

Survey of the world's languages

The languages of the world can be divided into a number of families of related languages, possibly grouped into larger stocks, plus a residue of isolates, languages that appear not to be genetically related to any other known languages, languages that form one-member families on their own. The number of families, stocks, and isolates is hotly disputed. The disagreements centre around differences of opinion as to what constitutes a family or stock, as well as the acceptable criteria and methods for establishing them.

Linguists are sometimes divided into lumpers and splitters according to whether they lump many languages together into large stocks, or divide them into numerous smaller family groups. Merritt Ruhlen is an extreme lumper: in his classification of the world's languages (1991) he identifies just nineteen language families or stocks, and five isolates. More towards the splitting end is *Ethnologue*, the 18th edition of which identifies some 141 top-level genetic groupings. In addition, it distinguishes 1 constructed language, 88 creoles, 137 or 138 deaf sign languages (the figures differ in different places, and this category actually includes alternate sign languages — see also website for Chapter 12), 75 language isolates, 21 mixed languages, 13 pidgins, and 51 unclassified languages. Even so, in terms of what has actually been established by application of the comparative method, the *Ethnologue* system is wildly lumping!

Some families, for instance Austronesian and Indo-European, are well established, and few serious doubts exist as to their genetic unity. Others are quite contentious. Both Ruhlen (1991) and *Ethnologue* identify an Australian family, although there is as yet no firm evidence that the languages of the continent are all genetically related. At least as contentious is Joseph Greenberg's (1987) putative Amerind stock of Native American languages.

In the following subsections we present a survey the languages of the world, organized on a regional basis. We adopt a conservative (splitting) approach, distinguishing well attested families, commenting here and there on possible larger genetic groupings.

Europe and neighbouring parts of Asia

Indo-European

The Indo-European languages have been recognized as forming a family since at least the late seventeenth century, when Andreas Jäger observed in 1686 that Persian and many of the languages of Europe are descendants of a single language. Since Jäger's time, many more languages have been shown to belong to the family. Indeed, Indo-European languages are spoken throughout most of Europe, across Iran, through Central Asia, and into India. With the colonial expansions of the fifteenth to nineteenth centuries, they spread into the Americas, Australia, New Zealand, Africa, and Asia, in the process, diversifying into numerous dialects. They have become major languages in many of the former colonies, and are spoken by almost three billion (i.e. three thousand million) speakers.

The family consists of just over four hundred languages (437 according to the latest edition of *Ethnologue*), which can be grouped together into a number of subfamilies or branches, as shown in the family tree representation of [Figure 1](#). [Map 1](#) shows the approximate locations of some of the main groups.

More historical-comparative work has been done on Indo-European than any other language family, and many lexemes have been reconstructed for proto-Indo-European, as well as some of its grammar. Proto-Indo-European was an inflecting language (like ancient Indo-European languages such as Latin, Hittite and Ancient Greek), with a complex verbal system with different inflections for different persons and numbers of the subject, tense, aspect, mood, as well as case-marking for nouns.

Proto-Indo-European is widely believed to have been spoken in the south-east of Europe, perhaps in the region of Turkey, some six to eight thousand years ago. Opinions differ, however, and some argue for a more northerly location in the steppes of Russia. From the homelands the language spread east and west, in the process fragmenting into numerous mutually unintelligible languages.

It is now widely believed that the early period of Indo-European expansion that took the languages as far as India in the east and Ireland in the west, was not via military style invasions like the Roman conquests of 2,000-odd years ago.

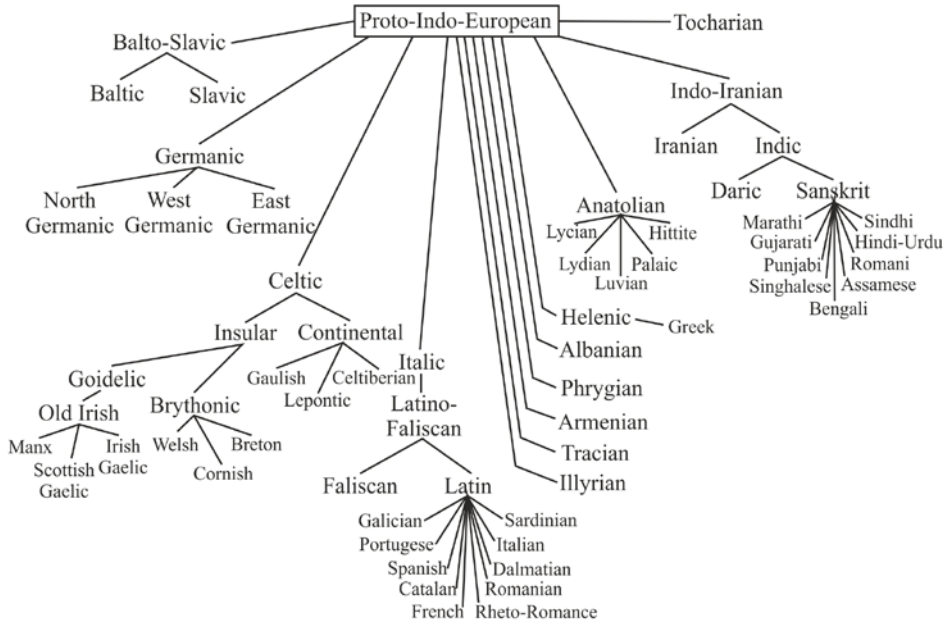


Figure 1 The Indo-European family tree (simplified and redrawn from Campbell 1998: 168).

