Instead, they invent their own ways of expressing at least some of their thoughts and desires through manual gestures.

These 'home sign' systems are of limited complexity, and certainly do not come to rival English and other natural languages in expressive capacity, but the signs that compose them are to a significant extent genuine symbols, as the words of a language are. More interestingly, perhaps, they are combinable into more complex assemblages of at least two or three gestures to make novel expressions, built up according to regular patterns. That is already a kind of expressiveness that is found nowhere in the communication systems of other animals, and which puts home signing definitely on the 'human language' side of a major conceptual divide. The age at which home signs emerge, and the age at which two to three sign combinations appear, are approximately comparable to the ages at which hearing infants produce words and basic combinations in acquiring a spoken language. This suggests that humans have a 'linguistic nature' which asserts itself even in inadequate environments, and that language is a by-product of the developing brain regardless of modality.

In isolation, spontaneously originated home sign systems remain quite limited, and disappear with their inventors, but in at least a few cases, we can see their much greater potential. A particularly fascinating and

well-documented example is provided by the experience of deaf individuals in Nicaragua since the late 1970s. Prior to that time, deaf people in Nicaragua had generally been marginalized, isolated as something of an embarrassment to their families, and not provided with an education. In 1977, however, a special education programme in Managua gathered some 50 or so children together, a number that grew over the following years and was further increased by the establishment of a vocational school for the deaf under the Sandinista government in 1980. These efforts brought deaf people from all over the country together into a more coherent community, and naturally they brought with them such individual home signing systems as they had.

The schools in Managua did not offer any instruction in a signed language such as American Sign Language (ASL), because there were no qualified signing teachers available in Nicaragua at the time, and the initial goal of the school was to provide basic teaching in Spanish. Before long, however, the teachers observed that the children were communicating with one another by manual gesture. Far from representing simply a failure on their part to learn Spanish, what was happening was that the linguistic richness of the children's home signing was growing rapidly through interchange with others. Subsequent generations of children arriving at the schools had available to them a means of communication that was already considerably developed beyond home signing, and they took that and developed it further. Rather strikingly, within a few generations, a complex sign language (known as *Idioma de Señas de Nicaragua*) has emerged. This language is quite independent of other languages, signed (like ASL) or spoken (like Spanish), and now serves for its users as a vehicle of communicative expression like any other.

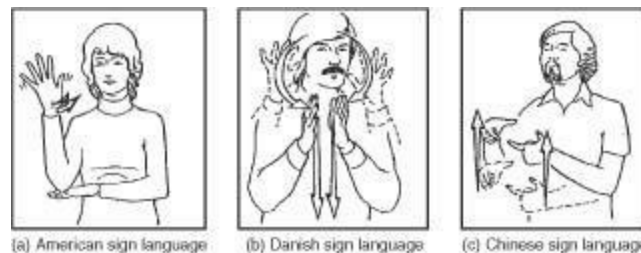
## **Some incorrect conceptions of signed language**

Since American Sign Language is the signed language that has been studied in much more detail than any other, most of the discussion below will be based on work on ASL. As a well-established language with a fairly large number of users (200,000–500,000), ASL has some properties that are not necessarily found in other signed languages, but

the differences appear to be ones deriving from its specific history and sociolinguistic position, and not matters of principle.

What do we mean when we say that signed languages are fully fledged languages in their own right? First of all, it is necessary to dispel a few widespread misconceptions. Signed languages are not simply pantomime, comparable to regular instances of what we do when we (or at least our ancestors) play(ed) ‘charades’. While it is true that many signs in a language like ASL have a basis in the iconic description of things in the world, most do not, or at least not one that is at all transparent. Different languages have different signs for even the most concrete aspects of the world, such as TREE, as illustrated in [Figure 10](#) (by convention, signs in ASL and other signed languages are referred to by means of an approximate English translation in capital letters).

Each of the signs in [Figure 10](#) can be seen to have some relation to the properties of a tree, but each plays the role it does because it is a word in a particular language. However evocative the ASL sign might seem, using that to refer to a tree when speaking Danish or Chinese sign language would be just as confusing as including a Danish or Mandarin Chinese word in an English sentence.



## 10. The sign TREE in three different signed languages

In fact, it is easy to exaggerate the degree of iconicity of signs. When presented with a series of ASL signs, for instance, naive English speakers (that is, those with no specific knowledge of ASL) perform essentially at chance in identifying their meaning, even on a task that involves a forced choice from among a small number of alternatives. Speakers of one signed language confronted with a similar task based on signs from another, unrelated language do not do much better.

In those few cases where we have evidence about the specific history of a signed language over time, we see that the forms of the signs in its lexicon change, just as we saw in [Chapter 3](#) that the phonetic shapes of spoken words change. An overall generalization about the nature of historical change in signs is that it is not constrained to maintain an iconic relation between the sign and its referent in the world. That is, many signs enter the language as highly iconic representations, but then tend over time to become less and less transparent. Once a sign constitutes a part of a language, its form is determined by the regularities of structure of that language, and not by its external meaning.

This is not to deny that iconicity plays a role in signed language: it is a valuable resource in a visual-spatial language, and that resource is exploited in ways that go beyond individual meanings. Even as iconicity fades over time in the lexicon, there are still ways that signed languages may reintroduce or revive it in their structure. What changes over time is the connection of iconicity to lexical content, on the one hand, and to grammatical structure, on the other. As sign languages age, and are transmitted over successive generations, the relationship becomes more complex and more layered, and is extended beyond the lexicon.

For example, verb classes in ASL are grounded in the underlying iconic relation of the sign to the action or state it represents. Verbs that are anchored to the body (including emotive or cognitive acts) are plain, in that they do not involve movement along a path determined by reference to their arguments (subject, object, etc.). Verbs of transfer, such as ‘GIVE’, involve movement on a path between positions corresponding to the two arguments (e.g., the subject and the indirect object of ‘GIVE’). A third class are neither plain or transfer, but spatial and locative, called spatial verbs, and serve to show very fine differences in spatial location and movement. The three classes differ systematically in the way their arguments are represented in the form of the sign. At this point, it can be said that iconicity has been remoulded into differences in the way verbs are inflected, and thus shifted from the meanings of individual signs into the grammar of the language.

Signed languages are also not just an alternative expression of the

spoken language of the surrounding community – the two are quite independent of each other. ASL, for instance, which is the principal signed language used in the United States, derives in substantial part from a language used in France in the 18th century, and is mutually comprehensible to some degree with French Sign Language. British Sign Language, on the other hand, the principal signed language in the UK, has quite different origins, and ASL and BSL are not mutually comprehensible, although the latter is quite close (though by no means identical) to the signed languages of Australia and New Zealand. The main signed language of Taiwan is quite close to Japanese Sign Language (as a consequence of earlier Japanese occupation of that island), but not mutually comprehensible with the signed languages of the Chinese mainland (including Hong Kong).

Anyone who has seen the common charts demonstrating a set of handshapes that can be used to represent the letters of the alphabet might get the impression that signing is just a matter of spelling out English (or some other language) with this mechanism. While it is true that words in a foreign language (English, or whatever) can be represented one letter at a time in this way within the system of ASL, that is quite peripheral as far as ordinary communication is concerned. Fingerspelling is quite unlike signing: it consists simply of the sequential production of a series of hand configurations, with none of the dynamic use of the visual-spatial medium that is essential to signing.

Furthermore, fingerspelling is a way of representing words in what is in fact a foreign language (e.g., English), and not a way to represent the signs of ASL itself. The same handshapes that are associated with letters of the alphabet in fingerspelling, along with others, are employed in signing, but in a completely different way, as formative elements of internally complex signs. The signed languages of East Asia also have ways of representing characters from languages like Chinese and Japanese, but it is not at all on that basis that normal communication takes place in these languages.

## **Signing as language**

If you have watched a television programme or film with a little box in one corner in which someone is signing, or attended a lecture with signers providing simultaneous ASL translation for deaf listeners, you may have wondered just how effective such translation is. The answer is: as effective as any translation from one language to another. ASL is perfectly capable of expressing anything that can be expressed in English, and a skilled translator can keep up with what is being said. In natural speech, the rate at which information is transferred (measured in terms of propositions per minute or the like) is roughly the same in signed and spoken language. In terms of overall efficacy, then, the two modalities are entirely comparable.

