WARNING CONCERNING COPYRIGHT RESTRICTIONS: The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproduction of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If electronic transmission of reserve material is used for purposes in excess of what constitutes "fair use", that user may be liable for copyright infringement.

CONTEMPORARY LINGUISTICS

AN INTRODUCTION

Fourth Edition

Edited by
William O'Grady
University of Hawaii at Manoa
and
John Archibald
University of Calgary

U.S. edition prepared by
Mark Aronoff
State University of New York at Stony Brook
and
Janie Rees-Miller
Marietta College

Bedford/St. Martin's
Boston ♦ New York

For Bedford/St. Martin's

Senior Editor: Talvi Laev
Production Editor: Harold Chester
Senior Production Supervisor: Cheryl Mamaril
Marketing Manager: Richard Cadman
Art Direction and Cover Design: Lucy Krikorian
Text Design: Kyle Gell
Graphics: Allan Moon, Valentino Sanna
Cover Art: The Stock Illustration Source, Inc., Joel Nakamura
Composition: the dotted i
Printing and Binding: RR Donnelley & Sons Company

President: Charles H. Christensen
Editorial Director: Joan E. Feinberg
Editor in Chief: Nancy Perry
Director of Marketing: Karen R. Melton
Director of Editing, Design, and Production: Marcia Cohen
Managing Editor: Erica T. Appel

Library of Congress Control Number: 00-104679

Copyright © 2001 by Bedford/St. Martin's
All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, except as may be expressly permitted by the applicable copyright statutes or in writing by the Publisher.

Manufactured in the United States of America.

6 5 4 3 2 1
f e d c b a

For information, write: Bedford/St. Martin's, 75 Arlington Street, Boston, MA 02116 (617-399-4000)

ISBN: 0-312-24738-9

Acknowledgments

Acknowledgments and copyrights appear at the back of the book on page 734, which constitutes an extension of the copyright page.

Convinced at once that, in order to break loose from the beaten paths of opinions and systems, it was necessary to proceed in my study of man and society by scientific methods, and in a rigorous manner, I devoted one year to philology and grammar; linguistics, or the natural history of speech, being, of all the sciences, that which best suited the researches which I was about to commence.

–PIERRE-JOSEPH PROUDHON, *What Is Property?* (1840)

eight

THE CLASSIFICATION OF LANGUAGES

Aleksandra Steinbergs

Everything it is possible for us to analyze depends on a clear method which distinguishes the similar from the not similar.
– LINNAEUS, *Genera Plantarum* (1754)

OBJECTIVES

In this chapter, you will learn

- how different languages can be classified according to similarities in their phonology, morphology, and syntax
- how languages are related to one another genetically
- what the major language families of the world are, and some representative languages of each family
- how language families may be grouped into larger phyla

In the world today there are thousands of different languages, each with its own sound patterns, grammar, and vocabulary. Regardless of how different these languages are, they have important similarities that allow linguists to arrange them into a fairly small number of groups. This chapter describes the methods of classification linguists use, and some of the findings that have resulted from this type of research.

1 SOME PRELIMINARIES

We will begin by considering two topics—the problem of distinguishing between a language and a **dialect**, and the chief methods of language classification used in linguistics today.

1.1 Dialect and Language

It is often difficult to determine whether two linguistic communities speak different languages or merely different dialects of the same language. One test that linguists use to decide this involves the criterion of **mutual intelligibility**. Mutually intelligible varieties of the same language can be understood by speakers of each variety. According to this criterion, the English of Toronto, the English of Milwaukee, and the English of London qualify as dialects of the same language. On the other hand, if two speakers cannot understand one another, then linguists normally conclude that they are speaking different languages. The Italian of Florence and the French of Paris are examples of varieties of speech that are not mutually intelligible.

Political, cultural, social, historical, and religious factors frequently interfere when determining linguistic boundaries. (In fact, it is sometimes said that a language is just a dialect with an army and a navy!) For example, Serbs and Croats, with their different histories, cultures, and religions, often claim that they speak different languages. However, even though they use different alphabets, Serbian and Croatian are actually mutually intelligible dialects of the same language, which linguists call Serbo-Croatian. In contrast, we often speak of Chinese as if it were a single language, even though it is actually a number of individual, mutually unintelligible languages (Cantonese, Mandarin, Hakka, and so on), each with a multitude of dialects of its own.

In addition to the problems presented by these nonlinguistic considerations, complications also arise when we try to divide a continuum of mutually intelligible dialects whose two end points are not intelligible. Dutch and German, for example, are mutually intelligible around the border area between Germany and Holland; however, the Dutch of Amsterdam and the German of Munich are not. Similarly, Palestinian Arabic and Syrian Arabic are mutually intelligible, but Moroccan Arabic and Iraqi Arabic are not.

Taking these considerations into account, how many languages are there in the world today? The best available estimate places the current figure at about six thousand five hundred. However, many of these languages have only a few hundred speakers and many others are in grave danger of demise, as indigenous peoples throughout the world lose their traditional cultures and homelands. Indeed, according to one estimate, only about three hundred of the world's languages have a secure future. The threat of such a massive loss of the world's linguistic diversity is of great concern to linguists, many of whom are actively involved in recording and studying languages on the verge of extinction and in finding ways to improve the prospects for endangered languages.

Section 3 of this chapter presents an overview of a few hundred languages and the families to which they belong. First, however, we will turn our attention to some of the methods that are used for classifying languages into a manageable number of types.

1.2 Types of Classification

Within the field of linguistics, three different approaches to language classification are used.

Genetic classification categorizes languages according to their descent. Languages that developed historically from the same ancestor language are grouped together and are said to be **genetically related**. This ancestor may be attested (that is, texts written in this language have been discovered or preserved, as in the case of Latin), or it may be a reconstructed proto-language for which no original texts exist (as is the case for Indo-European). Genetic classification is discussed further in Section 3.

Although genetically related languages often share structural characteristics, they do not necessarily bear a close structural resemblance. For example, Latvian and English are genetically related (both are descended from Indo-European), but their morphological structure is quite different. An English sentence like *It has to be figured out* can be expressed in Latvian by a single word.

1)

ja:izgudro
(one) must out figure (it)
'One must figure it out.'

Of course, Latvian and English are very distantly related, and languages that are more closely related will typically share a larger number of similarities. On the other hand, it is also necessary to recognize that even languages that are totally unrelated may share some structural similarities. Thus, English and Swahili, which are unrelated, both employ subject-verb-object word order in simple declarative sentences.

2)

Maria anapenda Anna
'Maria likes Anna.'

For this reason, another approach to language classification is useful. Known as **linguistic typology,** it classifies languages only according to their structural characteristics, without regard for genetic relationships. Thus typologists might group together languages with similar sound patterns or, alternatively, those with similar grammatical structures. Typological studies also endeavor to identify **linguistic universals**—that is, structural characteristics that occur in all or most languages. We discuss linguistic typology further in Section 2.

Finally, **areal classification** identifies characteristics shared by languages that are in geographical contact. Languages in contact often borrow words, sounds, morphemes, and even syntactic patterns from one another. As a result, neighboring languages can come to resemble each other, even though they may not be genetically related. Because of space considerations, this chapter will not deal with areal classification specifically; however, borrowing is discussed in Sections 1.2 and 5.1 of Chapter 7.

2 TYPOLOGICAL CLASSIFICATION

As just noted, the classification of languages according to their structural characteristics is known as linguistic typology. Typological studies group together languages on the basis of similarities in their syntactic patterns, morphological structure, and/or phonological systems. An important area of research within the study of linguistic

typology is the search for linguistic universals. Structural patterns and traits that occur in all languages are called **absolute universals**, while those that simply occur in most languages are known as **universal tendencies**.

Many typological generalizations involve **implicational universals**, which specify that the presence of one trait implies the presence of another (but not vice versa). For instance, languages with fricative phonemes (such as /f/ and /s/) will also have stop phonemes (such as /p/ and /t/), although the reverse is not necessarily true.

Another way to analyze linguistic universals is through **markedness theory**. Within this theory, **marked traits** are considered to be more complex and/or universally rarer than **unmarked** characteristics. In addition, a marked trait is usually found in a particular language only if its unmarked counterpart also occurs. Thus, markedness theory is closely related to the study of implicational universals.

An example can provide some clarification of these terms. Nasalized vowels are said to be marked, while nonnasalized (oral) ones are said to be unmarked. Phonologically, oral vowels can be considered less complex: oral vowels allow the airstream to exit only through the mouth, while nasalized vowels allow air to escape from both the mouth and the nose.

Cross-linguistically, we find that all languages have oral vowels, while only some languages have nasalized vowels. Even in the languages that have both, there are usually fewer nasalized vowels than oral ones. Thus, nasalized vowels (which are considered to be marked) are both rarer and phonologically more complex than (unmarked) oral vowels.

The following sections present some of the typological generalizations and universals that have been proposed in the areas of phonology, morphology, and syntax.

2.1 PHONOLOGY

In this section, we represent all vowel and consonant systems phonemically. This simplifies their presentation; note, however, that the exact phonetic realization of these systems may vary in the individual languages.

Vowel systems

Languages are often classified according to the size and pattern of their vowel systems. The most common vowel system has five phonemes—two high vowels, two mid vowels, and one low vowel (see Figure 8.1). The front vowels are unrounded, as is the low vowel, and the back vowels are rounded.

i		u
e		o
	a	

Figure 8.1 The most common vowel system

About half the world's languages, including Basque, Hawaiian, Japanese, Spanish, and Swahili, have such a system.

The majority of the world's other languages have vowel systems with three, four, six, seven, eight, or nine different vowels (disregarding contrasts based on length or nasalization, which can double or triple the number of phonemic vowels). Languages with fewer than three or more than nine distinctive vowels are rare. Some typical vowel systems are presented in Figure 8.2.

i	u	i		i		u
		e	o	e	ə	o
a		a			ʌ	
					a	
Three-vowel system		Four-vowel system		Seven-vowel system		
Gudanji (Australia)		Navajo (Arizona)		Ge'ez (Ethiopia)		

Figure 8.2 Common vowel systems

Analysis of many languages has led to the discovery of a number of universal tendencies pertaining to vowel systems. Some of these tendencies are listed here, along with a description of the most commonly occurring vowels.

- The most commonly occurring vowel phoneme is /a/, which is found in almost all of the languages of the world. The vowels /i/ and /u/ are almost as common as /a/.
- Front vowel phonemes (/i, e, ɛ, æ/) are generally unrounded, while nonlow back vowel phonemes (/ɔ, o, u/) are generally rounded.
- Low vowels (/æ, a, ɑ/) are generally unrounded.

Although English has an above-average number of vowels, they all conform to the above tendencies. Thus, English has only front unrounded vowels, all the low vowel phonemes are unrounded, and all of the back, nonlow vowels are rounded. The English vowel system is represented in Figure 8.3.

i		u
ɪ		ʊ
e		o
ɛ	ʌ	(ɔ)
æ		ɑ

Figure 8.3 The English vowel system

The relationship between contrasting vowel types (such as oral versus nasal, and long versus short) can also be expressed in terms of implicational universals, since

the presence of one vowel phoneme type implies the presence of another (but not vice versa).