One influential idea is that the expansion of the languages accompanied the spread of agriculture from a centre in the near east, beginning some six to eight thousand years ago (see Renfrew 1987, 1989). According to one version of the story, farmers gradually spread outwards, using land previously occupied by hunters and gatherers, eventually ousting them. Another version has it that agriculture and the language of the agriculturalists spread by diffusion, without major population movements. This story is not without difficulties, and it seems that there are some problems with the timing of some events. An alternative view is that Indo-European spread instead with the domestication of the horse and the invention of the wheel (e.g. Anthony 2007).



Map 1 Main groups in the Indo-European family.

Uralic

The much smaller Uralic family consists of some thirty-eight languages, of which Finnish and Hungarian are the best known members. Uralic languages were probably once spoken over a large area in the north-east of Europe and the south-west of Asia, but were split up by intrusions of speakers of Indo-European and Altaic languages, leaving many of them geographically isolated. Hungarian is geographically separated from its relatives as a result of migrations beginning in about the sixth century AD, and continuing until about the eleventh century.

The Uralic family splits into Finno-Ugric and Samoyedic. Finno-Ugric is further divided into a number of branches, including Balto-Finnic (e.g. Finnish and Estonian), Saamic (various Saami languages), and Ugric (Hungarian and

Khanty (Ob Ugric)). Samoyedic consists of just six languages spoken by small numbers of speakers on the northern fringe of Eurasia.

Altaic

Altaic is an uncertain grouping of at least three relatively well established families, Turkic, Tungusic and Mongolic. According to some, Korean and Japanese also belong to this genetic group, although this is contested; more usually Korean and Japanese are taken to be language isolates, although according to *Ethnologue*, Japanese represents a small language family (Japonic).

Turkic languages are spoken in a discontinuous region extending from the Balkans in the west through the Caucasus and into Central Asia and Siberia. Turkish (Turkey), Turkmen (Turkmenistan, Iran, Afghanistan), Tatar (Volga) and Uzbek (Uzbekistan) are members of this family, although the classification of the languages is not without problems.

The Tungusic family consists of a number of languages spoken by small numbers of speakers scattered over sparsely populated areas of eastern Siberia, Sakhalin Island and nearby parts of China and Mongolia. The highly endangered language Manchu is a Tungusic language.

Mongolic languages, of which *Ethnologue* lists thirteen, are spoken in Mongolia and adjacent parts of Russia and China. With about two and a half million speakers, Mongolian is the largest language in this group.

Other languages of the region

Also spoken in this large region are languages of the Caucasian families and the Chukotko-Kamchatkan family. Caucasian languages are spoken in the Caucasus region, along with Indo-European and Turkic languages. It is currently believed that Caucasian languages are constituted by two or three different families: Kartvelian (Southern Caucasian), spoken in Georgia and Turkey, and including Georgian; Northwest Caucasian spoken in the northwest of Georgia, and including Abkhaz; and Northeast Caucasian spoken in Chechnya, Ingushetia and Dagestan, and including Chechen. The Chukotko-Kamchatkan family is a small family of languages spoken on the two peninsulas with these names in far north-east Siberia. All of these languages are endangered, including the best known of them, Chukchi.

6 Survey of the world's languages

A number of language isolates are or were spoken in the regions covered in this section, including Basque (in the Pyrenees on the border of France and Spain), Etruscan (an extinct language spoken in northern Italy), Ket (spoken in western Siberia, and the only surviving member of the Yeniseian family), Yukaghir (in northeast Russia, and believed by some to be related to Uralic), Nivkh or Gilyak (spoken on Sakhalin Island and nearby parts of the mainland) and Ainu (today spoken mainly on Hokkaido, the northern island of Japan).

Larger groupings?

According to one proposal, Indo-European, Uralic, Altaic, Afroasiatic (see pp.394–395 of the textbook, repeated [below](#)), Kartvelian, Dravidian, possibly as well as others, form a single genetic family, dubbed Nostratic. A similar suggestion by Joseph Greenberg groups the first three of the languages together with Chukotko-Kamchatkan and Eskimo-Aleut in a Eurasiatic family. Neither proposal is widely accepted.

Africa and southwest Asia

Afroasiatic

Afroasiatic consists of some 366 languages (according to Ethnologue) spoken in northern Africa and southwest Asia by over 360 million people – see [Map 2](#). It is regarded as the best established of the four families that African languages are sometimes divided into; the other three families are Niger-Congo (on which see next subsection), the more contentious Nilo-Saharan and the highly contentious Khoisan (see [below](#)).

Afroasiatic is generally divided into six groups: Berber (consisting of around 30 languages spoken in Morocco, Algeria, Tunisia, Mali, including Tamazight, Zenaga, and Kabyle); Chadic (made up of nearly 200 languages spoken in Nigeria, Chad, Cameroon, including Hausa, Miya, and Ngizim); Cushitic (with about 50 languages in Ethiopia, Eritrea, Somalia, Kenya and Tanzania, including Somali, Dahalo and Afar); Egyptian (one language, Coptic, which became extinct in the fourteenth century, though it is still used as a language of religion); Semitic (consisting of some 80 languages spoken in Ethiopia and the Middle East, including Arabic, Hebrew, Aramaic, Amharic and Tigré); and Omotic (with 30 or



Map 2 Putative language families of Africa.

so languages spoken mainly in Ethiopia, including Dizi, Bench and Ganza).