But signing certainly appears to be very different, so the question might linger as to just how similar language is in the two modalities. Again, the answer is that the two are much more alike than might be suspected, although there are some inevitable differences. The structure of a spoken language involves a number of layers of organization: a division of utterances into (individually meaningless) sounds, grouped together in systematic ways into words (*phonology*); an internal organization of words by which pieces of their form convey components of their meaning (*morphology*, as when we analyse the single word *bakers* as *bake*+*-er*+*-s*); and the system that combines words into phrases and clauses (*syntax*). These same levels of organization are characteristic of signed languages as well, and the parallels are quite precise.

In phonology, a specific language makes use of sounds drawn from a particular inventory, selected from the rather larger range of sounds found across all of the world's languages. These sounds, in turn, can be seen as organized on a small number of dimensions such as place and manner of articulation, activity of the vocal folds, and so on. Sounds are combinable within the scope of language-particular regularities that are based on this categorization. Thus English allows a combination of the fricative consonant [s] plus a voiceless stop consonant [p, t] or [k], possibly followed by a liquid [r, l] at the beginning of a word, but nothing more complicated than that (and even here, combinations with [t] followed by [l] are excluded). Georgian, in contrast, allows much

more complex combinations, such as the initial sequence of consonants in the (monosyllabic!) word [gvprt<sup>s</sup>k'vnis] 'he is bleeding us (financially)', but does not allow the initial combination [sp] found in English *spot*. Furthermore, when sound combinations arise that would otherwise violate the regularities of a language, specific adjustments come into play, as when a short vowel is inserted to avoid the disallowed combination *ch*+*[z]* at the end of *churches*.

A signed language like ASL has a similar organization, though obviously not one based on sounds. Ever since the pioneering work of William Stokoe (pronounced 'stow-key') in the early 1960s, it has been known that ASL signs can be broken down into distinct components such as handshape, location, movement, orientation, and a few others. Each of these dimensions provides a small number of values: the handshape of a sign, for instance, is chosen from a set that includes the elements of the manual alphabet found on those cards plus a small number of variants of those and a few others. This set of handshapes is a fact about ASL, and other signed languages make use of slightly different inventories. A specific sign involves a particular handshape, made at a particular location with a particular orientation, perhaps involving a particular movement, and so on. The ways in which these components can be combined are governed by language-particular rules based on their internal organization, although that organization is not as well understood as the internal structure of speech sounds.

In general, the components of signs are not meaningful in themselves. A complication here is the use of specific handshapes with specific meanings as *classifiers* in a particular subset of signs; exploring this matter would take us much too far afield for present purposes, but it does not compromise the general point being made. It is signs as wholes, like words, that are associated with meanings. The components of signs, and their language-particular modes of combination, thus function as an organization entirely comparable to the phonology of spoken languages. While it seems somewhat strange to talk about the 'phonology' of a language that does not involve sound, the structural similarities of organization of language in the two modalities have led to this as standard terminology.



Some spoken words, like *tree*, are simple in that no subpart of their form corresponds to a part of their meaning, but many others are morphologically complex, like *bakers*. This complexity comes in several flavours. *Compounding* allows for the creation of a new word out of existing ones, like *doghouse*; *derivation* allows for the formation of a new word on the basis of another, as when we form *inflatable* from *inflate*; and *inflection* is responsible for the variation among forms of the same word used under different syntactic conditions, such as the singular and plural forms of nouns (*dog/dogs*) or present and past tenses of a verb (*wait/waited*). Most such cases in most spoken languages involve the addition of extra, affixal material to mark a morphological difference, but in some the marking is by some other means. Thus, *man/men* is parallel to *dog/dogs* but marks the plural by a difference in the vowel, just as *sit/sat* marks past tense by such a difference. Derivational cases can also be like this: think of *breathe*, related to *breath* by a difference in the vowel and the final consonant, or the relation between *food* and *feed*.

ASL also has complex words, parts of which correspond to parts of their meaning. Unlike English (but like languages of the Athabaskan family, such as Navajo, Carrier, Apache, and about 30 others), ASL does not make basic distinctions of tense in its verbs, but rather distinctions of aspect, a category that characterizes the way an event or state transpires or is distributed across time. ASL verbs can be inflected for a large number of distinct aspects, primarily by varying the movement component of the sign. Derivational morphology is also present in ASL to some degree, for instance in the form of a suffix that can be added to verbal signs to make nouns referring to someone who carries out the action of the verb: an exact parallel to pairs like *bake/baker*. And ASL, like most signed languages, makes extensive use of compounding to create new signs out of combinations of existing ones.

A serious exploration of the nature of syntax in natural languages would require much more discussion than we can engage in here. Suffice it to say that signed languages like ASL organize signs into coherent groups (or *constituents*), and these combine with one another in

hierarchical fashion to make larger and larger units. Each of these constituents is a member of some grammatical category such as NP ('Noun Phrase'), VP ('Verb Phrase'), PP ('Prepositional Phrase'), S ('Sentence'), and so on, and the membership of a phrase in such a category determines its possibilities of combination with others. The organization of a simple sentence of English in this way is sketched below, with a certain number of additional details omitted for clarity's sake.

[<sub>S</sub>[<sub>NP</sub>A man[<sub>S</sub>[<sub>NP</sub>who] [<sub>VP</sub>likes [<sub>NP</sub>long-haired cats]]]] [<sub>VP</sub>offered [<sub>NP</sub>this hand] [<sub>PP</sub>to [<sub>NP</sub>Felix]]]]

The organization of sentences into constituents in ASL is not as transparent as is (often) the case in English, because signs can appear in somewhat freer order in ASL than English words. In this respect, ASL is similar to languages like Latin or most of the Aboriginal languages of Australia, which also allow rather free word order; but the grouping of words (or signs) into constituents at a more abstract level shows itself in various ways in the grammar of such languages.

Part of what allows for this freedom of order in spoken languages is the fact that the relations between words in the sentence may be indicated by other means. In Latin, the subject and direct object have different case forms (e.g., nominative versus accusative), allowing the listener to tell whether the dog is biting the man or *vice versa* regardless of the order of the words, unlike the situation in English. In ASL, as mentioned above, verbs of transfer and some others describing a relation between two individuals indicate the subject and object by the path of movement of the sign. As a result, expressions directly referring to these individuals can come in any order without impairing the intended meaning. With verbs that do not indicate their subject and object in this way, word order is much stricter.

Overall, then, signed languages display the same basic structural architecture as spoken ones. There are certainly differences, at least of degree, that derive from the difference between modalities. One important one concerns the balance between simultaneous signalling of

parts of the meaning of an utterance as opposed to sequential placement of meaning-bearing items (words or parts of words). Language in the oral/aural mode makes use of rather small articulators that can move quite rapidly from one configuration to another, but the amount of information that can be transmitted at any given time (the 'bandwidth' of the channel) is limited. As a result, complex meanings tend to be transmitted one piece at a time. In the manual/visual mode, in contrast, the articulators involved are considerably larger and comparatively slower, but on the other hand, the bandwidth is greater, and so the transmission of information tends to proceed in terms of fewer sequential elements, each of which carries rather more information. Signed languages tend to structures like that of *sat*, where both the verb and its tense are part of the same unit, rather than *seated*, where the two parts occur one after the other. This is a matter of degree, however, and not by any means an absolute difference between language types, since both signed and spoken languages display both sequential and simultaneous structures.

The implementation of signed language in its users is also entirely parallel to the way spoken languages are housed in their speakers. Signers show activity in the same brain regions (apart from effects due purely to differences in motor control of the specific organs involved) as speakers. Damage to left hemisphere areas associated with language produces aphasia in signers that are comparable to the effects of similar lesions in speakers. Damage to these areas does not, however, impact the ability of signers to produce or interpret non-linguistic gestures or pantomime, any more than it does for users of a spoken language.