- If a language has contrastive nasal vowels, then it will also have contrastive oral vowels. For example, French contrasts different nasal vowels (/lɔ̃/ 'long' and /lɑ̃/ 'slow'), and contrasts oral vowels with nasal vowels (/la/ 'weary' and /lɑ̃/ 'slow'). Predictably, French also contrasts different oral vowels, as in /klo/ 'shut' and /klu/ 'nail'. English shows contrasts among oral vowels but does not contrast nasal vowels with oral vowels. There are no contrasts in English like /bɔt/ 'bought' and */bɔ̃t/.
- If a language has contrasting long vowels, then it will also have contrasting short vowels. For example, Finnish shows contrasting long vowels and, predictably, contrasting short vowels (see Table 8.1).

Table 8.1 Finnish vowel contrasts

Long versus long	/vi:li/	'junket'	/va:li/	'election'
Short versus short	/suka/	'bristle'	/suku/	'family'
Short versus long	/tuli/	'fire'	/tu:li/	'wind'

The reverse is not necessarily the case. English has contrast between short vowels (/bɪt/ vs. /bɛt/) but does not contrast long vowels with short ones, since vowel length in English is predictable.

Consonant systems

It is not particularly useful to classify languages according to the number of consonants that they contain, since languages may have as few as eight consonant phonemes (as in Hawaiian) or more than ninety. (!Kung, a language spoken in Namibia, has ninety-six consonant phonemes.) Nevertheless, typological analysis of consonant systems has produced a number of well-substantiated universals:

- All languages have stops.
- The most common stop phonemes are /p, t, k/. Very few languages lack any one of these, and there are no languages that lack all three. If any one of these three stops is missing, it will probably be /p/; for example, Aleut, Nubian, and Wichita have no /p/ phoneme. The most commonly occurring phoneme of the three is /t/.
- The most commonly occurring fricative phoneme is /s/. If a language has only one fricative, it is most likely to be /s/. It is the only fricative found in Nandi (a language of Kenya) and Weri (a language of New Guinea). The next most common fricative is /f/.
- Almost every known language has at least one nasal phoneme. In cases where a language has only one nasal phoneme, that phoneme is usually /n/ (as in Arapaho, spoken in Wyoming). If there are two contrasting nasals, they are normally /m/ and /n/.

- The majority of languages have at least one phonemic liquid. However, a small number of languages have none at all—for example, Blackfoot, Dakota, Efik (spoken in Nigeria), and Siona (found in Ecuador). English, of course, has two: /l/ and /r/.

Consonant phonemes are also subject to various implicational universals:

- If a language has voiced obstruent phonemes (stops, fricatives, or affricates), then it will also have voiceless obstruent phonemes. The reverse is not necessarily true; for example, Ainu (a language of northern Japan) has only voiceless obstruent phonemes: /p, t, k, tʃ, s/.
- Sonorant consonants are generally voiced. Very few languages have voiceless sonorants; those that do always have voiced sonorants as well. For example, Burmese contrasts voiced and voiceless nasals and laterals.
- If a language has fricative phonemes, then it will also have stop phonemes. There are no languages that lack stops; however, there are some languages that lack fricatives. For example, Gilbertese (Gilbert Islands), Kitabal (eastern Australia), and Nuer (southeastern Sudan) have no fricatives.
- Languages that have affricates will also have fricatives and stops. This is not surprising, since an affricate is, in essence, a sequence of a stop followed by a fricative. However, many languages lack affricates altogether. Note that while European French, for example, has fricatives and stops but no affricates, English has all three.

Suprasegmental systems

Languages can also be classified according to their suprasegmental (or prosodic) type. Languages that use pitch to make meaning distinctions between words are called **tone languages**. (The phonetics and phonology of tone were introduced in Chapters 2 and 3.)

A great many of the world's languages are tone languages. Mandarin Chinese, for instance, has four contrastive tones (see Table 8.2).

Table 8.2 Tone contrasts in Mandarin Chinese

High tone	dā	'build'
Low rising tone	dá	'achieve'
Falling-rising tone	dǎ	'hit'
High falling tone	dà	'big'

The other Chinese languages—as well as many languages of Southeast Asia, Africa, and the Americas—are also tone languages. A few tone languages are also found in Europe; for example, one of the dialects of Latvian makes a three-way tonal distinction (see Table 8.3).

As noted in the chapter on phonetics, there are two types of tones: level tones and contour tones. Tone languages most often contrast only two tone levels (usually

Table 8.3 Tone contrasts in Latvian

Falling tone	loks	|lûoks|	'arch, bow'
Level (high) tone	loks	|lûoks|	'green onion'
Rising-falling (broken) tone	loks	|lûoks|	'window'

high and low). However, contrasts involving three tone levels (such as high, low, and mid tones) are also relatively common. Five or more levels of tonal contrast are extremely rare.

Tone systems, too, exhibit various universal tendencies:

- If a language has contour tones (such as rising tone or falling tone), then it will also have level tones (such as high, mid, or low tone). Burmese, Crow, Latvian, and Mandarin are examples of languages that fit this pattern. The reverse pattern (languages with contour tones but no level tones) is extremely rare (although Dafla, spoken in northern India, has such a system).
- If a language has complex contour tones (such as rising-falling or falling-rising), then it will also have simple contour tones (like rising or falling). Both the Mandarin and Latvian examples fit this pattern.

Differences in stress are also useful in classifying languages. **Fixed stress languages** are those in which the position of stress on a word is predictable. For example, in Modern Hebrew and K'iché (a Mayan language), stress always falls on the last syllable of a word; in Polish, Swahili, and Samoan, stress falls on the penultimate (second-to-last) syllable of a word; while in Czech, Finnish, and Hungarian, the stressed syllable is always the first syllable of a word. In **free stress languages**, the position of stress is not predictable and must be learned for each word. Free stress is also called phonemic stress because of its role in distinguishing between words. Russian is an example of a language with free stress, as shown in Table 8.4.

Table 8.4 Stress contrasts in Russian

múka	'torture'	muká	'flour'
zámok	'castle'	zamók	'lock'
rúki	'hands'	rukí	'hand's' (genitive singular)

Syllable structure

All languages permit V and CV syllable structures (where V normally stands for a vowel, and C for a consonant). These syllable types are unmarked, in the sense that they are permitted in all languages. They are also simpler than most other syllable structures, such as CVC or VCC. Note, however, that VC is just as simple a structure as CV, but that only the latter is universally permitted. The presence of an onset (as in a CV syllable) is apparently more valued than the presence of a coda (as in a VC syllable), perhaps because an onset may help to signal the beginning of a new syllable.

In any given language, onsets may be structured differently from codas. For example, in English, a nasal + stop sequence is permitted in the coda (in a word like *hand*), but not in the onset (there are no English words that begin with the sequence *nd*). However, Swahili has precisely the opposite restrictions: the *nd* sequence is permitted in onset position (in words like *ndizi* 'banana'), but not in coda position. In fact, Swahili syllables are coda-less—they can only end in vowels.

Differing syllable structure constraints can have interesting consequences when languages come in contact. For example, in Hawaiian only V and CV syllables are permitted. Thus, when a word is borrowed from a language like English, which allows more complicated syllable structures, vowels are inserted to produce the only allowed syllable structures. For example, when the phrase *Merry Christmas* was borrowed into Hawaiian, it was reformulated as follows: *mele kalikimaka*. (Of course, some consonant changes were made as well, since Hawaiian lacks /r/ and /s/ phonemes.)

Two examples of implicational universals for syllable structure are presented below. Both deal with the structure of onsets as opposed to codas.

- If a language permits sequences of consonants in the onset, then it will also permit syllables with single consonant onsets and syllables with no onset at all.
- If a language permits sequences of consonants in the coda, then it will also permit syllables with single consonant codas and syllables with no coda at all.

2.2 MORPHOLOGY

Both words and morphemes are found in all languages. However, there are clear differences in the ways in which individual languages combine morphemes to form words. Four types of systems can be distinguished.

The isolating type

A language that is purely an **isolating** or **analytic language** would contain only words that consist of a single (root) morpheme. In such a language there would be no affixes, and categories such as number and tense would therefore have to be expressed by a separate word. In Mandarin Chinese, which is primarily an isolating language, the morpheme *le* is often used to indicate a past or completed action. Although this morpheme is thus semantically similar to a past tense, it acts just like an independent word, since its position in the sentence may vary:

3)

Tā chī fàn *le*
he eat meal past
'He ate the meal.'

Tā chī *le* fàn
he eat past meal
'He ate the meal.'

Other languages that are primarily isolating include Cantonese, Vietnamese, Laotian, and Cambodian.

The polysynthetic type

In a **polysynthetic language**, single words can consist of long strings of roots and affixes that often express meanings that are associated with entire sentences in other languages. The following word from Inuktitut illustrates this.

4)

Qasuiirsarvigssarsingitluinarnarpuq
Qasu -iir -sar -vig -ssar -si -ngit-luinar -nar -puq
tired not cause-to-be place-for suitable find not completely someone 3.sg
'Someone did not find a completely suitable resting place.'

Polysynthesis is common in many native languages of North America, including Inuktitut, Cree, and Sarcee to name but a few.

The terms *isolating* and *polysynthetic* refer to two extremes: words consisting only of single morphemes versus words that can be complete sentences. Few if any languages are either purely isolating or purely polysynthetic. Instead, the vast majority of languages are **synthetic languages**, in that they permit multimorphemic words.

Next we present two other morphological types that are sometimes distinguished.

The agglutinating type

An **agglutinating language** has words that can contain several morphemes, but the words are easily divided into their component parts (normally a root and affixes). In such languages, each affix is clearly identifiable and typically represents only a single grammatical category or meaning. The following examples are from Turkish (in standard Turkish orthography).

5)

a. köy
'village'

b. köy-ler
village-plural
'villages'

c. köy-ler-in
village-pl-genitive
'of the villages'

Turkish words can have a complex morphological structure, but each morpheme has a single, clearly identifiable function. In *5c*, for instance, *-ler* marks plurality and *-in* marks the genitive case, giving the meaning 'of the villages'.

The fusional type

Words in a **fusional** or **inflectional language** can also consist of several morphemes. However, in contrast to agglutinating systems, the affixes in fusional languages often mark several grammatical categories simultaneously. In Russian, for example, a single inflectional aftix simultaneously marks the noun's gender class (masculine, feminine, or neuter), its number (singular or plural), and its grammatical role (subject, direct object, and so on). This is illustrated in *6* for the suffix *-u*.

6)

mi vidlim ruk-u
we see hand-fem.sg/Accusative
'We see a/the hand.'

The distinction between agglutinating and fusional is sensitive to the number of semantic "bits" of information normally packed into an affix. In an agglutinating language, each affix normally contains only one element of grammatical or lexical meaning, while in a fusional language, affixes often denote several simultaneous functions.

Mixed types

Many (perhaps most) languages do not belong exclusively to any of the four categories just outlined. For example, English employs isolating patterns in many verbal constructions, where each notion is expressed by a separate word. The future, for instance, is indicated by the independent word *will* (rather than by an affix) in structures such as *I will leave*. On the other hand, English also exhibits considerable agglutination in derived words, such as *re-en-act-ment*, which consist of a series of clearly identifiable morphemes, each with its own unique meaning and function. However, the English pronoun system is largely fusional, since a single form can be used to indicate person, number, gender, and case. The word *him*, for instance, is used to express a third-person singular masculine direct object.