Semitic is the only group spoken widely outside of Africa. It is also the best-studied group. A notable feature of Semitic languages is a root structure consisting of three consonants; grammatical information is expressed largely through intervening vowels. For instance, the root form for 'book' in Arabic is *k-t-b*; thus *kitab* 'book', and *kutub* 'books'.

It is generally believed that proto-Afroasiatic was spoken around 10,000 years

8 Survey of the world's languages

ago in northeast Africa, and spread out from there, perhaps moving into southwest Asia around 8,000 years ago. Shared agricultural vocabulary is lacking across the family, suggesting that the languages dispersed before the advent of agriculture. On the other hand, cognate pottery terms are widely distributed, suggesting that pottery was known at the time of the proto-language.

Niger-Congo

Consisting of just over 1,500 languages, the Niger-Congo family is the largest language family in Africa, indeed in the world. This must be tempered by the observation that some linguists have expressed doubt concerning the status of Niger-Congo as a genetic unit. This is because the proto-Niger-Congo has not been reconstructed, and thus the genetic unity of the languages is not established fact. Nonetheless, the majority of specialists accept it as a viable genetic grouping, and regard it simply as a matter of time until this will be convincingly demonstrated.

Niger-Congo languages are spoken over a vast area of the African continent, as shown in [Map 2](#), and by almost 450 million speakers.

The composition of the family is not uncontroversial, and has been revised more than once. A recent classification of the family is shown in the family tree of [Figure 2](#). Some nodes on this tree represent individual languages (e.g. Pre/Berε), some represent small groups of languages (like Dogon), while others represent very large groups (e.g. Bantoid).

The well-known Bantu languages are a subgroup of the Bantoid group. They comprise over 400 languages (including, for example, Swahili, Fang, Setswana, Southern Sotho, Luganda and Shona), with perhaps 60 million speakers. It is believed that Bantu is a rather young group, that began diverging when speakers spread out from Cameroon perhaps only about 4,000–5,000 years ago. Bantu speaking people migrated over the next 3,000 years, taking West African yam agriculture with them. Today Bantu languages are spoken across a third of the African continent.

One characteristic of Niger-Congo languages is their possession of an elaborate system of noun classes (see note 2, Chapter 7, p.177), distinguishing humans, animals, plants, masses and liquids, abstracts, and so on. The classes are marked by affixes, usually prefixes, that occur sometimes on the noun, but usually

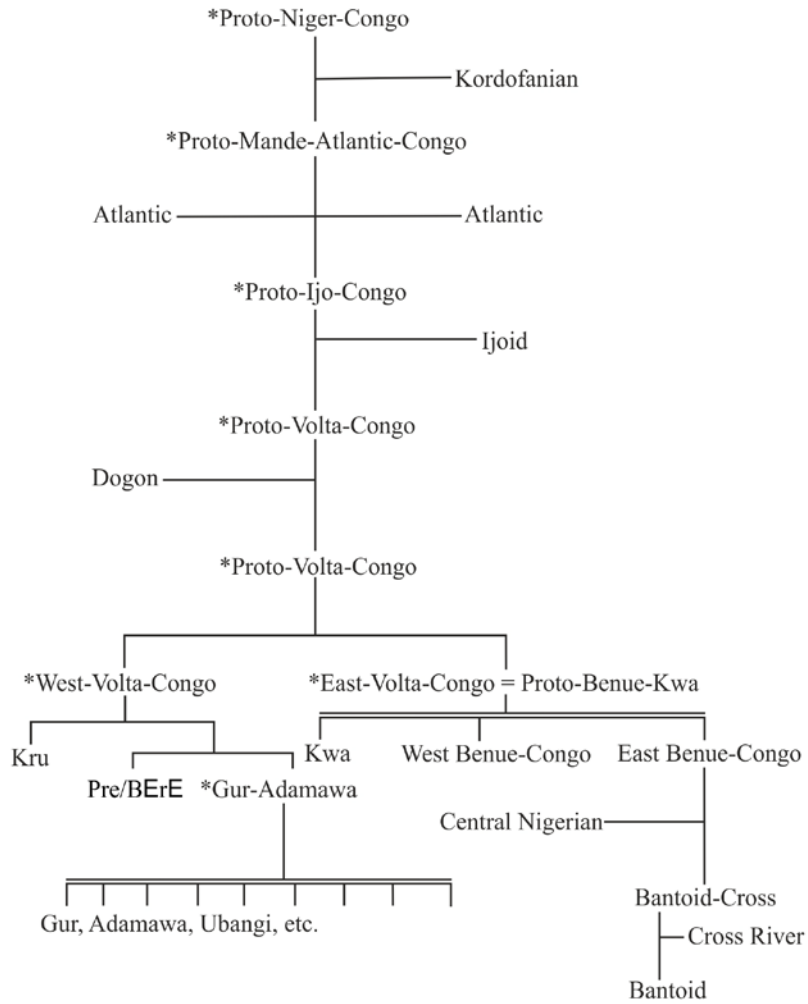


Figure 2 One representation of the Niger-Congo family (based on Williamson and Blench 2000: 18, redrawn and simplified)

on adjectives and verbs in agreement with the noun they apply to, as shown by the following example, where *ki-* and *-ki* are the class markers:

- (1) *ki-tu hi-ki ki-kubwa ki-lianguka* Swahili
 ki-thing this-ki ki-large ki-fell
 'This large thing fell.'