The course of acquisition of a signed language is also just like what is observed in children learning a spoken language. Signing mothers tend to use a simplified and exaggerated style – 'Motherese' – when interacting with their deaf infant. The relation between the age at which a child is exposed to language and the degree of fluency acquired – so-called 'critical period' effects – can actually be much better documented in the case of signed languages than for spoken ones, because deaf children tend to have their first exposure to sign at a variety of ages, while hearing children virtually always are surrounded by spoken

language from birth.

The timing of appearance of the first meaningful signs, the first multi-sign utterances, and so on, is essentially indistinguishable from what is observed in the development of spoken language, provided the child has access to signed input. Even babbling has a direct correlate in the development of signed language, with deaf children in a signing environment passing through a stage in which they make meaningless gestures with their hands in a way that is interestingly similar to actual signing, and different from the random hand movements made by hearing children.

In every way, then, we are driven to recognize that signed languages like ASL are just *languages*, despite their implementation in a very different modality from that of spoken languages like English. And that means that any description of the world's linguistic diversity has to include them along with all the others.

## **How many signed languages are there?**

If signed language really was just a pantomimic depiction of the world, there would be no reason to expect to find more than one, but as soon as linguists began to look seriously at the question, it became apparent that this was far from being the case. Scientific investigation of ASL brought out the fact that other communities of deaf individuals had other signed languages, different from ASL just as the spoken languages of hearing communities differ.

Given what we know about the histories of signed languages, their diversity has an added importance beyond that of spoken languages. Because all of the signed languages that are currently in use have arisen independently and relatively recently in human history (though there have surely been many others in the past), we know that they are not in general related to one another. There are of course families of related languages, such as the ones derived largely from 18th-century French Sign (including ASL), the Japanese Sign Language family (including Taiwan Sign Language and Korean Sign Language), and some others.

But we can be reasonably sure that these families are not in turn related to one another within larger groupings, simply because as far as we know, until quite recently, the emergence and disappearance of signed languages was a rather episodic matter, tied to the contingencies of communities of deaf individuals and lacking any particular historical continuity.

Spoken languages, in contrast, may well all be related historically at some great remove. Researchers on the evolutionary origin of language generally find it plausible to suggest that language was invented only once, and that all modern spoken languages are thus in some way related, even if that relation can no longer be recovered (as noted in [Chapter 3](#)) because of limitations on the methods available for reconstruction. But if that is indeed the case, then any particular feature which we find across all spoken languages might just be a residue of a contingent fact about ‘proto-world’: a property of that original language which historical change happens never to have altered, but whose broad distribution has no significance beyond that.

For a parallel in the biological world, consider the fact that the mechanism of heredity across a vast range of life forms makes use of a common genetic code, based on the biochemistry of DNA and its associated organic molecules. Indeed, a small set of basic genes can be found across an extremely broad spectrum of life, resulting in the fact that humans share about 44% of their specific functional genes with fruit flies, and 26% with yeasts. We can attribute this commonality to the innovation of such a system in very early living organisms, and its conservation (with modification and elaboration, but leaving its essential nature intact) across millions of years of evolutionary change. We do not, in particular, need to say that this specific mechanism is a necessary characteristic of living things, because the historical record suffices to account for its widespread distribution today.

Consider what we would say, though, if we were to discover that life has emerged independently on Mars or some other planet, and that life there also makes use of the same genetic code as the system governing inheritance on Earth. In that case, it would seem that this really must be

something of a necessary condition of life, and not just one possibility out of many, one which happens to have been widely conserved over time.

When we find properties in common across signed languages, and especially when we find properties that are common to signed and spoken languages, there can similarly be no question of a comparatively trivial explanation based on shared history, because these languages do *not* share their histories. That means that any such universal characteristics of signed and spoken languages must have some other basis, and must in some way be necessary properties of language, at least as used by members of the species *Homo sapiens*.

In discussion of signed languages, it has proven useful to distinguish two broad classes of language. *Village* sign languages ‘arise in an existing, relatively insular community into which a number of deaf children are born’ Meir et al. (2010). Under these circumstances, deaf and hearing individuals grow up together, and are related, and communication by signing emerges as a system shared by the entire community, or at least a substantial part of it not limited to its deaf members. *Deaf community* sign languages, in contrast, arise when individuals related only by their deafness are brought together and develop a common language, as seen, for instance, in the Nicaraguan example discussed earlier in this chapter. In occasional cases, such as the development of ASL, such a language may spread well beyond the initial community, and become a vehicle of communication across a much wider society. This is the case for a number of signed languages identified with particular countries.

The difference between these two situations plays out to some extent in the structure of the languages themselves. It appears that the full linguistic architecture described above is somewhat slower to develop in ‘village’ languages, where the communicative needs of the language’s users are localized within a small community with largely shared social and cultural backgrounds and a substantial basis of shared assumptions that constrain and facilitate communication. It may be that under these circumstances, the full richness of a developed language is not as

essential to serve the language's purposes. Perhaps the very fact of shared usage between deaf and hearing people (who also have access to a spoken language in all known cases) also has a role to play. Whatever the explanation, it seems to be the case that 'village' languages such as the signed language of the Al-Sayyid Bedouins in the Negev desert, when examined closely, show rather less internal organization in domains such as phonology and morphology than Nicaraguan Sign Language, for example, despite being older by a few decades.

We must remember that in all cases, the actual time depth of a signed language is extremely limited – a few hundred years, at most, compared to the 60,000 to 100,000 years estimated as the time depth of spoken languages. It is hardly surprising that not all signed languages have completed their development within their short lives: what is remarkable is the extent to which some, such as Nicaraguan Sign Language, have done so.

The distinction in language type just made is not an absolute one. ASL, for example, developed in the early 19th century at a school for the deaf in Hartford, Connecticut. A major contributor to this language was the French sign introduced to the school by Laurent Clerc, a deaf man brought by Thomas Gallaudet to help establish this institution. Other important influences, however, included the signing of students from Martha's Vineyard in Massachusetts, where a significant incidence of hereditary deafness in the population had led to the development of a 'village' sign system there.

Most of the 130 signed languages identified in the current edition of *Ethnologue* are 'deaf community' languages, identified with specific nations or with politically defined sub-groups within a country, such as Catalan Sign Language. A few 'village' languages are included, such as Kaapor Sign Language in Brazil and Jhankot Sign Language in Nepal, but it is virtually certain that there are many more such languages that remain to be discovered.

As the study of signed languages develops, we can hope to explore the I-languages of their users in the same way we do for spoken languages, in order to determine the dimensions of variation on which

an adequate individuation of these languages could be based. Until we do so, we cannot really assess the proportion of the world's languages that signed languages represent. What we can be sure of, though, is that these systems represent an important part of the world's overall linguistic diversity, both in their numbers and in what they show us about the generality of the human capacity for language.



## Chapter 8

### Conclusion: the unity of human language

So how many languages *are* there in the world? On the basis of the various notions of how to individuate them that have been explored in previous chapters, we have come up with a fairly wide range of answers to this question, and no one of them seems overwhelmingly correct in comparison with the others. Yet another possible answer, and a rather surprising one, is suggested by Noam Chomsky:

It must be, then, that in their essential properties, languages are cast to the same mold. The Martian scientist might reasonably conclude that there is a single human language, with differences only at the margins.

For our lives, the slight differences are what matter, not the overwhelming similarities, which we unconsciously take for granted. No doubt frogs look at other frogs the same way. But if we want to understand what kind of creature we are, we have to adopt a very different point of view, basically that of the Martian studying humans.

What is the sense that Chomsky has in mind here of the nature of language? At various points in this book, we have compared the problem of individuating languages to similar issues in biology, but let us adopt the perspective of the biologist on our basic question. If we ask, for example, how many human visual systems there are, the answer might be ‘nearly seven billion’ one for each person on earth; or it might be ‘one’.

Vision in members of the species *Homo sapiens* has a range of characteristic properties. These include the precise structure of the lens, the distribution of rods and cones in the retina, including a fovea where visual acuity is particularly sharp and a blind spot where it is not, the

way the optic nerve projects to specific regions of the brain, the way various areas of cortex function to extract information about the external world from the pattern of activity in the cells of the retina, and so on. Some of these are physical structures that develop in the embryo; others are aspects of neural development that proceed in early life; and yet others are patterns of cortical activity that emerge on the basis of early visual experience. All this is genetically governed, however, and the important point is that vision is essentially uniform across the species.