Since many, if not most, of the world's languages exhibit mixed patterns of this type, it has been suggested that terms like isolating, agglutinating, and fusional should be used to refer not to a language as a whole but to particular structures within a language.

It is also important to recognize that these classifications do not take into consideration morphological processes such as compounding (e.g., English *greenhouse*), reduplication (e.g., Tagalog *sulat* 'write' versus *susulat* 'will write'), grammatical use of stress or tone (e.g., the noun *présent* versus the verb *presénť* in English), and internal word change (e.g., vowel ablaut, as in English *run* versus *ran*).

Implicational universals: morphology

A variety of generalizations can be made about word structure in human language.

- If a language has inflectional affixes, it will also have derivational affixes. For example, English not only has inflectional affixes such as the past tense *-ed* and possessive *-'s*, but it also contains derivational affixes like *un-* (*unhappy*, *unwanted*) and *-ly* (*quickly*, *slowly*).
- If a word has both a derivational and an inflectional affix, the derivational affix is closer to the root (DA = derivational affix; IA = inflectional affix) (see Table 8.5).

Table 8.5 The ordering of derivational and inflectional affixes

English	
friend-ship-s	*friend-s -ship
Root DA IA	Root IA DA
Turkish	
iʃ -tʃi -ler	*iʃ -ler -tʃi
work -er -pl	work -pl -er
Root DA IA	Root IA DA

- If a language has only suffixes, it will also have only postpositions. (As noted in Chapter 5, postpositions are the equivalent of prepositions in languages that place the head at the end of the phrase.) Turkish, for example, has only suffixes; as expected, it also has postpositions rather than prepositions. This is illustrated in the following sentence.

7)

Ahmet Aiʃe itʃin kitab-i al-di
Ahmet Ayshe for book-Acc bought
'Ahmet bought a book for Ayshe.'

2.3 Syntax

Because we lack detailed descriptions for most of the world's languages, much of the work on syntactic universals has been restricted to the study of word order in simple declarative sentences such as *The men built the house*. Patterns are classified in terms of the order of the subject (S), direct object (O), and verb (V). The three most common word orders (in descending order of frequency) are SOV, SVO, and VSO. Over 95 percent of the world's languages use one of these patterns as their basic word order.

8)

SOV (Turkish, in standard orthography):
Hasan öküz-ü al-dı
Hasan ox-Acc bought
'Hasan bought the ox.'

9)

SVO (English):
The athlete broke the record.

10)

VSO (Welsh, in standard orthography):
Lladdodd y ddraig y dyn
killed the dragon the man
'The dragon killed the man.'

SOV, SVO, and VSO patterns all have one common trait: the subject appears before the direct object. The prevalence of the SO pattern may be due to the fact that the subject usually coincides with the topic of the sentence (i.e., what the sentence is about; see Chapter 6, Section 4.3), and therefore is more useful at an early point in the utterance.

While an overwhelming majority of the world's languages place the subject before the direct object in their basic word order, this pattern is not universal. There are a small number of VOS languages, of which the best-known example is Malagasy.

11)

VOS (Malagasy):
Nahita ny mpianatra ny vehivavy
saw the student the woman
'The woman saw the student.'

As well, there are a very few OVS or OSV languages, all of which seem to be spoken in South America:

12)

OVS (Hixkaryana):
Kana yanimno biryekomo
fish caught boy
'The boy caught a fish.'

13)

OSV (Apuriña)
Anana nota apa
pineapple I fetch
'I fetch a pineapple.'

Word-order universals

Sometimes the order of elements within one kind of structure has implications for the order of elements in other structures. Many of these implications concern the relationship between the verb and its (direct) object.

- If a language has VO word order, then it will have prepositions rather than postpositions. Languages of this type include Berber (spoken in Morocco), Hebrew, Maori (spoken in New Zealand), Maasai (spoken in Kenya), Welsh, and Irish Gaelic.

14)

Irish Gaelic

a. VSO pattern:
Chonaic mé mo mháthair
saw I my mother
'I saw my mother.'

b. Preposition pattern:
sa teach
in house
'in the house'

- If a language has OV word order, then it will probably have postpositions rather than prepositions. Languages with this structural pattern include Basque, Burmese, Hindi, Japanese, Korean, Quechua, Turkish, and Guugu Yimidhirr, an aboriginal language of Australia (Erg = Ergative).

15)

Guugu Yimidhirr

a. SOV pattern:
Gudaa-ngun yarrga dyinday
dog-Erg boy bit
'The dog bit the boy.'

b. Postposition pattern:
yuwaal nganh
beach from
'from the beach'

- PPs almost always precede the verb in OV languages, and usually follow the verb in VO languages (Nom = Nominative; Acc = Accusative).

16)

Japanese

a. SOV pattern:
Gakusei-ga hon-o yonda
student-Nom book-Acc read
'The student read a book.'

b. PP precedes verb:
Taroo-ga [$_{PP}$ nitiyoobi ni] tsuita.
Taroo-Nom Sunday on arrived
'Taroo arrived on Sunday.'

17)

English

a. SVO pattern:
I like candy.

b. PP follows verb:
George left [$_{PP}$ on Sunday].

- Manner adverbs overwhelmingly precede the verb in OV languages and generally follow the verb in VO languages.

18)

Japanese (SOV pattern, as seen in *16a*):
Manner adverb precedes verb:
hayaku hasiru
quickly run
'run quickly'

19)

English (SVO pattern, as seen in *17a*):
Manner adverb follows verb:
John runs well.

- With respect to possessive structures, there is an overwhelming preference for Genitive + N order in OV languages, and a (somewhat weaker) preference for N + Genitive order in VO languages.

20)

Japanese (SOV pattern, as seen in *16a*):
Genitive structure precedes head N:
Taroo-no hon
Taroo-Gen book
'Taroo's book'

21)

French

a. SVO pattern:
Pierre aime Marie.
'Pierre likes Marie.'

b. Genitive structure follows head N:
la maison de Marie
the house of (Gen) Marie
'Marie's house'

English, although an SVO language, exhibits both Genitive + N and N + Genitive patterns:

22)

a. Genitive + N pattern:
the country's laws

b. N + Genitive pattern:
the laws of the country

Examples such as this are rare, however, and do not invalidate the universal tendencies we have been considering.

Grammatical hierarchies

Implicational universals are often stated in terms of **hierarchies** of categories or relations. One of the most important hierarchies of this type refers to the grammatical relations of subject and direct object (see Chapter 5). Hierarchies represent degrees of markedness, with the least marked option at the top and the most marked at the bottom. According to the hierarchy in Figure 8.4, then, a process that applies only to subjects is less marked than a process that applies to direct objects, and so on. Given the definition of markedness outlined at the beginning of Section 2, it follows that if a particular phenomenon applies to direct objects, it should also apply

Least marked

Subject
Direct object
Other

Most marked

Figure 8.4 Hierarchy of grammatical relations

to subjects. The opposite, however, need not be true: it would not be surprising to find a process that applies to subjects but not to direct objects.

Among the many typological phenomena that conform to this hierarchy is verb agreement, first mentioned in Chapter 4 (Section 6.4). As the following examples show, there are languages in which the verb agrees only with the subject, and there are languages in which it agrees with both the subject and the direct object (3 = 3rd person, SG = singular, PL = plural, Pst = past).

23)

Agreement with subject only (Spanish):

Subject
Juan parti-ó
Juan leave-3.SG.Pst
'Juan left.'

24)

Agreement with subject and direct object (Swahili):

Subject *Direct object*
Juma a- li- wa- piga watoto
Juma 3.SG Pst- 3.PL-hit children
'Juma hit the children.'

However, as predicted by the hierarchy, there are no languages in which the verb agrees only with the direct object.

2.4 Explaining Universals

Linguists are still uncertain about how to explain the existence of many linguistic universals. Nonetheless, a number of interesting proposals have been made, and it is worthwhile to consider some of them here.

Phonology

Perceptual factors play a role in shaping phonological universals. For example, the fact that /s/ is the most commonly occurring fricative may have to do with its acoustic prominence: varieties of /s/ are inherently louder than other kinds of fricatives.

Vowel systems (discussed in Section 2.1) develop so as to keep vowel phonemes as different from each other as possible. A three-vowel system such as the following allows for plenty of "space" around each vowel, which probably makes each vowel easier to distinguish from the others.

i u

a

Figure 8.5 A three-vowel system

The same holds true for the distribution of stop phonemes. It may be that /p/, /t/, and /k/ are the three most common stops, because they occur at three maximally distant places of articulation within the supralaryngeal vocal tract. These three stops are much easier to distinguish perceptually than a sequence of dental, alveolar, and palatal stops, for example, all of which are produced in the central region of the oral cavity (i.e., in the center of the mouth).

It has been recently suggested that consonant systems in general respond to the articulatory pressures that give rise to unmarked sounds and systems. Basic obstruents such as [p], [t], and [k] are found much more commonly than more complex articulations such as [tɬ] and [q^{w}]. Table 8.6 shows the set of obstruents that is most widely used across human languages.

Languages tend to have consonant systems that consist of about 70 percent obstruents and 30 percent sonorants no matter what the total size of their consonant inventories may be. These figures reflect the articulatory possibilities available for contrast: more distinctions can be made among obstruents than among sonorants. There are, for example, no nasal fricative sonorants, because the air pressure needed to force air through a narrow opening (which is necessary for the production of fricatives) cannot be built up when so much air is flowing through the nasal passage at the same time. For reasons such as this, the number of obstruent consonants in any language is potentially much larger than the number of possible sonorant consonants. This is just one example of how considerations involving articulation can play a role in the shaping of consonant systems.

Table 8.6 Obstruents most often found cross-linguistically

p	t	k	ʔ
b	d	g	
f	s		h
	tʃ		

Morphology

Other types of explanations are appropriate for morphological universals. For example, the fact that languages with suffixes but no prefixes always have postpositions (Section 2.2) may have a historical explanation. In these languages, some postpositions became attached to a preceding word and were thereby converted into suffixes. Because suffixes in such languages evolved from postpositions, the link between the two elements can be traced to their common origin.

An example of this very phenomenon can be seen in the closely related languages Finnish and Estonian. The ancestor language (Proto-Baltic-Finnic) contained a postposition **kanssa* 'with', which is still evident in Standard Finnish but has evolved into a suffix in Estonian (see Table 8.7).

Table 8.7 Proto-Baltic-Finnic postposition *kanssa* becomes suffix *-ga*

Standard Finnish: postposition *kanssa* 'with'				
poika	'boy'	poja-n	kanssa	'with the boy'
		boy+Gen	with	
Estonian: case suffix *-ga*				
poeg	'son'	poja-ga		'with the son'
		son+Committative		

The requirement that derivational affixes occur closer to the root than inflectional affixes has another type of explanation. As noted in the morphology chapter, derivation typically forms new words, while inflection marks the subclass (for example, plural for Ns, past tense for Vs) to which a word belongs. Given that a word must be formed before its subclass can be determined, it follows that derivational processes will precede inflection. This is reflected in word structure, where derivational affixes appear closer to the root than inflectional markers. In Figure 8.6, for instance, the verbal root *treat* is converted into a noun by the affix *-ment* before the plural inflectional marker is added.