Nilo-Saharan

The Nilo-Saharan grouping, consisting of some two hundred languages, is generally regarded as more controversial than either Afroasiatic or Niger-Congo. [Figure 3](#) shows one opinion of the constitution of Nilo-Saharan.

A few linguists have proposed a single genetic family consisting of Nilo-Saharan and Niger-Congo. Thus, according to a suggestion by Roger Blench (1999), the entire Niger-Congo family fits under a single low node in the Nilo-Saharan family tree.

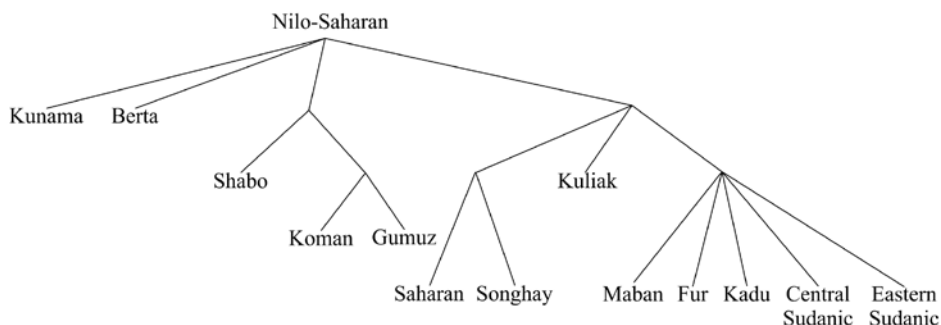


Figure 3 Structure of Nilo-Saharan according to Blench (2010) (redrawn)

Khoisan

The term Khoisan (also spelt Khoesan) is a cover term for a small group of around thirty languages of Africa that have clicks as part of their normal phoneme inventory, but are not Bantu or Cushitic. Greenberg's classification of African languages (1963) identified Khoisan as one of four genetic macro-units covering the continent. However, there is no evidence that the languages form a genetic unit, and specialists use the label Khoisan as a convenience label for a residue

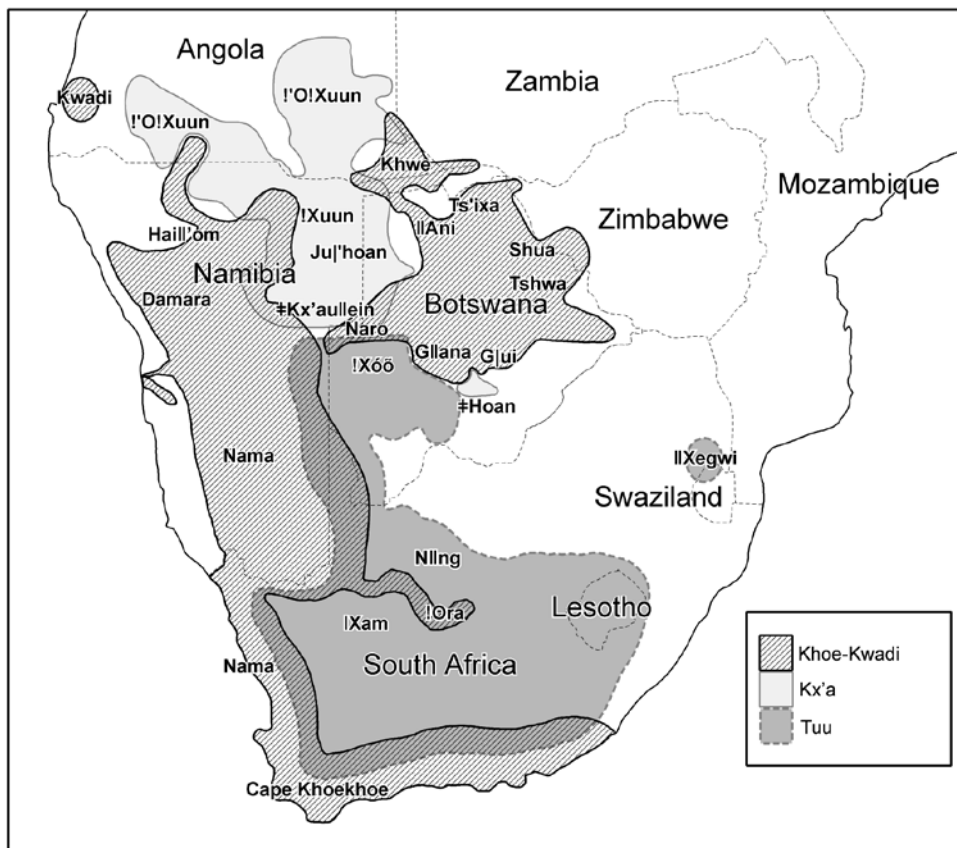
class of languages that don't fit into the better supported family units.¹