The specific visual experience of individuals is immensely varied, though, and this may impact our visual processing in the way we interpret certain stimuli. For example, an important component of visual processing is our ability to recognize faces, but the implementation of this capacity depends heavily on experience; indeed, the emergence of the ability to distinguish faces bears some similarity to the development of language. Six-to-eight-month-old infants are known to have the capacity to produce and distinguish a wide range of sounds (including ones that do not occur in the language of their care-givers), but by the age of one year, this ability has seriously declined, and the child is primarily tuned in to the sounds that occur in the specific language(s) being acquired. Similarly, six-month-old infants can distinguish the face of one macaque monkey from another, but by the age of nine months they can no longer do this unless they have been presented regularly with pictures of monkeys.

Physically, the iris of some people's eyes is blue, in others brown, and for some people the irises of the two eyes are coloured differently. Some people have more acute vision than others, in a variety of senses. Experiments on animals of other species have shown that when the visual pathway is 're-wired' to connect to a different part of the brain, visual processing can develop there instead of in the region where it is normally found. We do not say, though, that any of these matters, or the specific range of faces that a person can easily identify, and so on, imply the existence of a variety of distinct visual systems. Rather, we say that other aspects of the environment and life experience of a person impact the way a genetically based system that is largely uniform initially develops in that individual, within the boundaries of limited phenotypic

variation that does not affect its overall functions.

Similarly, it makes sense to say that there is a single biologically based capacity for language that is essentially uniform across our species, but which develops in different ways in different individuals depending on their specific environment and experiences. Language will develop in roughly similar ways in individuals with broadly similar experiences: we think of these broad classes of comparable developmental paths as leading to the world's distinct languages. Nonetheless, they are all grounded in the same basic capacity from a biological point of view. In the absence of pathology, any child from any background, raised in the presence of speakers of any of the world's languages, will acquire that language and not one determined by the child's biological parents.

On the other hand, if we investigate the I-languages of speakers in enough detail, as suggested in [Chapter 6](#), we will almost certainly find some differences between virtually any two individuals in the precise form taken by their knowledge of language. In that sense, there might be said to be nearly seven billion languages in the world (and probably more, once we take multilingualism into account), but surely it is the uniformity of the human language capacity, in relation to the capacities of other species with which it might be compared, that is most significant.

## **The 'languages' of other species compared with ours**

When we look at the languages of the world, they may seem bewilderingly diverse. From the point of view of communication systems more generally, however, they are remarkably similar to one another. Human language differs from the communicative behaviour of every other known organism in a number of fundamental ways, all shared across languages.

The communicative behaviour of other animals is uniformly based on fixed sets of messages, essentially limited to responses to the here and

now. For any particular species, those messages are drawn from a limited list – typically fewer than a few dozen, for all species that have been seriously investigated – and not combinable with one another to express novel notions.

In nearly all cases, animal communication systems emerge in the individual with no dependence on the animal's experience: that is, they are innate, rather than learned, although in some instances, there may be a limited amount of 'fine tuning' possible with regard to the precise conditions under which a specific signal in the system is appropriate. Even where some learning occurs, of which by far the most robust examples are the songs of oscine birds (and, to some poorly understood extent, hummingbirds, parrots, and perhaps a few other bird species), the system that is acquired still has this same character of an essentially fixed list.

The most basic properties of human language show only the most general similarities to the communicative resources of other species. In parallel with every other communication system, language is deeply embedded in human biology, just as other animals' communication systems are part of their specific biological nature. The details of human language are learned, in the sense that experience affects which possibility from within a limited space will be realized in a given child. Song in oscine songbirds provides really the only parallel, while in most animals, including all of the other primates, communication is entirely innate, and develops in a fixed way that is independent of experience.

But where other species have fixed, limited sets of messages they can convey, humans have an unbounded range of things that can be expressed in language. And here there is no analogy with birdsong, since a bird's song always carries the same message even in species that learn a number of distinct variants. Even a bird that sings hundreds of different songs, like the nightingale (*Luscinia megarhynchos*), will use all of them to the same ends, to establish his claim to a territory in relation to his neighbours and to attract a member of the opposite sex.

Apart from the fact that it does not suffer from such a limitation in its scope of application, human language is also unusual in that its use is

voluntary, controlled mainly by cortical centres. In contrast, and perhaps with the exception of some ape gestures, other animals produce communicative signals under various sorts of non-voluntary (sub-cortical) control.

In more specific terms, human languages are distinctive in that individual messages are based on the combination of discrete elements into larger wholes. Any kind of meaningful combination is virtually unknown in the communicative behaviour of other species, and what flexibility such systems do show is usually obtained by variation of some properties of the signal along continuous dimensions, as in the famous ‘dances’ of European honeybees.

An important property of human language that is sometimes undervalued is what Charles Hockett christened ‘duality of patterning’: in a human language, as we saw above in [Chapter 7](#), individually meaningless sounds (or components of signs) combine to make meaningful words through the system of a language’s *phonology*. This is not just an ornament: phonological structure is what makes large vocabularies practical, by avoiding the requirement that every distinct word have its own completely distinct external shape, and replacing such holistic differences with systematic combinations of a limited number of distinct formative elements. Understanding the structure and emergence of the properties of phonology is thus another important task for those who would understand the evolutionary emergence of language, in addition to that of grammar.

Words in turn can have an internal organization by which parts of their form correspond to parts of their meaning (organized by the language’s *morphology*). They also combine by a completely different system (*syntax*) in a recursive, hierarchically organized fashion to make phrases, clauses, and sentences.

As we saw in [Chapter 7](#), these properties of the architecture of language that are found across spoken languages are also characteristic of signed languages like ASL. These too display duality of patterning, with individually meaningless formational elements combining to make meaningful signs through one system, and those signs combining in

hierarchical, recursive fashion into larger structures through another system. This sort of structure is thus characteristic of human language in the most general sense, and not just of individual languages or even spoken language alone.

With all due respect to the late Alex the parrot and Washoe the chimpanzee, and even Kanzi the bonobo, efforts to teach a system that genuinely displays the properties of a human language to other animals have not succeeded. This assertion is controversial in some quarters, but it would take us much too far afield to defend it here in any detail, and I simply refer to Anderson (2004a) for discussion (see the References). The bottom line is that there is no evidence that any other animal has the cognitive *capacity* to acquire and use a system with the core properties of human language: a discrete combinatorial system, based on recursive, hierarchical syntax and displaying two independent levels of systematic structure, one for the composition of meaningful units and one for their combination into full messages.

Really, though, there is no more reason to expect that our means of communication should be accessible to animals with a different biology than we should expect ourselves to be able to catch bugs by emitting short pulses of sound and listening for the echo, like a bat. In each case, the relevant capacity is grounded in the biology of the species, and the abilities of another animal with a different biology are unlikely to provide the required support.

By comparison with the communicative devices of zebra finches, honeybees, dolphins, or any other non-human animal, then, language provides us with a system that is not stimulus-bound and ranges over an infinity of possible distinct messages, on the basis of very specific organizational properties. No other system of communication found in nature has these properties. Language, and especially the combinatorial system of syntax, is quite unique in the animal world.

## **The human language faculty**

The overall architecture and essential properties of human language

(spoken or signed) do not vary across our species, in comparison with the communicative capacities of other animals. Can we be more specific than this, though, in defence of something like Chomsky's claim that 'there is a single human language'?

Humans, as opposed to other animals, are capable of learning a range of linguistic systems. Indeed, as we see in cases such as those of deaf children in hearing families, the drive to do so is extremely strong even in the absence of relevant experience. It probably makes more sense to think of the development of language as a matter of growth, like the descent of the larynx in early childhood and its further descent at puberty in males, rather than as a matter of learning of the sort illustrated by what happens when we come to know calculus or how to play the oboe.