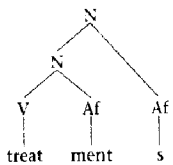

Figure 8.6 The structure of a word containing a derivational affix and an inflectional affix

Syntax

At least some syntactic universals may be explained in terms of the way that the human brain processes sentence structure. Consider the summary of word order patterns in Table 8.8, which is based on the implicational universals discussed in Section 2.3.

Table 8.8 Word-order patterns

Constituents	*Order in VO language*	*Order in OV language*
P & NP	preposition-NP	NP-postposition
V & PP	verb-PP	PP-verb
V & manner Adv	verb-manner Adv	manner Adv-verb
N & Gen	noun-genitive	genitive-noun

One recent explanation as to why the word-order properties in the second and third columns cluster together involves the contrast between right-branching and left-branching languages. In right-branching languages, the more elaborate part of a phrase's structure occurs on its right branch; in left-branching languages, it occurs on the left. Thus, a verb-object pattern is right-branching since a phrasal constituent (an XP) appears on its right branch, but an object-verb pattern is left-branching, as shown in Figure 8.7.

a. Right-branching (VO) *b. Left-branching (OV)*

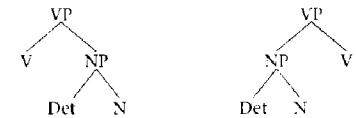

Figure 8.7 Right-branching and left-branching patterns

As you can easily determine for yourselves, the P-NP, V-PP, V-Adv, and N-Gen patterns commonly associated with VO languages are also all right-branching (both genitives and adverbials are a type of phrase). In contrast, the NP-P, PP-V, Adv-V, and Gen-N patterns typically found in OV languages are all left-branching. In other words, it seems that languages are fairly consistent in using one or the other type of branching structure. This sort of uniformity may make it easier for speakers and hearers to process syntactic structure. Thus, just as some human beings are right-handed and others left-handed, it appears that some languages prefer to use consistently right-branching systems, while others prefer consistently left-branching systems.

The study of linguistic typology and language universals is a relatively new field within linguistics. There is obviously much still to be learned about linguistic universals, and it must be admitted that some of the current work is speculative and incomplete. No doubt many interesting new facts will eventually come to light.

3 GENETIC CLASSIFICATION

The world's languages can be grouped into a relatively small number of language families. However, genetic classification is sometimes difficult for a number of reasons.

Perhaps the biggest problem is simply the amount of data that must be collected before linguists can be confident about the status of a group of languages. It is only in the last two or three decades, for example, that enough information has been accumulated to propose a detailed classification of the languages of Africa. Moreover, many of the languages of South America, New Guinea, and Australia are still relatively unknown.

In many cases, linguists face the problem of establishing the tests or criteria to be used in proposing genetic relationships. There is some disagreement over the degree of similarity that should exist among languages before a genetic relationship can be proposed. This issue arises because unrelated languages are often typologically similar (that is, share some structural characteristics). This is particularly likely if languages have been in contact long enough to have borrowed a large number of words, sounds, morphemes, or syntactic structures from one another.

Additional difficulties stem from the fact that genetically related languages need not be typologically similar. This is especially true if the relationship is a distant one, as is the case with English and Russian. Russian has numerous inflectional affixes, an extensive case system, and free word order, while English has relatively few inflectional affixes, virtually no case marking, and fixed word order. Yet, both belong to the **Indo-European family**.

To complicate matters even further, linguists also disagree as to how much evidence is required in order to establish a genetic relationship between languages. The more distant the genetic relationship between languages, the less likely it is that a large number of obvious cognates will be found, especially since sound changes can obscure similarities between cognate words. English and Latin are related (though distantly), but the similarity between cognates like Latin *unda* 'wave' and English *water* is certainly not striking.

Research is also hampered by the fact that words that may be excellent indicators of a genetic relationship can drop out of the lexicon. For example, Old English had a word *leax* 'salmon' (which was cognate with German *Lachs* and Yiddish *lox*), but this lexical item has since been lost from the native English lexicon (although *lox* has, of course, been borrowed back into some varieties of English as the name for a popular delicatessen food).

Since word loss is a common historical event, linguists prefer to use the oldest available form of a language for their research; thus, our knowledge of Proto-Indo-European is drawn from the study of Old English, Sanskrit, Latin, etc., rather than English, Hindi-Urdu, French, and their other modern descendants. Of course, languages that are genetically related do share many similarities, particularly if their common ancestor is not too distant.

Some language families contain many hundreds of languages. In other cases, only one language may remain to represent a family. In still other cases, families have become extinct. The following sections present some information about the makeup and membership of a few of the language families represented in the world today.

3.1 The Indo-European family

With only about a hundred languages, Indo-European is not a large family in terms of the total number of languages. However, it is the largest language family in the world in terms of the total number of speakers: there are about 1.7 billion native speakers of an Indo-European language.

If we consider only living languages, the Indo-European family currently has nine branches, which are listed in Table 8.9.

Table 8.9 Main branches of the Indo-European family

Germanic	Hellenic	Baltic
Celtic	Albanian	Slavic
Italic	Armenian	Indo-Iranian

Germanic

The Germanic branch of Indo-European can be divided into three sub-branches: East, North, and West. The East Germanic branch included Gothic, the oldest Germanic language for which written texts exist (dating from the fourth century A.D.). Gothic and any other languages belonging to this branch of Germanic have long been extinct.

The North Germanic (or Scandinavian) branch originally included Old Norse (also known as Old Icelandic)—the language of the Vikings and the ancestor of modern Icelandic, Norwegian, and Faroese (spoken on the Faroe Islands, north of Scotland). Swedish and Danish are two other familiar North Germanic languages.

The West Germanic branch includes English, German, Yiddish, Dutch, Frisian, and Afrikaans. Afrikaans is descended from the Dutch spoken by seventeenth-century settlers (known as Boers) in South Africa.

Frisian is the language most closely related to English. It is spoken on the north coast of Holland, on the Frisian Islands just off the coast, as well as on the northwestern coast of Germany. English descended from the speech of the Angles, Saxons, and Jutes—Germanic tribes who lived in northern Germany and southern Denmark (in an area just east of the Frisians) before invading England in A.D. 449 and settling there.

The organization of the Germanic family of languages is illustrated in Table 8.10. (In this and other tables, parentheses are used to indicate languages that no longer have any native speakers. The tables are intended to illustrate the membership and organization of the families; they do not necessarily provide a complete list of the languages in each family.)

Table 8.10 The Germanic family

(East Germanic)	*North Germanic*	*West Germanic*
(Gothic)	Icelandic	English
	Faroese	German
	Norwegian	Yiddish
	Swedish	Dutch
	Danish	Frisian
		Afrikaans

Celtic

The Celtic branch of Indo-European (see Table 8.11) has two main sub-branches: Insular and Continental (now extinct). Gaulish, a member of the Continental branch, was once spoken in France (the Gauls were the tribe Julius Caesar defeated), but it has long been extinct.

The Insular sub-branch can be subdivided into two groups of languages: Brythonic and Goidelic. Brythonic languages include Welsh and Breton (which is spoken in northwestern France) as well as Cornish, which was formerly spoken in southwest Britain but no longer has any native speakers. The Goidelic branch contains Irish (or Irish Gaelic), which is still spoken natively in the western parts of Ireland, and Scots Gaelic, which is native to some of the northwestern parts of Scotland (especially the Hebrides Islands) and, to a lesser extent, Cape Breton Island in Nova Scotia.

Table 8.11 The Celtic family

Insular		*Continental*
Brythonic	*Goidelic*	
Welsh	Irish [= Irish Gaelic]	(Gaulish)
Breton	Scots Gaelic	
(Cornish)		

Italic

The Italic family originally had a number of branches, which included several now-extinct languages spoken in the area corresponding roughly to modern-day Italy. However, the only Italic languages that are presently spoken are all descended from Latin, the language of the Roman Empire (hence the term "Romance languages").

These languages can be divided into an Eastern group, consisting of Italian and Rumanian, and a Western group, containing all of the other Romance languages except Sardinian, which stands alone.

The Western group is further divided into Ibero-Romance (Spanish, Portuguese, and Catalan—the latter is spoken in northeastern Spain, around Barcelona) and Gallo-Romance, which includes French, Occitan (spoken in southern France), and Romansch (one of the four official languages of Switzerland). These divisions are illustrated in Table 8.12.

Table 8.12 The Romance family

Eastern	*Western*		
	Ibero-Romance	*Gallo-Romance*	*Sardinian*
Italian	Spanish	French	Sardinian
Rumanian	Portuguese	Occitan	
	Catalan	Romansch	

Hellenic

The Hellenic branch of Indo-European has only one living member, Greek. All modern Greek dialects are descended from the classical dialect known as Attic Greek, which was the speech of Athens during the Golden Age of Greek culture (approximately 500 to 300 B.C.).

Hellenic Greek, which was used in subsequent centuries, was the language of commerce throughout the Middle East. (Hellenic Greek was also Cleopatra's native language; she was descended from one of Alexander the Great's generals.)

Albanian

The Albanian branch of Indo-European has only one member—Albanian—which is spoken not only in Albania but also in parts of the former Yugoslavia, Greece, and Italy.

Armenian

The Armenian branch also has only one member—Armenian. This language is centered in the Republic of Armenia (once part of the former Soviet Union and located between the Black Sea and the Caspian Sea) but is also spoken in Turkey, Iran, Syria, Lebanon, and Egypt.

Baltic

The Baltic branch contains only two surviving languages—Latvian (or Lettish) and Lithuanian. They are spoken in Latvia and Lithuania (located just west of Russia and northeast of Poland). Lithuanian has an elaborate case system, which resembles the one proposed for Proto-Indo-European.

Slavic

The Slavic branch of Indo-European can be divided into three sub-branches: East, West, and South. The East Slavic branch is made up of Russian (also called Great Russian), Ukrainian, and Byelorussian (or White Russian). The latter is spoken in Byelorussia, which is just east of northern Poland. The West Slavic branch includes Czech, Slovak, and Polish.

The South Slavic branch consists of Bulgarian, Macedonian, Serbo-Croatian, and Slovene (or Slovenian). The latter three languages are all spoken in the former Yugoslavia. Note that although Alexander the Great was king of Macedonia, he spoke Hellenic Greek, not (Slavic) Macedonian; Slavic-speaking tribes did not move into that area until several centuries later.

The organization of the Slavic group of languages is represented in Table 8.13.

Table 8.13 The Slavic family

East Slavic	*West Slavic*	*South Slavic*
Russian	Czech	Bulgarian
Ukrainian	Slovak	Macedonian
Byelorussian	Polish	Serbo-Croatian
		Slovene

Indo-Iranian

The Indo-Iranian branch of Indo-European is divided into the Iranian and Indic sub-branches. The Iranian sub-branch contains about two dozen different languages, including Modern Persian (also called Parsi or Farsi, spoken in Iran), Pashto (the principal language of Afghanistan), and Kurdish (found in Iran, Iraq, Turkey, and Syria). Other Iranian languages are spoken in Pakistan, southern parts of the former Soviet Union, and China.