Khoisan is a discontinuous group made up of two languages spoken in Tanzania, Hadza and Sandawe, the remainder being spoken in southern Africa. Those spoken in southern Africa are generally accepted by experts as comprising three distinct genetic lineages (e.g. Vossen 2013: 3): Khoe-Kwadi (fourteen languages, including Hailom, Nama-Damara, Khwe, Shua, Ts'ixa and Kwadi),² Kx'a (e.g. seven languages, including Ju|'hoan and †Kx'aul'ei) and Tuu (six languages, including !Xóǀ, Nǀng and !Xegwi). [Map 3](#) shows the locations of most southern African Khoisan languages. Current consensus among experts is that Hadza and Sandawe are language isolates; there is however some evidence that Sandawe may be very distantly related to Khoe-Kwadi.

Khoisan languages are famous for their possession of click consonants. Some languages, for instance !Xóǀ – the language with the largest known consonant inventory in the world, with well over 100 distinct consonant phonemes – distinguish five different click phonemes, bilabial (ǀ), dental (ǃ), (post)alveolar (ǁ), palatal (ǂ) and lateral (ǁ̥). Each of the clicks may be accompanied by some further modification by changes to the manner of the velar or uvular closure. (Check your understanding of click articulation, p.40.) In !Xóǀ each click admits up to sixteen accompaniments, including voicing, aspiration, nasalization and glottalization. Vowel systems in Khoisan languages usually distinguish five phonemic vowel qualities. However, additional colouring of vowels is normally possible. Almost all Khoisan languages show distinctive nasalization of vowels; in addition, glottalized, breathy and pharyngealized vowels are commonly phonemic.

1 Unlike previous editions of *Ethnologue*, the latest edition no longer identifies *Khoisan* as a genetic unit.

2 Kwadi is an extinct language of which very little is known. Available evidence suggests that it belongs in a genetic group with Khoe languages (Güldemann 2004).



Map 3 Location of Khoisan languages of southern Africa

Southern and eastern Asia

Dravidian

The seventy-five Dravidian languages are spoken mainly in India, but also in Pakistan and Nepal; they are the dominant languages of Southern India. The family is divided into four groups: Northern (e.g. Brahui, with 2 million speakers in Pakistan, and Nepali; Kurux with some 28,000 speakers in Nepal); Central (all with smallish numbers of speakers, including Northwestern and Southeastern Kolami, with about 50,000 and 10,000 speakers, respectively); South-Central (e.g.

Telugu with about 70 million speakers); and Southern (e.g. Tamil with around 60 million speakers, and Malayalam with about 35 million).

Austro-Asiatic

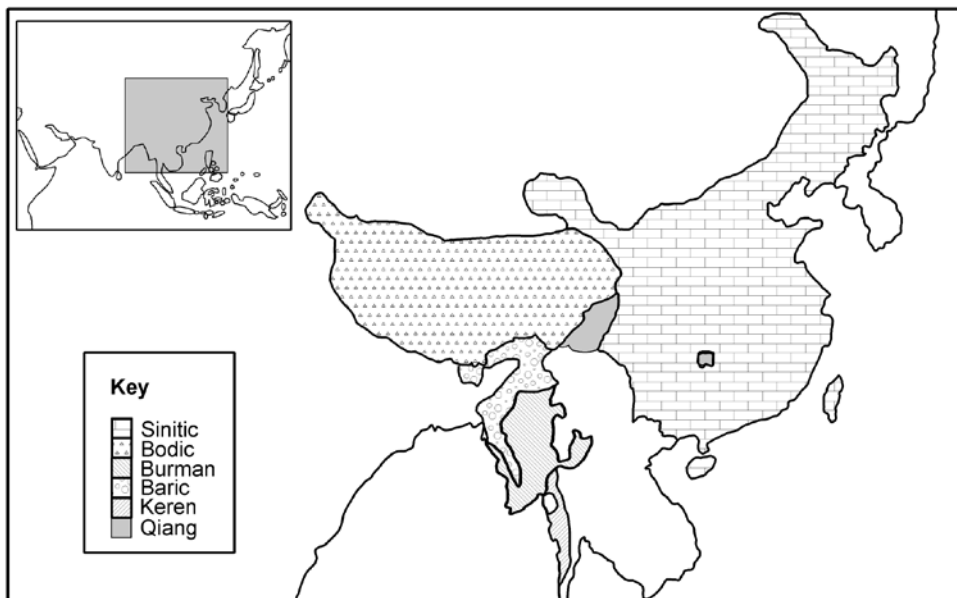
Austro-Asiatic languages are spoken from eastern India to Vietnam, and south into Malaysia, though interspersed amongst other more widely spoken languages. There are two main branches. In the west is the Munda group (21 languages), spoken in eastern India and neighbouring regions. Mundari and Santali are among the few Munda languages with fairly large communities of speakers, around 2 and 5 million respectively. In the east are the Mon-Khmer languages, which number almost 150, and occur in greatest concentration in Thailand, Malaysia, Cambodia, Laos and Vietnam. Except for Vietnamese and Khmer (Cambodia), Mon-Khmer languages have few speakers.