What is the content of the systems that we are thus predisposed to develop? Recall the suggestion from [Chapter 6](#) that an individual's I-language, or knowledge of language, can be seen as having two components. Part of what we know stems from a set of *principles* that are common across languages, such as the principle governing the interpretation of pronouns tentatively formulated as (6), while the remainder implements a language's specific set of choices with respect to a set of dimensions, or *parameters*. Collectively, the set of principles taken together with the collection of parameters and their possible variation define the range of natural languages, and so can be said to characterize the human language faculty.

This account is sometimes seen as controversial, and associated with one particular school of thought in linguistics as opposed to many others, but such a view is surely misguided. In the worst case, even if diverse languages had nothing at all in common, that would simply show that the set of cross-linguistically valid principles was empty, and that the characterization of any particular I-language was entirely a matter of valuing the parameters of possible variation. The logical division of what languages have in common from what is variable among them is simply a framework for analysis, not a substantive empirical claim. The interest comes from what research on languages

can turn up that leads to the discovery of principles with real content, and also to an interestingly narrow range of parameters with a limited set of values.

In assessing the question of what is common across languages, we naturally look for properties that are true of every language we can examine. It is sometimes suggested, therefore, that a proposal about a potentially general property of language is falsified if we can find at least one language that does not display it. To reason in this way, however, is to misunderstand the nature of the language faculty, which is fundamentally a capacity to acquire and use systems of a particular kind. As such, it provides a sort of tool kit, a range of possibilities from which languages can select, but of course not all possibilities need to be implemented in every language.

When some particular language does not make use of one or another of the possibilities made available by the language faculty, that does not in itself show that the property in question is not part of the nature of language, seen as a general capacity of human beings. The fact that some language does not make use of click consonants comparable to those of the Khoi and San languages of southern Africa does not show that these are not among the possibilities universally *available* to languages (as opposed, say, to hiccups or laughter, which apparently never play a role in linguistic sound systems). Similarly, if a language does not form information questions (like *What did you say?*) by placing a question word at the front of the sentence, this does not show that a proposed universal condition on the kinds of sentence structure from which such words can be displaced in that way is incorrect or invalid.

Let us assume that linguistic investigation leads us to an interesting set of cross-linguistically valid principles and a framework of variation in terms of parameters. This theoretical apparatus can then be taken as a characterization of the human language faculty. A matter of great contentiousness in the study of language has been the extent to which the principles thereby uncovered, and the set of possible parametric choices provided, should be seen not only as *true* of language in general but also as *specific* to this domain. To what extent, that is, is the content



of the language faculty derived from the intersection of capacities that have relevance and application going beyond language? Some have maintained that to the extent one or another component of the mind that plays a role in language also plays a role elsewhere, it is not interestingly a part of the *language* faculty. On this line, if we could show that every aspect of language is related to something that is not limited to language, we would have shown that there is in fact no content to the notion of the language faculty itself.

This does not seem to be a particularly cogent way to think about what is at stake here. Of course, it is extremely interesting to ask how much of what we discover about language is domain-specific in the relevant sense. But even if we did show that our ability to develop and use a language was entirely cobbled together from other cognitive systems with other, broader purposes, that would hardly show that it was uninteresting to study the properties of the language faculty. After all, we do know that even if it results from the combining of other things, the human capacity for language taken as a whole is unique to our species, and so it is surely worthwhile to study its properties.

Furthermore, it is surely likely that even if other, more general facilities have been recruited for use in language, the particular form they take has been influenced by the role they play there. This is immediately evident when we consider the physical apparatus supporting speech. All of the parts of the vocal tract are also components of systems supporting other functions such as eating, drinking, and breathing, but it is clear that the specific form of the structures involved has been strongly influenced by suitability for speech, rather than just for these other functions.

The position of the human larynx low in the throat, with a corresponding lengthening and reshaping of the oral and pharyngeal cavities, is extremely advantageous from the point of view of rapid sequential production of a wide variety of speech sounds. It is much less advantageous from the point of view of the other functions of the vocal tract, since it leads to a greatly increased risk (*vis-à-vis* our other primate relatives) of choking on our food – a significant cause of death in

humans, but not in other apes. Although recruited from other functions, the human vocal tract has the form it has at least in part because of its role in speech. To suggest that because the vocal tract has other, broader purposes it is not part of the human capacity for language is just to miss the point.

The same line of reasoning is undoubtedly applicable to other human capacities, physical and cognitive, that have been recruited from broader purposes to play a role in language. This aspect of their function has surely had an effect in shaping them in particular ways to suit the novel purpose represented by the language faculty as well their original place in the life of *Homo sapiens*.

## **The search for universals of language**

A survey of the results of the search for the cross-linguistic content of the language faculty would require at least a book of its own (in fact, there are quite a few such books already in existence) and no summary will be attempted here. Two main approaches to the discovery of universals of language are to be found in the extensive literature on the topic, paralleling in some ways other pervasive divisions within the science of language.

One group of scholars, particularly associated with the late Joseph Greenberg, attempts to find such principles inductively. This involves the comparison of grammatical descriptions from (what is hoped to be) a representative sample of the world's languages, with an eye to discovering commonalities. Given the rather heterogeneous nature of such descriptions, it is commonly impossible to compare anything beyond the most superficial properties (e.g., usual order of subject, verb, and direct object; presence versus absence of tonal contrasts) in this way, and in fact some adherents of this approach have reached the conclusion that in fact there are *no* interesting universals of language to be found. Others have taken (what seems to the present author to be) a more constructive line, using the statistical tendencies that do emerge from these comparisons as the starting point for deeper investigation of the languages involved in the search for explanations both of the

regularities and of the apparent exceptions.

Another approach, particularly associated with Noam Chomsky, proceeds in quite a different way. When we can demonstrate that speakers of a language have knowledge of a certain sort, we can then ask how that knowledge might have arisen. In many cases, of course, we can see that the data available to the language-learning child provide everything necessary to account for what is acquired. Most overt properties of word form, inflectional shape, sound contrast, and much else fall into this category. In other cases, however, it appears that adequate evidence for what speakers can be shown to know is not available to the learner. Where this can be plausibly argued, then, another explanation is required, and the natural one is to say that the speaker's I-language takes the form it does because this is required by some substantive principle(s) of the language faculty. This logic is referred to as an argument from the *poverty of the stimulus*.

A substantial set of arguments of this sort concern things that speakers can be shown to know are *not* part of their language. This is because data about what is not possible are virtually never present in the child's input, except for things like the forms of irregular verbs ('no, Mary, not *taached* but *taught!*'). Nonetheless, many sequences of words in a language are immediately recognized by its speakers as impossible, despite their superficial conformity to its grammar.

For example, let us suppose that Fred has been charged to bring refreshments to the reception following this afternoon's seminar, and specifically, to bring tortilla chips, guacamole dip, and some Mexican beer. Now Fred, not being totally reliable, seems to have gotten confused, and from the looks of the table in the seminar room, has brought the wrong things. I can see the beer, and the chips, but I can't identify what's in that bowl over there. I can inquire about what has happened by asking (1-ai). However, I cannot say (1-aii) or (1-aiii). And a possible answer to my inquiry might be (1-bi), but surely not (1-bii) or (1-biii).

(1) a. i. What has Fred brought with those chips?

- ii. \*What has Fred brought and those chips?
- iii. \*What has Fred brought those chips and?
- b. i. *Hot salsa*, Fred brought with those chips, and not the guacamole we wanted.
  - ii. \**Hot salsa*, Fred brought and those chips, and not the guacamole we wanted.
  - iii. \**Hot salsa*, Fred brought those chips and, and not the guacamole we wanted.

What is going on here? The sentences that are marked (with a ‘\*’) as ungrammatical are not excluded because of their meaning, which is quite clear. Rather, the problem must relate to their form.

The question construction in English illustrated by the sentences of (1-a) involves placing the question word (e.g. *what*) at the beginning of the sentence, rather than in the position within the sentence where it would naturally be interpreted (for instance, as part of the direct object of the verb *bring*). Another construction in English, illustrated by the sentences in (1-b), allows us to place a word or phrase that is particularly focused (in a sense which can be made reasonably precise) at the front of the sentence, rather than in its natural position (cf. ‘The beer, Fred brought as we asked’). Both of these constructions involve relations of *displacement* between the element at the front of the sentence and some other position within it.