There are about thirty-five different Indic languages. Most of the languages spoken in northern India, Pakistan, and Bangladesh belong to this branch of Indo-European. Some of the most widespread (in terms of number of speakers) are Hindi-Urdu, Bengali, Marathi, and Gujarati. Although Hindi and Urdu are two dialects of the same language, they have totally different writing systems and are associated with different cultures; Urdu is spoken principally in Pakistan by Muslims while Hindi is spoken primarily in India by Hindus.

Less well known as an Indic language is Romany, or Gypsy. It is now believed that the Gypsies (or Roma) fled to Turkey from northwestern India during the Middle

Ages, after being defeated by Islamic invaders. Subsequently they spread throughout Europe: Gypsies are found as far west as Ireland and as far east as Russia. Many now live in North America. Romany contains many borrowed words—particularly from Greek, which was the language primarily spoken in Turkey during the Middle Ages.

Table 8.14 depicts the organization of Indo-Iranian.

Table 8.14 The Indo-Iranian family

Iranian	*Indic*
Persian [= Farsi]	Hindi-Urdu
Pashto	Bengali
Kurdish	Marathi
	Gujarati
	Romany [= Gypsy]

The map in Figure 8.8 illustrates the geographic location of the Indo-European families identified in this chapter.

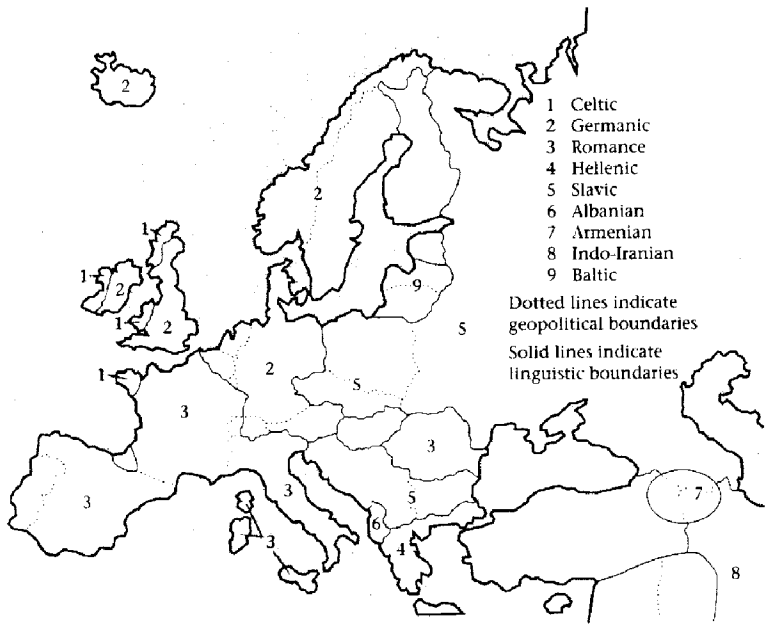

Figure 8.8 Location of Indo-European languages

3.2 Some other families

Although no introductory text could hope to present a complete survey of all of the world's language families, some further discussion of this topic is worthwhile in order to illustrate the extraordinary variety of human language.

Uralic

The Uralic family (see Table 8.15) contains about twenty languages and has approximately twenty-two million speakers. Uralic languages are spoken in a band across the northern part of Europe, all the way from northern Norway to Siberia. Uralic has two major branches: Samoyed and Finno-Ugric. The Samoyed branch contains a handful of languages spoken in northern Russia, particularly in areas around the Ural mountains, and also in Siberia.

The most widely spoken Finno-Ugric language is Hungarian. Other Finno-Ugric languages are Finnish, Lapp (also known as Lappish or Saame, spoken in northern Scandinavia and northwestern Russia), Estonian (Estonia), Livonian (Latvia), Karelian (eastern Finland and northwestern Russia), Erzya, and Cheremis (both spoken in the former Soviet Union).

Table 8.15 The Uralic family

Finno-Ugric		*Samoyed*
Finnic	*Ugric*	
Finnish	Hungarian	Nganasan
Lapp [= Saame]		Selkup
Estonian		Nenets
Livonian		Enets
Karelian		
Erzya		
Cheremis [= Mari]		

Uralic languages are primarily agglutinating, and most have postpositions with SOV or SVO word order. The nouns often have many cases (Finnish has fifteen), which appear to have developed historically from postpositions that became attached to nouns as suffixes.

Altaic

Languages belonging to the Altaic family are spoken in a continuum from Turkey to Siberia, as well as in China. The membership of the Altaic family (see Table 8.16) includes three main branches—Turkic, Mongolian, and Tungusic. Recent scholarship has collected evidence that Korean and Japanese are also members of the Altaic family.

The best-known Turkic language is Turkish, which is spoken by over fifty million people. Other Turkic languages (most of which are spoken in central Asia in the former Soviet Union) include Uzbek (Uzbekistan), Azeri (Azerbaijan Republic and Iran), Kazakh (Kazakhstan, China, and Mongolia), Uighur (China and Kazakhstan), and Volga Tatar (also called Tatar, in the Tataristan Republic, Kazakhstan, and China).

Table 8.16 The Altaic family

Turkic	*Mongolian*	*Tungusic*	*Korean*	*Japanese*
Turkish	Khalkha	Evenki	Korean	Japanese
Uzbek	Buriat	Manchu		Ryukyuan
Azeri [= Azerbaijani]	Chakhar	Even		
Kazakh	Kalmyck	Nanai		
Uighur	Monguor	Orik		
Volga Tatar [= Tatar]				

The Mongolian languages are spoken by around ten million people, primarily in Mongolia and China, while the Tungusic languages are spoken by approximately 50,000 people in central and east Siberia and Mongolia.

Altaic languages are usually agglutinating, often with several suffixes in the same word. They normally employ SOV word order and typically use postpositions rather than prepositions. Many Altaic languages have vowel harmony—a phonological phenomenon in which all vowels of a word share certain features, such as [round] or [back].

Caucasian

The languages that are normally grouped together as Caucasian have not yet been assigned to families in a definitive way. These languages are primarily found in northeastern Turkey and in the former Soviet Union (between the Black Sea and the Caspian Sea, in and around the Caucasus Mountains). The best evidence so far points to three distinct language families—South Caucasian, Northwest Caucasian, and Northeast Caucasian (see Table 8.17).

Table 8.17 The Caucasian languages

The South Caucasian (Kartvelian) family
Georgian
Laz [= Mingrelian]
Svan
The Northwest Caucasian family
Adyghe [= West Circassian]
Kabardian [= East Circassian]
The Northeast Caucasian family
Chechen
Lezghian
Avar [= Daghestani]

South Caucasian (sometimes called Kartvelian) consists of Georgian, Laz (also called Mingrelian), and Svan. Georgian has the largest number of speakers. It was also the native language of Joseph Stalin, dictator of the former Soviet Union in the 1930s and 1940s. Northwest Caucasian contains a handful of languages, including Adyghe and Kabardian (also called West Circassian and East Circassian, respectively). Northeast Caucasian consists of about two dozen languages. Of these, Chechen, Lezghian, and Avar have the largest number of speakers.

Altogether there are about thirty-five languages in the three separate families, with a total of approximately five million speakers. Although no genetic relationship has been proven to exist between these three families, they do seem to share a number of areal features (probably brought about through mutual borrowing): many Caucasian languages have glottalized consonants, complex consonant clusters, a very large consonantal inventory, but very few vowel phonemes. It has recently been claimed that the Northwest and Northeast Caucasian languages are part of a single family, but this grouping is not yet widely accepted among Caucasianists.

Dravidian

There are twenty-five Dravidian languages (see Table 8.18), which are primarily found in the southern half of India, but also in Sri Lanka, Pakistan, and Nepal. About 175 million people are native speakers of a Dravidian language. The most widely spoken languages in this family are Telugu, Tamil, Kannada, and Malayalam. Dravidian languages are normally SOV. They are agglutinating and nontonal, and usually have initial stress.

Table 8.18 The Dravidian family

North	*Central*	*South-Central*	*South*
Kurux	Kolami	Telugu	Tamil
Malto	Naiki	Savara	Kannada
Brahui	Parji	Konda	Malayalam
	Gadaba	Gondi	Tulu

Austroasiatic

The Austroasiatic family of languages (see Table 8.19) consists of about 150 languages with approximately fifty million speakers.

Mon-Khmer is the largest branch of Austroasiatic and contains languages such as Vietnamese, Cambodian (also called Khmer), and many other languages of India, Cambodia, Vietnam, Burma, and southern China. The Munda branch of Austroasiatic includes languages spoken in central and northeastern India, such as Mundari and Santali. Other Austroasiatic languages are spoken in Malaysia and on the Nicobar Islands (northwest of Sumatra).

Table 8.19 The Austroasiatic family

Mon-Khmer	*Munda*
Vietnamese	Mundari
Cambodian [= Khmer]	Santali
Mon	
Parauk	

Some Austroasiatic languages are tonal (for example, Vietnamese) and some are characterized by large and complex vowel systems. Word order is generally SVO or SOV.

Tai-Kadai

The Tai-Kadai family includes Thai (formerly called Siamese), Laotian, Shan (spoken in Burma and Thailand), and several other languages of China, Thailand, and Vietnam. Typical salient features of these languages include SVO order, a general lack of inflectional morphemes, and the widespread use of tone (with the number of contrasting tones varying from three to nine).

Sino-Tibetan

In terms of numbers of speakers, the Sino-Tibetan family (see Table 8.20) is the largest language family after Indo-European. There are about three hundred Sino-Tibetan languages, with well over a billion native speakers.

Table 8.20 The Sino-Tibetan family

Tibeto-Burman	*Sinitic*						
	Mandarin	*Wu*	*Min*	*Yue*	*Xiang*	*Hakka*	*Gan*
Tibetan	Mandarin	Wu	Taiwanese	Cantonese	Hunan	Hakka	Gan
Burmese			Amoy				
Yi [= Nyi]			Hokkian				
Sharpa			Fukian				

There are two major branches: Tibeto-Burman and Sinitic. The first branch includes Tibetan, Burmese, and many other languages spoken in northeastern India, Nepal, Burma, Tibet, and China. For the most part, these languages employ SOV word order.

The Sinitic branch contains several different subgroupings, including Mandarin (with major dialects in and around Beijing, Szechuan, and Nanking), Wu (with dialects in Shanghai and Suchow), Min (which includes Taiwanese, Amoy, Hokkian, and Fukian), Yue (Cantonese), Xiang, Hakka, and Gan. The Sinitic languages typically have SVO order and are usually tonal. They are predominantly isolating, having many monomorphemic (and usually monosyllabic) words. Consonant clusters are normally avoided.

The Sinitic family of languages is sometimes referred to as "Chinese" by nonlinguists, as if it were a single language with several dialects rather than a group of related, mutually unintelligible languages. This confusion is based on the fact that the same writing system is used across China and can be understood by speakers of different Chinese languages (see Chapter 15, Section 4.1).