Sino-Tibetan

Comprising over 400 languages, Sino-Tibetan is the second largest language family of the world in terms of numbers of speakers, with a bit under half the number of speakers of Indo-European. It includes Mandarin Chinese, the language with the largest number of native speakers.

Sino-Tibetan falls into two groups. One, Sinitic, consists of just 14 languages, including Mandarin Chinese, Cantonese (Yue), Hakka, Northern Min, Southern Min and Gan. The other group, Tibeto-Burman, has some 450 languages, mainly spoken in China, Nepal and India. Groupings within Tibeto-Burman include, according to the traditional classification: Baric (e.g. Meithei in India), Bodic (e.g. Tibetan), Burmese-Lolo (e.g. Burmese), Keren (various Keren languages spoken in Myanmar and Thailand, the most widely spoken being S'gaw Karen), Nung (e.g. Norra, Nung) and Qiang (e.g. Northern and Southern Qiang, spoken in China). [Map 4](#) shows the location of the family and most of the groups.

With the exception of Baric languages, Sino-Tibetan languages are mostly tone languages. Tone cannot, however, be reconstructed for proto-Sino-Tibetan. Rather, certain syllabic endings of the proto-language gave rise to the tones of the modern languages.



Map 4 Location of the Sino-Tibetan family

Other languages of southern and eastern Asia

Tai-Kadai is a family of some 70 languages spoken in southeastern Asia, mainly in China, Vietnam, Laos, and Thailand, although a few are spoken in India and Myanmar. It is divided into three groups, Hlai, Kadai, and Kam-Tai. Thai (Thailand) and Lao (Laos) have the largest numbers of speakers. The Hmong-Mien (or Miao-Yao) family, consisting of some 30 languages spoken mainly in southern China and into Vietnam, is divided into two groups, Hmong and Mien. Hmong-Mien languages have smallish numbers of speakers; Hmong Njua is the most widely spoken language, with about a million speakers.

Language isolates in the region include Burushaski (northern Pakistan) and Nihali (central India, not to be confused with Nahali, an Indo-European language). Elamite, an extinct language spoken in southwestern Iran around 1000 BC and Sumerian, perhaps the first language to have been written, are both of uncertain genetic lineage, although some regard Elamite as related to Dravidian.

Oceania

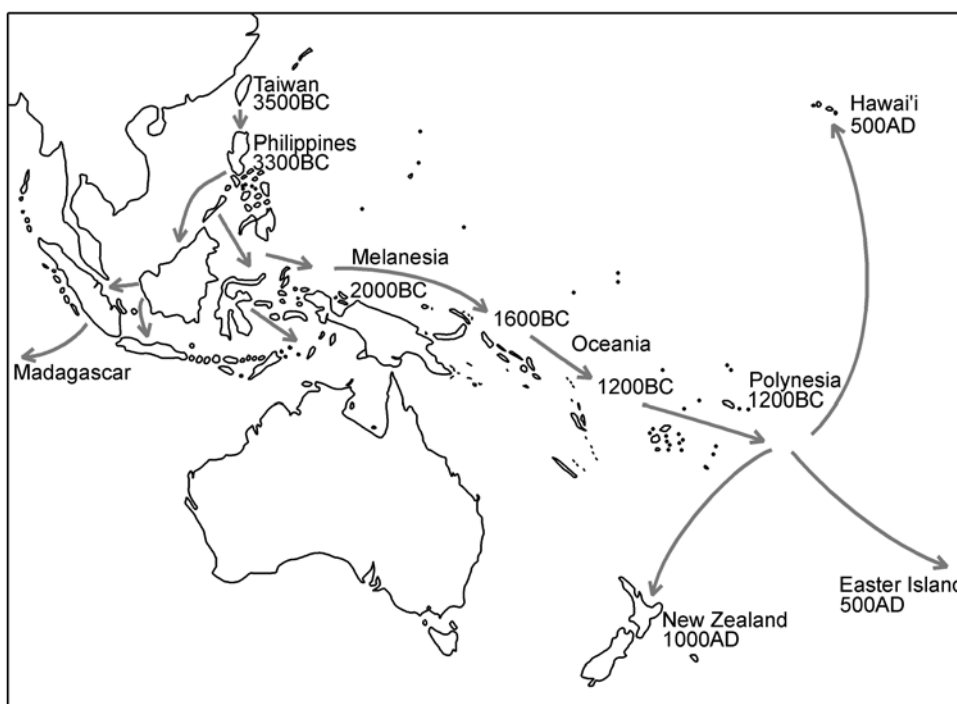
Austronesian

Austronesian is the largest universally accepted language family in the world with over 1,200 languages, spoken by some 300 million speakers from Madagascar in the west to Easter Island in the east, Taiwan in the north and New Zealand in the south, with the exception of Australia and most of the island of New Guinea. (The [Niger-Congo family](#) (see p.394) is the only larger family, but it is more contentious.)