The problem with the impossible sentences in (1) is that they violate a principle of grammar known as the *coordinate structure constraint*, which can be roughly formulated as (2).

- (2) An element of a sentence cannot be displaced from a position within one branch of a coordinate structure (a structure of the form *X and/or/but Y* or the exact equivalent in other languages).

Principle (2) may actually be a subcase of a more general principle, but that refinement need not concern us here.

The point is that speakers recognize violations of (2) as ungrammatical in their language, in the face of the fact that language learners have no direct evidence for its validity. True, the data on which they base their learning will not have contained any violations of this principle, but there is a literally infinite range of sentence types that are not instantiated in the limited data available to the child, but which are produced and understood without problem on the basis of the I-language she develops. There is no reason to believe that the principles of formation of questions and focus sentences themselves are not quite general, and as such, they ought to accommodate the bad examples in (1). The problem does not reside in the intended meaning of the sentences, which is quite coherent (and substantially indistinguishable from the meanings of the good sentences (1-ai,bi)) but in their syntactic form.

Furthermore, if we were to consider a variety of other constructions involving displacement in English (which we will not do here), we would find that the restriction on such operations in (2) applies to all of them, suggesting that it is a general property of English. By an argument from the poverty of the stimulus, we conclude that this must be the consequence of a principle in the language faculty. And this conclusion is confirmed by the fact that the principle in (2) turns out to constrain displacement operations in every language that has been systematically investigated, to the extent they contain genuine coordinate constructions (note that 'X with Y' is a combination of X and a prepositional phrase, and not a coordinate construction in the relevant sense, even though such structures are employed in many languages to express the equivalent of 'X and Y').

A principle having (2) as a consequence is thus a good candidate for a universal of language, and the literature on syntax contains a number of other proposals of this general sort. The more content we can provide for the language faculty in such a way, the more plausible it is to point to human language in terms of its overall unity, rather than focusing on its superficial diversity.

Arguments like this become particularly persuasive when we can

show that the same principles that appear to be universally valid for spoken languages also apply in signed languages, despite the fact that no direct relation between these two classes of languages is possible, apart from their common manifestation of our species-specific capacity for language. And this is indeed the case when we explore the syntax of ASL. ASL has a question construction in which the question word appears at the front of the sentence, and also a construction with focused elements in initial position – comparable to the English constructions illustrated in (1). In both cases, the displacement of an element from a position in one branch of a coordinate structure is ungrammatical.

In support of this, consider the ASL sentences represented in (3). These are presented in a standard form, with capitalized English expressions representing the meaning of individual signs. Subscripts represent the reference of pronouns; with verbs, an initial subscript indicates the subject of an agreeing verb, and a following subscript the direct object. When a verb agrees with both its subject and its object, the sign is made on a path from the one to the other, as mentioned above in [Chapter 7](#).

(3) a. i. PRO<sub>1st</sub> MAJOR LINGUISTICS AND PSYCHOLOGY

I'm majoring in Linguistics and Psychology

ii. \*WHAT PRO<sub>2nd</sub> MAJOR LINGUISTICS AND?

\*What are you majoring in Linguistics and?

b. i. EXERCISE-CLASS<sub>1st</sub> HOPE SISTER SUCCEED PERSUADE MOTHER TAKE-UP

The exercise class, I hope my sister manages to persuade my mother to take (it)

ii. \*FLOWER<sub>i</sub> GIVE<sub>1st</sub> MONEY BUT<sub>j</sub> GAVE<sub>1st</sub>

\*Flowers, he gave me money but she gave me (those)

We can show that the ungrammaticality of the bad examples in (3) is indeed a consequence of the displacement they involve. This is because

ASL allows a question word to remain in place without being displaced (comparable to the English ‘echo question’ construction illustrated by *He brought chips and what?*), and in that case the question corresponding to (3-ai) is well formed.

#### (4) PRO<sub>2nd</sub> MAJOR LINGUISTICS AND WHAT?

You’re majoring in Linguistics and what?

The validity of principles like (2) across diverse languages and across linguistic modalities shows us that there is indeed non-trivial content to the human language faculty, content that is independent of the way this faculty is implemented in the I-languages of different individuals and language communities. While there is obviously much about this implementation that is dependent on the specific linguistic experience of the individual, there is also enough that is common to all members of our species to make it reasonable to focus our attention on the overall unity.

## **The unity of human language**

The particular linguistic system that each individual controls goes far beyond the direct experience from which the knowledge underlying it arose. And the principles governing these systems of sounds, words, and meanings are largely common across languages, with only limited possibilities for difference (such as the parameters described in [Chapter 6](#), as well as specific vocabulary items and their meanings). In its most basic aspects, human language is so different from any other known system in the natural world that the narrowly constrained ways in which one grammar can differ from another fade into insignificance. For a native of Milan, the differences between the speech of that city and that of Turin may loom large, but for a visitor from Kuala Lumpur both are ‘Italian’. Similarly, the differences we find across the world in grammars seem very important, but for an outside observer – say, Chomsky’s scientist from Mars, come to study terrestrial biology – all are relatively minor variations on the single theme of *human* language.

As the 11th edition of the *Encyclopedia Britannica* put it (in 1911),

[...] all existing human speech is one in the essential characteristics which we have thus far noted or shall hereafter have to consider, even as humanity is one in its distinction from the lower animals; the differences are in nonessentials.



## References

- Stephen R. Anderson, *Doctor Dolittle's Delusion: Animals and the Uniqueness of Human Language*. (New Haven: Yale University Press, 2004a).
- Stephen R. Anderson, How many languages are there in the world? 'FAQ' produced and distributed by the Linguistic Society of America, Washington, DC. 2004b; available on their website, <<https://lsadc.org/info/ling-faqs-howmany.cfm>>
- Mark C. Baker, *The Atoms of Language*. (New York: Basic Books, 2001).
- Claire Bower and Harold Koch (eds.), *Australian Languages: Classification and the Comparative Method*. (Philadelphia: John Benjamins, 2004).
- Diane Brentari (ed.), *Sign Languages*. (Cambridge: Cambridge University Press, 2010).
- Lyle Campbell and Verónica Grondona, Who speaks what to whom? Multilingualism and language choice in Misión La Paz. *Language in Society* 39 (2010): 617–46.
- Noam Chomsky, Language and mind: current thoughts on ancient problems (parts i and ii). *Pesquisa Linguística* 3(4) (1998). Lectures at Universidad de Brasilia.
- Bernard Comrie, Stephen Matthews, and Maria Polinsky, *The Atlas of Languages*. (New York: Facts on File, revised edn., 2003).
- Jerry A. Coyne and H. Allen Orr, *Speciation*. (Sunderland, MA: Sinauer Associates, 2004).
- Nicholas Evans, *Dying Words: Endangered Languages and What They Have To Tell Us*. (London: Wiley-Blackwell, 2010).
- Scott Freeman and Jon C. Herron, *Evolutionary Analysis*. (Upper Saddle River, NJ: Prentice Hall, 4th edn., 2007).
- Claude Hagège, *On the Death and Life of Languages*. (New Haven: Yale University Press, 2009).

- K. David Harrison, *When Languages Die*. (Oxford: Oxford University Press, 2007).
- K. David Harrison, *The Last Speakers*. (Washington: National Geographic, 2010).
- Edward Klima and Ursula Bellugi, *The Signs of Language*. (Cambridge, MA: Harvard University Press, 1979).
- Gaurav Mathur and Donna Jo Napoli (eds.), *Deaf around the World*. (Oxford: Oxford University Press, 2011).
- John McWhorter, *The Power of Babel*. (New York: Henry Holt, 2001).
- Irit Meir, Wendy Sandler, Carol Padden, and Mark Aronoff, Emerging sign languages. In Marc Marschark and Patricia Elizabeth Spencer (eds.), *Oxford Handbook of Deaf Studies, Language, and Education*, vol. 2, pp. 267–80. (New York: Oxford University Press, 2010).
- Robert McCall Millar, *Trask's Historical Linguistics* (London: Hodder Arnold, 2007).
- Carol Padden, Sign language geography. In Mathur and Napoli (2011), pp. 19–37.
- Sarah G. Thomason, *Language Contact: An Introduction* (Washington: Georgetown University Press, 2001).
- Nikolai Vakhtin, Copper Island Aleut: a case of language 'resurrection'. In Lenore A. Grenoble and Lindsay J. Whaley (eds.), *Endangered Languages: Current Issues and Future Prospects*, pp. 317–27. (Cambridge: Cambridge University Press, 1998).
- The blog 'Language Log' can be found at:  
<<http://languagelog.ldc.upenn.edu/nll/?p=3004>>

## Further reading

The problem of speciation in biology is addressed in a great many books over a long period. It forms part of the discussion in any basic text on evolution, as well as being the focus of specialized works such as Coyne and Orr (2004).