Austronesian

The Austronesian family (see Table 8.21) contains approximately one thousand languages, which are spoken from the island of Madagascar halfway across the world to Southeast Asia, Hawaii, Easter Island, and New Zealand. Some of the languages of Taiwan belong to the Formosan branch of this family; however, Taiwanese, which is spoken by most of the island's residents, is a Sinitic language (see above).

Table 8.21 The Austronesian family

Formosan	*Malayo-Polynesian*	
	Western	*Oceanic*
Paiwan	Malagasy	Samoan
Amis	Malay [= Indonesian]	Tahitian
Atayal	Tagalog [= Pilipino]	Maori
Seediq	Javanese	Hawaiian
	Sundanese	Fijian
	Balinese	Motu
		Ponapean

The largest branch within the Austronesian family is Malayo-Polynesian, which contains all the Austronesian languages outside of Taiwan. These include Malagasy (spoken on Madagascar), Malay (and the mutually comprehensible Indonesian), Tagalog (the basis for Pilipino, the official language of the Philippines), Javanese, and many other languages spoken in the Philippines, Malaysia, Indonesia, Vietnam, Cambodia, Taiwan, and the islands of the Pacific Ocean.

The Polynesians were intrepid ocean travelers who colonized Hawaii, Easter Island, and New Zealand sometime between A.D. 500 and 1000. Well-known Polynesian languages include Samoan, Tahitian, Maori, and Hawaiian (which now has only a few hundred first-language native speakers, although efforts are underway to revive it).

A characteristic feature of Austronesian languages is the extensive use of reduplication. Many of these languages also make liberal use of infixes, which are extremely rare in other language families. Word order is usually SVO, although VSO is more prevalent in the Austronesian languages spoken in Taiwan, the Philippines, Northern Borneo, and Polynesia.

Some research has attempted to link the Austronesian family with the Austroasiatic family of India and Southeast Asia, forming a larger Austric family. However, this relationship is still very tentative.

Indo-Pacific

Indo-Pacific (or Papuan) languages are all spoken on the island of New Guinea, on nearby islands such as New Britain or Bougainville, or on the Andaman Islands (just southwest of Burma). Little is known about many of these languages, but they appear to be about seven hundred in number, with just under three million speakers.

Two languages with relatively large speaker populations are Enga (165,000 speakers, spoken in the western highlands of New Guinea) and Bunak (50,000 speakers, spoken on the island of Timor, west of New Guinea).

Indo-Pacific languages are normally tone languages. Nouns are often marked for case but not always for number. Word order is usually SOV.

Australian

Recent studies have established that all of the aboriginal languages of Australia belong to the same family. There are about 170 such languages, but many have very

few speakers. There are currently only about thirty thousand speakers of aboriginal Australian languages.

The majority of Australian languages are spoken in Arnhem Land (north central Australia) and the northern part of Western Australia. The languages with the largest number of speakers are Mabuiag (seven thousand speakers on the Torres Straits Islands, north of Australia) and the Western Desert Language (five thousand speakers in Western Australia).

Australian languages are characterized by simple vowel systems. Nouns are normally marked for case, sometimes in unusual and intricate ways, and word order can be very free.

Afroasiatic

Afroasiatic languages (see Table 8.22) are spoken primarily in a band across the northern half of Africa and in the Middle East. There are about 250 Afroasiatic languages and 175 million speakers of these languages.

Table 8.22 The Afroasiatic family

(Egyptian)	*Cushitic*	*Berber*	*Chadic*	*Semitic*
(Coptic)	Somali	Tachelhit	Hausa	(Babylonian)
	Oromo	Tamazight		(Assyrian)
		Kabyle		(Old Canaanite)
		Rift		(Moabite)
		Tuareg		Aramaic
				Arabic
				Amharic
				Modern Hebrew

Afroasiatic has five main branches, one of which—Egyptian—no longer contains any living languages. Although Old Egyptian was spoken from 3000 B.C. onward (including during the time of Rameses II [1290–1224 B.C.], who was probably Pharaoh at the time of the Exodus), it has long been extinct. Its descendant, Coptic, is now used only as the liturgical language of the Coptic Church.

A second branch of Afroasiatic is Cushitic, whose member languages are spoken in Somalia, Kenya, Ethiopia, and the Sudan. A third branch, Berber, includes several languages of Algeria, Morocco, and Niger, such as Tamazight and Tuareg. Still another branch, Chadic, contains many of the languages of Chad and Nigeria, such as Hausa. Unlike other Afroasiatic languages, Chadic languages are tonal.

The fifth and largest branch of Afroasiatic (in terms of number of speakers) is the Semitic branch. Many (now extinct) languages mentioned in the Bible were of Semitic origin, such as Babylonian, Assyrian, (Old) Canaanite, Moabite, Classical Hebrew, and Biblical Aramaic. Biblical (or Palestinian) Aramaic was spoken in Palestine at the time of Jesus, and may have been his native language.

Classical Hebrew died out several centuries before the birth of Jesus, although it was maintained as a written language within Judaism. Modern (or Israeli) Hebrew is not directly descended from Classical Hebrew; rather, it was created (or re-created) at the beginning of this century by regularizing some aspects of Classical Hebrew and adding new vocabulary. Modern Hebrew has only had a community of native speakers for the past few decades.

Still another Semitic language, Arabic, has various dialects (not all of which are mutually intelligible) spoken across North Africa and throughout the Middle East. All of these are descended from Classical Arabic, which was the language of Mohammed, the founder of Islam, and is the language of the Koran, the holy book of Islam.

The Semitic languages are characterized by a system of consonantal roots. Most roots consist of three (sometimes two) consonants, with vowels being inserted to indicate various inflectional and derivational categories (see Chapter 4, Section 1.3). For example, Arabic has the root *k-t-b* (denoting the concept of writing) from which a variety of words can be formed, including *kitaabun* 'book', *kaatibun* 'writer', *kataba* 'he wrote', and *yaktubu* 'he is writing'. The Semitic languages frequently have complex consonant clusters and pharyngeal or pharyngealized consonants.

Niger-Congo

Most of the languages spoken in sub-Saharan Africa belong to the Niger-Congo family of languages (see Table 8.23). In all, this family contains over nine hundred languages, with a total of (approximately) 180 million speakers. There are three major branches: Kordofanian, Mande, and Atlantic-Congo.

Table 8.23 The Niger-Congo family

Kordofanian	*Mande*	*Atlantic-Congo*	
Tegali	Maninka	Swahili	Xhosa
Koalib	Bambara	Shona	Yoruba
Katla	Mende	KinyaRwanda	Igbo
		Zulu	Wolof

The Kordofanian branch includes only a handful of languages spoken in the Sudan, such as Tegali, Koalib, and Katla. The Mande branch contains a number of families of languages spoken in West Africa, such as Maninka, Bambara, and Mende. (Alex Hailey's famous African ancestor, described in the novel *Roots*, was probably a speaker of Maninka.)

Atlantic-Congo, on the other hand, is much larger and can be divided into several branches and numerous sub-branches. One of the largest sub-branches contains the more than one hundred languages of the Bantu family, with about fifty-five million speakers. Some of the principal Bantu languages are Swahili (Tanzania and Kenya), Shona (Zimbabwe and Zambia), KinyaRwanda (Rwanda, Uganda, and Congo), Zulu (South Africa and Lesotho), and Xhosa (South Africa). Other Atlantic-Congo languages include Yoruba (Nigeria, Togo, and Benin), Igbo (Nigeria), and Wolof (Senegal).

Niger-Congo languages are typically SVO and usually have tone systems (with the notable exception of Swahili). The Bantu languages are usually agglutinating with verb-subject and verb-direct object agreement. Languages in the Bantu group also exhibit a complex system of noun classes, each of which is marked by a separate set of prefixes.

Nilo-Saharan

The Nilo-Saharan family is primarily found in eastern and central Africa and includes approximately 120 languages, with about thirty million speakers. Languages

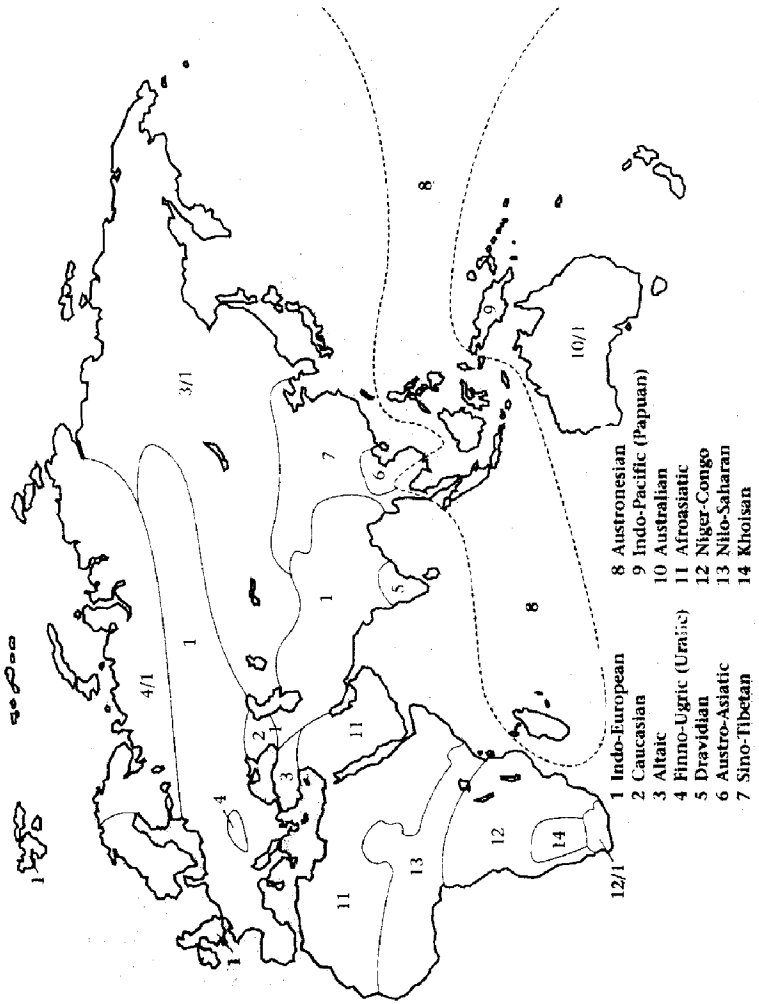

Figure 8.9 Location of some major language families

in this family include Luo and Maasai (both spoken in Kenya), Dinka (Sudan), Kanuri (Nigeria), and Nuer (Sudan and Ethiopia).

Nilo-Saharan languages generally have tonal systems, and nouns are usually marked for case and often use internal change (as in English *foot* vs. *feet*).

Khoisan

The Khoisan family is quite small, containing only about thirty languages spoken by 120,000 speakers. The majority of Khoisan languages are spoken in the southern and southwestern areas of Africa.

Some Khoisan languages are Hottentot (= Nama), !Kung, and Sandawe (one of only two Khoisan languages spoken in east Africa). Khoisan languages have unusual click sounds in their consonantal systems. These clicks have been borrowed by a few neighboring Bantu languages, such as Zulu and Xhosa.

3.3 North, Central, and South America

Contrary to popular belief, not all native American Indian (usually called **Amerindian**) languages belong to the same family. Although many of the genetic relationships are still unclear, it appears that there are well over a dozen different language families in the Americas (see Table 8.24). (Languages in parentheses in Table 8.24 are extinct.)