As for Indo-European, a good deal of proto-Austronesian has been reconstructed. There are, however, differences of opinion concerning how the family is structured. One view is that it is divided into four groups, three of which – Atayalic, Tsouic and Paiwanic – are located on the island of Taiwan. Other proposals identify up to nine groups on Taiwan. Just one branch, Malayo-Polynesian, accounts for the bulk of the languages of the family, and includes all Austronesian languages outside Taiwan. Malayo-Polynesian is subdivided into four groups, Central Malayo-Polynesian, South Halmahera-West New Guinea, Oceanic (eastern group) and Western Malayo-Polynesian. Regardless of the actual structure of the family, it is clear that there is considerably greater diversity in the languages of Taiwan than in all of the rest of the languages. It is generally assumed that the region of greatest diversity is the most likely homeland, the region where the proto-language was spoken, since it is in this region that the languages have been longest that they have had the most opportunity to diversify. Taiwan is thus the most likely homeland for Austronesian.

Evidence from archaeology is largely in agreement with linguistic evidence that Taiwan was the homeland of Austronesian, and that the languages began spreading from there some 5,500 or so years ago. The languages spread via migrations of people travelling over the sea, and taking farming with them. The island of New Guinea was reached some 4,000 years ago, and New Zealand about 1000 AD. [Map 5](#) gives an idea of the dispersal of Austronesian languages.

It has recently been proposed that the Austronesian languages are genetically related to the [Sino-Tibetan languages](#) (see p.398), forming a large Sino-Tibetan–Austronesian family. Laurent Sagart (2005) makes a plausible (though not widely

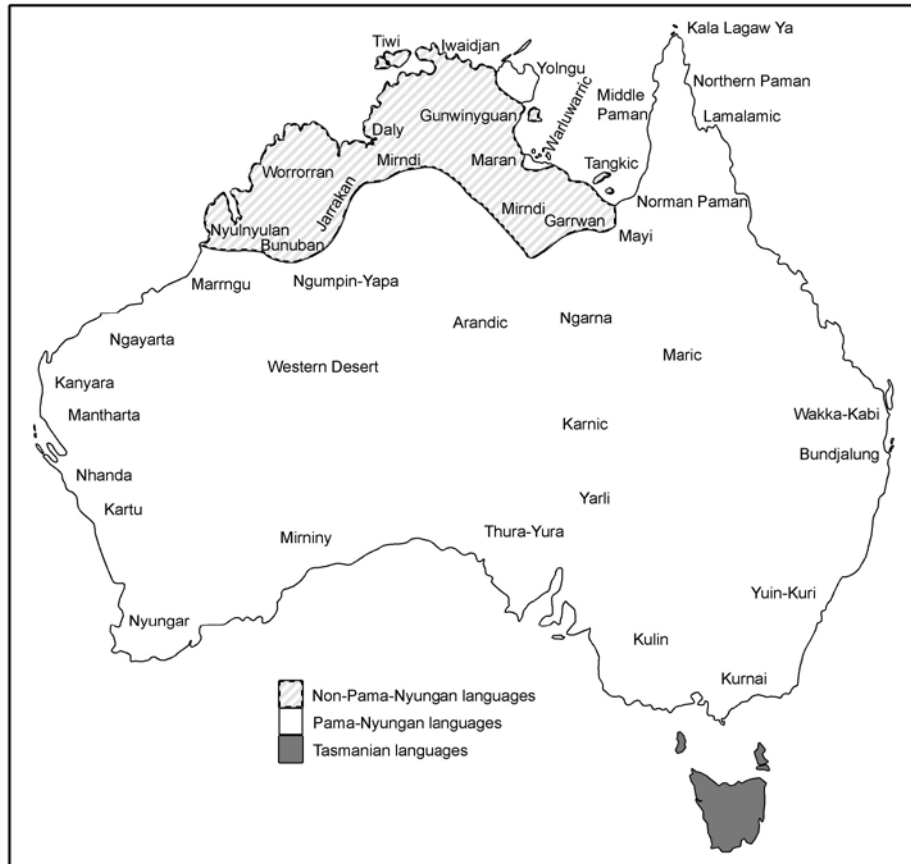


Map 5 Dispersal of Austronesian languages
(adapted from Gray and Jordan 2000: 1053)

accepted) case for this macro-group, identifying some 60 cognates in basic vocabulary among Austronesian and Sino-Tibetan languages, as well as recurrent sound correspondences. He avers that there is archaeological evidence in agreement with his proposals, and that the initial spread of the proto-language for this family was from mainland China to Taiwan, accompanying a migration of agriculturalists driven by population expansion. The archaeologist Peter Bellwood is in basic agreement, though he places the ultimate mainland China homeland in a different location (Bellwood and Sanchez-Mazas 2005).

Australia

Probably around 260 distinct and mutually unintelligible languages were spoken in Australia prior to European colonization. (The number would be closer to 500 if we define languages by political criteria; the 18th edition of *Ethnologue* gives an



Map 6 Languages of Australia, showing some non-Pama-Nyungan families and Pama-Nyungan subgroups

intermediate figure, 378.) A number have since passed out of everyday usage as primary codes of communication; almost all are currently endangered, and only a few are being passed on to children as their mother tongues.

As already mentioned, it is not known for certain whether or not the languages are ultimately genetically related (and a few linguists believe that we will never know). The current consensus is that the languages belong to some twenty or so families, one of which, Pama-Nyungan, is spread over three-quarters of the continent. The other quarter of the continent is home of the remaining families, which are conventionally grouped together as non-Pama-Nyungan — there is no implication that this is a genetic family. (The status of Pama-Nyungan as a

genetic family is also disputed by a few Australianists. Most Australianists, however, believe that ultimately it will be shown that they are related languages, by reconstruction of proto-Pama-Nyungan.) [Map 6](#) shows the geographical division between Pama-Nyungan and non-Pama-Nyungan.