Among many published surveys of the world's languages, one that strikes a useful balance among comprehensive coverage, linguistic substance, and accessibility is Comrie, Matthews, and Polinsky (2003). Most of the numbers and other data about the distribution of the world's languages cited in the present book, unless otherwise indicated, come from the (2009) online edition of *Ethnologue* found at <<http://www.ethnologue.com/web.asp>>. This extremely important and comprehensive compilation has no real competitors. Various comments in this book that suggest caution in using its data and conclusions should not be taken to disparage the invaluable resource it provides.

A number of textbooks describe the techniques of historical linguistics, and the comparative method in particular. One of these is Millar (2007), a revision of a book originally by Larry Trask, but there are many others. The specific case of the Pama Nyungan family in Australia referred to in the text is dealt with in great detail in Bown and Koch (2004), in the context of the history, techniques, and limitations of the classic comparative method.

Recent years have seen a large number of books devoted to raising public awareness of language endangerment and the issues it raises. Two particularly good examples of this work are by K. David Harrison (2007, 2010). Harrison's description of Kallawaya is in the last of these works. Evans (2010) is another highly persuasive, if idiosyncratic, account, coloured to some extent by the author's aversion to formal linguistic

analyses. A French perspective on these issues is provided by Hagège (2009).

The UNESCO atlas of endangered languages referred to in the text can be found (at the time of writing) at <<http://www.unesco.org/culture/languages-atlas/en/atlasmap.html>>

The discussion of language contact in [Chapter 4](#) is based heavily on Sarah G. Thomason's 2008 Presidential Lecture to the Linguistic Society of America, 'Safe and Unsafe Language Contact'. Many of the same themes are treated in Thomason (2001).

McWhorter (2001) provides an accessible and engaging discussion of the importance of political and social factors (as opposed to purely linguistic ones), along with a number of other matters touched upon in the present work.

The approach to linguistic variation in [Chapter 6](#) is related to that of Baker (2001), a book that presents for a general audience a rather ambitious programme that aims to reduce differences between grammars to a small number of parameters. One does not have to believe in the particular set of parameters advocated by Baker, or even in the plausibility of a reduction on the scale proposed there, in order to take seriously the notion that differences among languages are fundamentally differences among grammars.

The study of signed languages has become a major research topic in linguistics since the 1960s, and there are many available sources from which to obtain more information. A highly accessible early classic of this literature is Klima and Bellugi (1979), which covers much of the basic evidence supporting the essentially linguistic nature of these languages, especially ASL. Two recent collections that deal more directly with the diversity of the world's signed languages are Brentari (2010) and Mathur and Napoli (2011). Padden (2011) provides a very useful introduction to the questions of sign language diversity addressed in [Chapter 7](#).

# Index

## A

accidental similarities between languages 36  
African-American students 66–7  
agreement, knowledge about 48  
Aleut language 52  
Altaic family 35  
American Sign Language (ASL)  
  iconicity in 96  
  misconceptions 94  
  morphology 100–1  
  origin 107  
  question constructions 122  
  syntax 101  
  *see also* signed languages  
Anderson, Stephen R. 113–14  
Anglo-Irish 67  
animals  
  and human language 113–14  
  language in 111–12  
anthropological interest in languages 48–50  
Arabic 19, 60  
‘arbitrariness of the sign’ 33–4  
ASL *see* American Sign Language  
assimilation and language change 26

## B

Basque (isolate language) 13

Bible, translations 10  
Big Nambas (Vanuatu language) 47–8  
bilingualism  
and contact 53–4  
and ethnicity 56  
and marriage 54–5  
biological connections between speakers of different languages 24  
biological species concept 4  
biological systems for speech 117–18  
Black Vernacular English 66–7  
Bloomfield, Leonard 31  
borrowing 52  
vs. inherited words 35  
boundaries, physical 25  
brain regions and signed languages 102–3  
Brazil 17  
Breton (Celtic language) 13  
British Sign Language 96  
*see also* signed languages  
Bulgarian and Macedonian 70

## C

Campbell, L. and V. Grondona 55  
Cape York peninsula, Australia 55  
case 101–2  
categorization of languages 3  
census data 59–60  
change  
and assimilation 26  
inevitability 23–4  
and misperceptions 26–8  
regularity 28–31  
sources 25–6  
systematic 32–3

Chickasaw 44  
children's words 35–6  
Chinese 19  
“dialects” 62–4  
Chomsky, Noam 108, 119  
classification of languages 23  
Clerc, Laurent 107  
clusters of languages 2–3  
common ancestor 20–21, 104–5  
and the comparative method 32–7  
common properties of languages 115–16  
signed languages 105  
communication among animals 111–12  
community identity 47  
comparative method 32–7  
competence, linguistic and knowledge 44–50  
competition between languages 51  
compounding 100  
connections between languages 20–1  
contact  
and bilingualism 53–4  
and borrowing 52  
and diversity 51  
and hybridization 51–2  
continuum of dialect changes 71–2  
coordinate structure constraint 121  
co-reference rule 82  
critical period for signed language acquisition 103  
cultural identity 47  
cultures, understanding 48–50

## D

Danish 19  
and Norwegian, mutual intelligibility 69

Darwin, Charles 3, 5, 23  
deaf children, home signing 92–4  
deaf community sign languages 106, 107  
death of languages 41–2  
and multilingualism 56–7  
and political policy 57–8  
Denya (African language) 88  
derivation 100  
*Descent of Man* (Darwin) 5, 23  
dialect  
vs. language 61–7  
and mutual intelligibility 68–71  
dialect continua 71–3  
displacement 121–2  
distinctions between languages  
distinguishing languages 2–3, 61–7  
and mutual intelligibility 68–71  
divergence of languages 23–4  
diversity of languages 1–2  
and contact 51  
in France 13, 16–17  
in North America 13, 16–19  
in Papua New Guinea 11, 13–15  
dominance and language shift 57–8  
duality of patterning 112–13

## E

Ebonics 66–7  
E-language and I-language, distinction 76–7  
*Encyclopedia Britannica* (1911) 10, 124  
endangered languages 39–41, 42–3  
and loss of knowledge 44–50  
preservation 50–1  
endangered species 38–9, 43–4



English  
French influence on 25  
historical approach 22  
and Italian, subject parameters 85–8  
and Latin, comparison 33  
past-tense verbs 77–8  
phonology 98  
sentence construction 120–3  
verb agreement in 122–3  
Estonia, multilingualism in 56–7  
ethnicity and multilingualism 56  
*Ethnologue* 11, 13, 20, 59–60, 68  
exceptionless sound laws 28–31  
exogamy 54–5

## F

families of languages 11  
largest 20  
signed languages 91, 104  
and sub-families 34–5  
field methods in linguistics 49  
fingerspelling 97  
France, linguistic diversity 13, 16–17  
Freeman, S. and J. C. Herron 4  
French, influence on English 25  
French Sign Language 96

## G

generational change 28  
genetic code, common 105  
genetic relationships between languages 24–5  
genetic species concept 4, 77  
genotypes and phenotypes 76–7  
geographical boundaries 25

geographical changes, continuum of dialect 71–2  
Georgian 36  
phonology 98–9  
Gothic 29  
grammar  
knowledge of 48, 77–80  
parameters 82–5  
rules 80–2  
grammatical sentences, knowledge of 120–3  
Greek 29  
Greenberg, Joseph 118  
Grimm's Law 29–31  
Guatemala 17  
Guugu Yimidhirr 55