Table 8.24 North, Central, and South American families

Language family	*Some member languages*
Eskimo-Aleut	Inuktitut [= Inuit]
Athabaskan	Navajo, Apache, Chipewyan, Dogrib, Slavey
Algonquian	Blackfoot, Micmac, Cree, Ojibwa, (Mohican)
Siouan	Dakota, Lakota, Winnebago, Crow
Iroquoian	Cherokee, Mohawk, Cayuga, (Huron)
Caddoan	Caddo, Witchita, Pawnee
Wakashan	Nootka, Kwakiutl, Nitinat
Salish	Flathead, Halkomelem, Okanagan, Shuswap
Klamath-Sahaptin	Yakima, Nez Perce, Sahaptin, Klamath
Penutian	Patwin, Wintu, Nomlaki
Muskogean	Choctaw, Muskogee
Hokan	Diegueno, Yuma, Mohave
Coahuiltecan	Comecrudo, Cotoname, Pakawa, Carrizo
Uto-Aztecan	Hopi, Nahuatl, Papago, (Classical Aztec)
Oto-Manguean	Mazahua, Zapotec, Mixteco, Otomi
Mayan	Yucatec, Kekchi, Maya, Tzeltal, Tojolabal
Andean-Equatorial	Quechua, Aymara, Arawak, Guarani
Ge-Pano-Carib	Carib, Bororo, Witoto, Mataco
Macro-Chibchan	Cuna, Cayapa, Epera, Warao, Talamanca

Figure 8.10 shows the location of groups found in North and Central America.

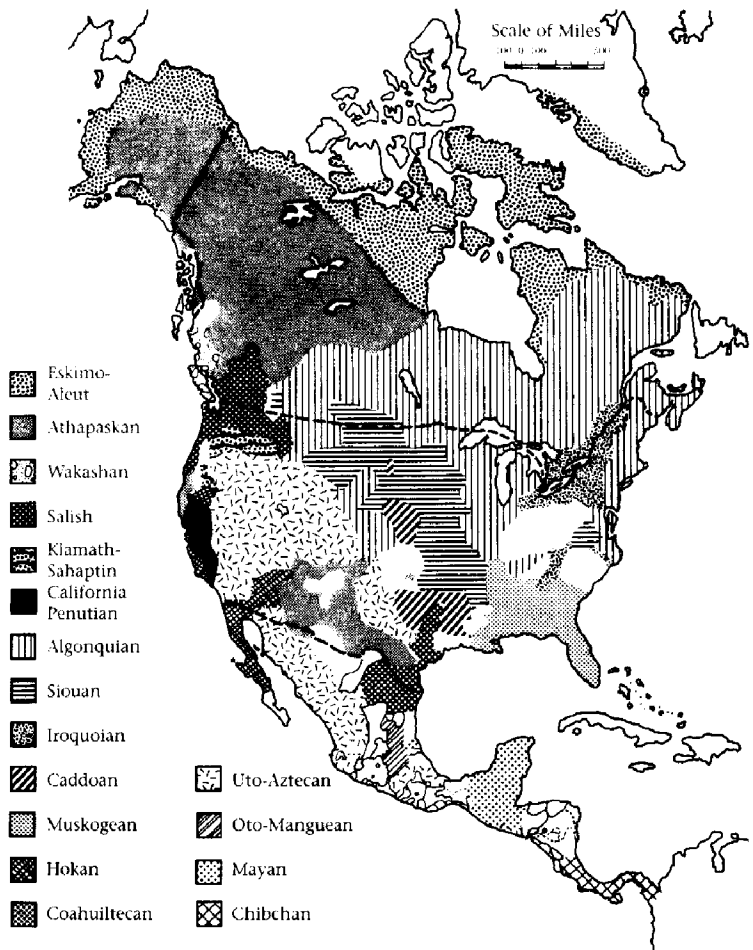

Figure 8.10 North and Central American language groups

Since the next chapter describes the native languages of the United States in some detail, it necessarily includes discussion of many of the Amerindian language families of North America. Therefore, we will restrict our focus here to some facts about a few of the indigenous languages of South America.

There are at least six hundred different Amerindian languages spoken in South America, by about eleven million people. However, our knowledge of these languages is often minimal, and some linguists estimate that there may be over a thousand South American Amerindian languages. Most of these languages belong to one of three subfamilies: Andean-Equatorial, Ge-Pano-Carib, and Macro-Chibchan.

The Andean-Equatorial subfamily contains languages that are found throughout South America, and may have as many as ten million speakers all together. The principal language in this family is Quechua, which has over six million speakers. Dialects of Quechua are spoken in Peru, Ecuador, and Bolivia. This was the language of the Inca empire, which reached its height in the sixteenth century A.D., before being destroyed by the Spanish conquistadors. Other languages belonging to this family are Aymara (Peru), Arawak (Surinam), and Guaraní (the major language of Paraguay). An interesting typological feature of some Andean-Equatorial languages is that they have no lateral consonants whatsoever.

The Ge-Pano-Carib subfamily is also spread over much of South America. Some of the languages belonging to this family are Carib (Surinam), Bororo (Brazil), Witoto (Peru), and Mataco (Argentina). Languages of the Ge-Pano-Carib family also often lack laterals; the dominant word order in these languages is usually SOV.

Languages of the Macro-Chibchan subfamily are found in Central America and the northwestern part of South America. Some languages belonging to this family are Cuna (Panama), Cayapa (Ecuador), Epera (Colombia), and Warao (Venezuela). Macro-Chibchan languages generally have SOV word order.

3.4 LANGUAGE PHYLA

In recent years attempts have been made to group many of the language families presented in Sections 3.1, 3.2, and 3.3 into even larger groupings called **phyla** (singular phylum) or **macrofamilies**. These attempts are controversial, as they challenge established views within linguistics. However, they also afford a number of intriguing possibilities. In this section we will attempt to provide a balanced view of these ventures into long-range comparison.

One of the best known of the proposed phyla is called Nostratic (also called Eurasiatic). It includes Indo-European, Uralic, Altaic, and (depending on the linguist) various other languages and language families. A number of reconstructed forms have been proposed for this phylum; two of the most convincing are the reconstructed first- and second-person singular pronouns: **m-* 'I' and **t-* 'you (sg)'. These forms are particularly persuasive, since pronoun systems are normally extremely stable and, thus, are among the most likely forms to have remained constant for the extended period of time since the existence of Proto-Nostratic (about 20,000 years ago).

Another proposed phylum is Dene-Caucasian. It includes Sino-Tibetan, Na-Dene (which includes Athabaskan), North Caucasian, and a number of other individual languages. A third proposed phylum is Austric (mentioned in Section 3.2), which includes at least Austroasiatic and Austronesian, and perhaps Daic (the family to which Thai belongs) and the Hmong-Mien (Mia-Yiao) group of languages of southern China and southeast Asia.

If all the phyla proposed to date were to be accepted by linguists, the total number of linguistic groupings would be reduced to a mere twelve (as opposed to the fifty or more that have previously been proposed) (see Table 8.25).

Table 8.25 Twelve phyla of the world's languages

1. Khoisan	7. Dene-Caucasian:	11. Nostratic/Eurasiatic:
2. Niger-Congo	a) Sino-Tibetan	a) Indo-European
3. Nilo-Saharan	b) Na-Dene	b) Uralic
4. Australian	c) North Caucasian	c) Altaic
5. Indo-Pacific	d) Nahali	d) Korean-Japanese-Ainu
6. Austric:	e) Basque	e) Gilyak
a) Austroasiatic	f) Yeniseian	f) Chuckchi-Kamchatkan
b) Austronesian	g) Burushaski	g) Eskimo-Aleut
c) Daic	8. Afroasiatic	12. Amerind
d) Hmong-Mien	9. Kartvelian (South Caucasian)	
	10. Dravidian	

Venturing still further, some linguists have even gone so far as to reconstruct a single, common ancestor for all human languages, which has been called Proto-World, or Proto-Sapiens. This ancestor language would have been spoken approximately 60,000 to 70,000 years ago. For the sake of interest, we provide a (simplified) example of one of the more than two dozen Proto-World etymologies that have been reconstructed to date:

> *Proto-World* *mena *'to think (about)'* Proposed cognates: Latin *men(s)* 'mind', Basque *munak* (pl) 'brains', Hungarian *mon(-d)* 'say', Telugu *manavi* 'prayer, humble request', Shawnee *menw* 'prefer, like', Bambara *me* 'know', Tumale *aiman* 'think', Songhai *ma* 'understand', Masa *min* 'wish'.

It is certainly possible that all human languages have descended from a single ancestor language. It is generally agreed, for example, that all human beings are closely biologically related, and recent genetic studies have even proposed that all living human beings are descended from a particular *Homo sapiens* woman (dubbed "Eve") who lived in Africa some 200,000 years ago.

Most linguists would probably agree that all human languages must have descended from a small number of languages, if not a single mother language. However, many believe that the tools used for reconstruction (the historical-comparative method) are not able to provide any linguistic evidence for long-range comparisons that go back more than about 8,000 or 10,000 years. There are a number of reasons for this.

First, the pronunciation and meaning of words can change so radically over even a much shorter period of time that cognates can become completely unrecognizable. Thus, Latin *aqua* [akwa] 'water' developed into French *eau* [o] 'water' in less than 2,000 years, Proto-Indo-European (PIE) **dwo* 'two' developed into Armenian *erku* 'two' in approximately 5,000–8,000 years, and Old English *hūswīf* 'housewife' radically changed its meaning, to become *hussy* 'a strumpet, or trollop' in less than 1,500 years. Of course, as noted in Section 3, languages also lose words altogether, which

can make it all the more difficult to uncover the cognates needed to establish a genetic relationship.

Another argument against long-range comparison has to do with the complications introduced by borrowings. For instance, for many years Thai was thought to be a Sino-Tibetan language because it contained so many Chinese loan words. However, painstaking research has finally made it clear that Thai is not Sinitic but belongs to the Daic family.

A further difficulty arises when words that appear to be cognates may have evolved independently as instances of onomatopoeia. Since onomatopoeic words (e.g., *cuckoo, vroom*) are intended to sound like real world noises, it would not be surprising to find such similar words even in unrelated languages.

Another argument against long-range comparison is that certain cross-linguistic similarities among sounds may stem from the fact that all human beings have the same vocal tract configuration. Thus, the presence of universally common sounds like /p, t, k, s, i, a, u/ in suspected cognates may be due not to a shared ancestor but rather to the fact that these sounds have articulatory and acoustic properties that favor their frequent use.

In defense of their endeavor, supporters of long-range comparison claim that loan words and onomatopoeic words are easily identifiable by the experienced researcher and can thus be easily discounted. Furthermore, language change need not affect all of the words of a language; some words retain a similar meaning and pronunciation for thousands of years. Thus, Old English *fisc* [fiʃ] apparently had the same pronunciation and the same meaning as modern English *fish* [fiʃ], despite a lapse of 1,500 years; the meaning and pronunciation of Latin *aqua* [akwa] 'water' have been maintained in Italian *acqua* [ak:wa] 'water' (2,000 years); Proto-Indo-European **dwo* 'two' changed very little in pronunciation in Latin *duo* 'two' (about 6,000 years); and Proto-Indo-European **nepot* 'nephew, grandson' had a meaning and pronunciation almost identical to those of its descendant *nepot* 'nephew, grandson' in modern Rumanian (about 8,000 years).