## H

handshapes in signed language 99  
Harrison, K. David 45  
healing, knowledge in specific languages 45–6  
Hebrew 41–2  
Hindi and Urdu 63  
historical approach to languages 22–3  
Hockett, Charles 112  
home sign systems 92–4  
Hungarian 19, 24–5  
hybridization 51–2

## I

Icelandic 87–8  
iconicity in signed languages 96  
identifying languages 3, 5  
identity 47  
and multilingualism 56  
I-language 77–80, 119

distinguishing 83–5, 89–90  
and E-language, distinction 76–7  
rules 80–2  
of signed languages 107  
India  
bilingualism in 53  
multilingualism in 56  
number of languages 1–2  
and Pakistan 63  
individual differences in language 108–11  
Indo-European languages 11, 12  
inflection 100  
inherited words vs. borrowed words 35  
innateness of language 93, 113–14, 115–18  
International Union for the Conservation of Nature (IUCN) 43  
invasions, influence on languages 25  
inventiveness of language 112  
Italian  
and English, subject parameters 85–8  
and Piedmontese 83–4

## K

Kalabari and Nembe (African languages) 70–1  
Kallawaya (Bolivian language) 45–6  
Kannada (Dravidian) 56  
Kayardild (Australian language) 49–50  
kinship terms 49–50  
knowledge  
of grammar 48, 77–80  
preservation of 44–50

## L

Labov, William 31  
language acquisition,

innateness 93, 113–14, 115–18  
language faculty 115–18, 119, 122  
‘Language Log’ 1–2  
Latin 29  
and English, comparison 33  
levels of organization in signed languages 98–102  
linguistic imperialism 57  
local traditions 47

## M

Maasai (Nilotic language) 57  
Macedonian and Bulgarian 70  
macro- and micro-parameters 88–9  
Maharashtra state, India 56  
Mandarin 19, 62–4  
manual gestures *see* signed languages  
Marathi (Indic language) 56  
marriage and multilingualism 54–5  
Mayr, Ernst 72  
Mbabaram (Australian language) 36  
meaningfulness 113  
medicine, knowledge in specific languages 45–6  
Mednyj (Copper Island) Aleut 52  
Michif (“mixed” language) 52  
micro- and macro-parameters 88–9  
Milošević, Slobodan 65  
minimal differences between grammars 85  
Misión La Paz, Argentina 55  
misperceptions and language change 26–8  
morphological species concept 4, 77  
morphology  
and meaningfulness 113  
of signed languages 100–1  
Motherese and signed language acquisition 103

- multilingualism
  - and contact 53–4
  - and ethnicity 56
  - and marriage 54–5
- mutual intelligibility 62
- and distinguishing languages 68–71
- of signed languages 96–7
- over time 73

## N

- national languages and multilingualism 53–4
- Navajo 19
- negation and word order 83–4
- Nembe and Kalabari (African Languages) 70–1
- Neolinguisti 29
- Nicaragua, deaf children in 93–4
- Norman French 25
- North America
  - endangered languages 39–40
  - linguistic diversity 13, 16–19
- Northern Spotted Owl 38–9
- Norwegian and Danish, mutual intelligibility 69
- null subject parameter 85–8
- number of languages 1–2, 10–11
  - in France 13
  - in North America 16–19
  - and number of parameters 89–90
  - in Papua New Guinea 14–15
- signed languages 21, 104
- numbers of speakers 19–20
- difficulty of counting 59–61
- of North American languages 17, 19
- nursery words 35–6

## O

Oakland California School Board 66  
official languages and multilingualism 53–4  
Old English 29, 30  
onomatopoeic words 35

## P

Pakistan and India 63  
Papua New Guinea, linguistic diversity 11, 13–15  
Paraguay, bilingualism in 53  
parameters  
and distinguishing languages 89–90  
of grammar 82–5  
macro- and micro- 88–9  
principles 115–16  
subject 85–8  
past-tense verbs  
in English 77–8  
in Scots 78–80  
phenotypes and genotypes 76–7  
phonological anomalies 47–8  
phonological change, regularity 28–31  
phonology 98–9  
internal knowledge about 78–80  
and meaningfulness 113  
of signed languages 99–100  
phylogenetic species concept 4  
physical boundaries 25  
Piedmontese and Italian 83–4  
place of articulation and language change 26  
plants, knowledge about in specific languages 45–6  
political policy  
and distinguishing languages 61–7  
and language death 57–8

- and mutual intelligibility 70–1
- poverty of the stimulus 119, 122
- predisposition to learning language 115–18
- prestige and language shift 57–8
- principles and parameters 115–16
- pronouns
  - rules of 80–2
- subject 85–8
- pronunciation, variation in 26–7

## Q

- question constructions in American Sign Language (ASL) 122

## R

- replacement of languages 57–8
- reproduction
  - of languages 41
  - of separate species 68
- ring species 71–3
- Romance languages spoken in France 13
- Ross, J. and R. Langacker 82
- rules, grammatical 80–2
- Rumantsch 53–4
  - number of speakers 60

## S

- salamander, subspecies 72
- Sanskrit 29, 30
- Saussure, Ferdinand de 33
- Scandinavian languages, mutual intelligibility 69–70
- Scots dialect/language 78–80
- sentence construction 120–3
- Serbo-Croatian 65

- shift of languages 57–8
- signed languages
  - in babies 92–3
  - and brain regions 102–3
  - common properties 105
  - differences between 94–5, 96–7
  - iconicity in 96
  - misconceptions 94
  - morphology 100–1
  - number 21, 104
  - phonology 99–100
  - possible origin 91–2
  - question construction 122
  - structure 102
  - syntax 101
  - translation from spoken language 97–8
  - types 106–7
  - word order 101–2
- SIL International 10–11
- similarities between languages, accidental 36
- simplification of language and language change 26
- Sinitic language family 62–4
- social anthropology 48–50
- social variation 31
- songbirds 111–12
- sound change, regularity 28–31
- sounds in language, anomalies 47–8
- speaker numbers 19–20
  - difficulty of counting 59–61
  - of North American languages 17, 19
- species
  - dialect continua 71–3
  - endangered 38–9, 43–4
  - identifying 3–5
  - language in other animals 111–12



- as metaphor for languages 5
- reproduction between 68
- spoken language
  - connections 20–1
  - translation into signed language 97–8
- Stokoe, William 99
- structure of signed languages 102
- subject parameters 85–8
- Summer Institute of Linguistics (SIL International) 10–11
- Swiss German 54, 68
- Switzerland
  - languages spoken 59–60
  - multilingualism in 53–4
  - Swiss German in 68
- syntax
  - knowledge of 48, 120–3
  - and meaningfulness 113
  - of signed languages 101
- systematic change 32–3

## T

- Tagalog 19
- tense marking, knowledge about 48
- Toronto *Globe and Mail* 1–2
- traditional cultures 47
- Trans-New Guinea Phylum 11, 14
- translations
  - of the Bible 10
  - from spoken language into signed language 97–8

## U

- uniformity of language 110–11
- universal conditions 116
- universals for language 118–19, 122

Urdu and Hindi 63  
USA *see* North America

## V

Vakhtin, Nikolai 52  
variation  
  between grammars 82–5  
  in pronunciation 26–7  
  social 31  
verbs  
  agreement in English 122–3  
  in American Signed Language (ASL) 96  
  past-tense 77–80  
Vernacular English (African-American) 66–7  
Verner, Karl 30  
Viking invasion 25  
village sign languages 106–7  
visual systems, number of 109–10  
Votic (Fenno-Ugric language) 56–7  
vowels, dropping and adding 27

## W

Weinreich, Max 61  
Western Desert, Australia (language continuum) 71  
Wichita (Caddoan language) 48  
word order 120–23  
  and negation 83–4  
  in signed languages 101–2  
writing systems in related languages 62–5

## Y

Yaaku (Cushitic language) 57  
Yiddish 61

Yugoslavia 65