Supporters of long-range comparison claim that both anthropological and biological evidence show that all human beings are at least distantly related and that it is therefore plausible to believe that all human languages must be related as well. It only remains, they claim, for linguists to determine the degree of relationship among the individual language families.

This controversy is far from being resolved, and it is fair to say that most linguists remain very skeptical about the evidence and conclusions associated with comparative research involving a time depth greater than 8,000 or 10,000 years. Nonetheless, the debate has presented new and intriguing possibilities in the study of linguistic classification.

Summing up

The focus of this chapter is on the criteria that linguists use to classify languages and on the enormous variety of languages found throughout the world. Linguists sometimes attempt to classify languages solely in terms of their structural similarities and

differences (that is, in terms of their **linguistic typology**). Analysis of cross-linguistic data has identified a number of **linguistic universals**, indicating the most common characteristics of human language. The other major type of classificatory work in linguistics is concerned with **genetic relationships**—establishing language families whose members are descended from a common ancestor. While research in this area is hampered both by the large number of languages involved and the scarcity of the available data, a sizable portion of the world's several thousand languages has been placed in families. Finally, we present the controversial work recently done on linguistic **phyla** or **macrofamilies**. Research in these areas can shed light on the nature of language change, as well as the movement of peoples throughout the world.

KEY TERMS

General terms and terms concerning types of classification

areal classification
dialect
genetic classification
genetic relationships
genetically related (languages)
linguistic typology
linguistic universals
mutual intelligibility

Terms concerning typological classification

absolute universals
agglutinating language
analytic language
fixed stress languages
free stress languages
fusional language
hierarchies
implicational universals
inflectional language
isolating language
markedness theory
marked traits
polysynthetic language
synthetic languages
tone languages
universal tendencies
unmarked

Terms concerning genetic classification

Amerindian (languages)
Indo-European family

Terms concerning larger groupings of language families

macrofamilies
phyla

SOURCES

The estimate that approximately sixty-five hundred languages are spoken in the world today comes from the Web site of the Foundation for Endangered Languages at <www.bris.ac.uk/Depts/Philosophy/CTL/FEL/> (email: nostler@chibcha.demon.co.uk). The suggestion that only three hundred of these languages have a secure future is due to K. Hale, professor of linguistics at MIT, cited in *Time* magazine (23 September 1991, p. 48). For further discussion, see "Endangered Languages" by K. Hale et al., *Language* 68 (1992): 1–42.

The section on linguistic typology draws on data from the books by B. Comrie and J. Greenberg cited in Recommended Reading. Other material for this section comes from *Tone: A Linguistic Survey*, edited by V. Fromkin (New York: Academic Press, 1978); J. Hawkins's "On Implicational and Distributional Universals of Word Order," *Journal of Linguistics* 16 (1980): 193–235; M. Dryer's "The Greenbergian Word Order Correlations," *Language* 68 (1992): 81–138; *Patterns of Sounds* by I. Maddieson (Cambridge: Cambridge University Press, 1984); M. Ruhlen's *A Guide to the Languages of the World* (Stanford, CA: Language Universals Project: Stanford University, 1976); *The World's Major Languages*, edited by B. Comrie (Oxford: Oxford University Press, 1990); and the four-volume series *Universals of Human Language*, edited by J. Greenberg (Stanford, CA: Stanford University Press, 1978).

The discussion of morphological typology draws on information presented in B. Comrie's book cited below. The estimate of the relative frequency of languages in which the subject precedes the direct object is based on information in the book by W. Croft cited below. The data on OVS and OSV languages are from "Object-Initial Languages" by D. Derbyshire and G. Pullum, *International Journal of American Linguistics* 47 (1981): 192–214. The discussion of consonant systems in Section 2.4 is based on "Phonetic Universals in Consonant Systems" by B. Lindblom and I. Maddieson, *Language, Speech and Mind: Studies in Honor of Victoria Fromkin*, edited by L. Hyman and C. Li (New York: Routledge & Kegan Paul, 1988), pp. 62–78.

The section on language families is based on B. Comrie's book *The Languages of the Soviet Union* (London: Cambridge University Press, 1981); J. Greenberg's *The Languages of Africa* (Bloomington, IN: Indiana University Press, 1966); the book by M. Ruhlen cited previously; another book by Ruhlen entitled *A Guide to the World's Languages, Volume 1: Classification* (Stanford, CA: Stanford University Press, 1987); and C. F. and F. M. Voegelin's *Classification and Index of the World's Languages* (cited below). Additional data derive from C. D. Buck's *A Dictionary of Selected Synonyms in the Principal Indo-European Languages* (Chicago: University of Chicago Press, 1949); *The American Heritage Dictionary of Indo-European Roots*, revised and edited by C. Watkins (Boston: Houghton Mifflin Company, 1985); the three-volume *Russisches Etymologisches Wörterbuch*, compiled by M. Vasmer (Heidelberg: Carl Winter Universitätsverlag); "Syntactic Reconstruction and Finno-Ugric," an article by L. Campbell in *Historical Linguistics* (1987), edited by H. Andersen and K. Koerner (Amsterdam: John Benjamins, 1990); and the Proto-Baltic dictionary database developed by the author of this chapter. The maps in Figures 8.8 and 8.9 are adapted from *Problems in the Origin and Development of the English Language*, 3rd ed., by John Algeo, copyright © by Harcourt Brace & Company, reprinted by permission of the publisher. The map in Figure 8.10 is adapted from *A Guide to the World's Languages, Vol. 1*, by M. Ruhlen (Stanford, CA: Stanford University Press, 1987).

The section on language phyla is based on the book by M. Ruhlen (cited below) and on *Sprung from Some Common Source*, edited by S. Lamb and E. Mitchell (Stanford, CA: Stanford University Press, 1991).

The data for exercises 1 to 3 are found in *A Guide to the Languages of the World* by M. Ruhlen. The data for exercise 6 are from *Malagasy: Introductory Course* by C. Garvey (Washington: Center for Applied Linguistics, 1964).

Recommended Reading

Comrie, Bernard. 1989. *Language Universals and Linguistic Typology*. 2nd ed. Oxford: Blackwell.

Croft, William. 1990. *Typology and Universals*. New York: Cambridge University Press.

Greenberg, Joseph, ed. 1966. *Universals of Language*. 2nd ed. Cambridge, MA: MIT Press.

Lyovin, Anatole V. 1997. *An Introduction to the Languages of the World*. New York: Oxford University Press.

Ruhlen, Merritt. 1994. *On the Origin of Languages*. Stanford, CA: Stanford University Press.

Voegelin, C. F., and F. M. Voegelin. 1977. *Classification and Index of the World's Languages*. New York: Elsevier.

Questions

1. Which tendencies and universals are manifested in the following vowel systems? (*Hint:* Look at the pattern of the vowel systems and at rounding.)
   a) Afrikaans (South Africa) ([y] and [ø] are front rounded vowels)

i	y		u
	ø	ə	o
ɛ			ɔ
		a	

   b) Squamish (Washington State)

i		u
	ə	
	a	

2. As noted in Section 2.1, the presence of long and nasal vowel phonemes is governed by implicational universals. In what ways do the vowel systems below comply with the implicational universals that make reference to length and nasality?
   a) Maltese Arabic

i		u	i:	u:
e		o	e:	o:
	a		a:	

   b) Awji (North New Guinea)

i		u	ĩ		ũ
e	ə	o	ẽ	ə̃	õ
	a			ã	

3. Consider the following consonant systems. Do these consonant systems comply with the implicational universals mentioned in this chapter? Explain your answer.
   a) Tahitian (Tahiti)

p	t	ʔ
f		h
v	r	
m	n	

   b) Palauan (Palau Islands)

	t	k	ʔ
b			
	ð		
	s		
m		ŋ	
	l, r		

   c) Nengone (Loyalty Islands, South Pacific)—stop and nasal system only

pʰ	tʰ	ʈʰ		kʰ	ʔ
b	d	ɖ		g	
m	n		ɲ	ŋ	
m̥	n̥			ŋ̊	

   (*Note:* [ʈ] and [ɖ] are retroflex consonants; [̥] marks a voiceless nasal; [ɲ] represents a palatal nasal.)
   d) Mixe (South Mexico)

p	t		k	ʔ
	d		g	
	ts	tʃ		
	s		x	h
v			ɣ	
m	n			

4. Morphological phenomena can be classified into four types: analytic, polysynthetic, agglutinating, and fusional. Which type does each of the following languages belong to? Why?
   a) Siberian Yupik
   Angya-ghlla-ng -yug -tuq
   boat -big -get-want-3SG
   'He wants to get a big boat.'
   b) Latvian

las-u	las-ām	rakst-u	rakst-ām
read-1SG.Pres	read-1PL.Pres	write-1SG.Pres	write-1PL.Pres
'I read'	'we read'	'I write'	'we write'

   c) Japanese
   gakusei-wa homer-are-na-i
   student-Topic praise-Pass-neg-Pres
   'The student is not praised.'

5. Do a morphological analysis of the following data from Latvian. Single out each morpheme and identify its meaning. After you have segmented and identified the morphemes, describe how the data reflect the implicational universals in Section 2.2.

a)	lidotājs	'aviator (nominative)'
b)	lidotāju	'aviator (accusative)'
c)	lidotājam	'to the aviator (dative)'
d)	lidot	'to fly'
e)	rakstītājs	'writer (nominative)'
f)	rakstītāja	'writer's (genitive)'
g)	rakstīt	'to write'

6. Note the following data from Malagasy, an Austronesian language spoken on the island of Madagascar. Does Malagasy comply with all the word-order tendencies mentioned in Section 2.3?

a) amin' ny restauranta
'to the restaurant'

b) Enti'n ny labiera ny mpiasa.
brings the beer the waiter
'The waiter brings the beer.'

c) Avy any Amerika izy.
come from America he
'He comes from America.'

7. For each of the following languages on the left, determine which of the languages on the right is most closely related to it. Use the information presented in Section 3, and give reasons for your answers.

i) Macedonian	a) Albanian
	b) Ukrainian
	c) Greek
ii) Gypsy	a) French
	b) Old Egyptian
	c) Egyptian Arabic
iii) French	a) Spanish
	b) Italian
	c) Romansch
iv) Irish	a) English
	b) Icelandic
	c) Breton
v) Vietnamese	a) Laotian
	b) Mundari
	c) Burmese
vi) Modern Hebrew	a) Turkish
	b) Somali
	c) Yiddish

8. To which families do the following languages belong?

a)	Gujarati	j)	Yuma
b)	Hakka	k)	Volga Tatar
c)	Lapp	l)	Georgian
d)	Uzbek	m)	Mohican
e)	Sandawe	n)	Aramaic
f)	Quechua	o)	Flathead
g)	Faroese	p)	Telugu
h)	Maninka	q)	Javanese
i)	Santali	r)	Navajo

9. Make a list of up to fifteen languages spoken by friends and acquaintances, and identify the language family to which each belongs.