Precious little is known about the now extinct languages of Tasmania, and insufficient information was recorded on them to permit any viable classification. The notion that they belong to an Indo-Pacific genetic family that includes most of the languages, other than Austronesian ones, spoken in this vast region, from the Andaman Islands in the Bay of Bengal through New Guinea, Australia, and Tasmania (Greenberg 1971) has no empirical foundation. Nor is there even strong evidence that they are related to any languages of mainland Australia — or that they are not.

New Guinea and Melanesia

New Guinea and nearby islands is perhaps the most linguistically diverse region on earth, with around 1,200 languages spoken by some six to seven million people. About a quarter of the languages, mostly located on the smaller islands and on parts of the coast of the island of New Guinea, are [Austronesian](#), and belong to the Central-Eastern Malayo-Polynesian subgroup. The remainder, which number over 800, are called Papuan, a term that includes all of the languages of the region other than Austronesian. (Like the term non-Pama-Nyungan, this is a negative characterization, and does not imply the languages form a genetic unit.)

Papuan languages fall into 30 or more distinct genetic families and about two dozen isolates. The average family consists of 25 languages, each with an average of less than 3,000 speakers. The largest language, Enga, has around 200,000 speakers; many languages have less than 100, even less than 50 speakers.

The largest Papuan family (or stock), Trans-New Guinea, consists, according to *Ethnologue*, of some 480 languages belonging to dozen groups and spoken mainly along the mountainous cordillera of New Guinea. Proto-Trans-New Guinea has not been reconstructed; indeed, there is considerable diversity of opinion as to the structure and composition of the putative family. The *Ethnologue* grouping is a 'lumping' one basically following Wurm (1975); more recently, Foley (2000) has suggested that it consists of around 300 languages, and

Ross (2005) has proposed a version with 400 or so languages. Other Papuanists accept the core of the Trans-New Guinea grouping, with a good deal of uncertainty as to which languages and groups belong to it. Even in its most reduced form Trans-New Guinea fits into the category of large language family.

The Americas

The classification of the indigenous languages of America has been the subject of much controversy in recent years. On the lumping side, Joseph Greenberg has proposed (1987) three stocks or phyla, i.e. large groupings of genetic families: Eskimo-Aleut (widely accepted), Na-Dene (see [below](#)), and Amerind, containing all other languages of North America, Meso-America and all languages of South America. Amerind in particular is highly contentious, and is not supported by application of the comparative method; it is almost universally rejected by Americanists. On the splitting side, Lyle Campbell recognises some 200 distinct families (1997), with no higher groupings into stocks.

North America

Some 300 distinct languages were spoken in what is now the USA and Canada before the advent of Europeans. A recent survey, Mithun (1999), recognizes 50 families; some of the main ones are shown in [Map 7](#).

In the far north is the Eskimo-Aleut family, which stretches from Eastern Siberia across the northern fringe of the continent, and then on to Greenland. The Aleut group is spoken on the Aleutian Islands in and near the Bering Strait. The Eskimo group is divided into Yup'ik, spoken in Siberia and Alaska, and Inuit, spoken in northern Alaska, Canada, and Greenland.

One of the largest families in north America is Athapaskan, spoken over a large area of Alaska and the northwest of Canada; Athapaskan languages are also found in two geographically separate regions, Pacific Coast Athapaskan in Oregon and California, and Apachean in southwestern USA. Navajo, one of the most vigorous of the traditional languages of north America, is an Athapaskan language. Athapaskan forms the core of Greenberg's proposed Na-Dene family, which includes Tlingit and Haida as well, although these are uncertain genetic relatives.



Map 7 Some major languages and language families of North America

Other major families are Algonquian (which includes Algonquian languages covering much of the northeast of the continent, and two languages in California), Iroquoian (spoken around the Great Lakes), Siouan (spoken in a large part of the plains region of the USA), and Uto-Aztecan (spoken in the Great Basin, in the western part of the USA, and extending into Meso-America). A number of small families, including Wakashan, Miwok-Costanoan, and Yuman are located along the Pacific Northwest, and a number of isolates are found mainly on the Pacific coast and south-east of the USA.

Meso-America

Aside from Uto-Aztecan, Meso-American languages belong to four main families: Otomanguan (a large family of around 170 languages spoken in

Mexico), Mixe-Zoquean (a small family of 16 languages spoken in a number of geographically separated areas in Mexico), Mayan (about 70 languages spoken in Mexico, Belize and Guatemala), and Chibchan (some 22 languages spoken in Panama, Costa Rica, Honduras and Colombia).

South America

Around 1,000 traditional languages are or were spoken on the continent of South America, of which some 300 are spoken in the Amazon Basin. Since the beginning of European colonization many have disappeared, some without trace; it is presumed that many more will disappear in the next century. Brazil is home to some two hundred indigenous languages, of which more than half have less than 1,000 speakers, and only four have more than 10,000 speakers. The Indo-European languages English, Spanish, Portuguese, or French now dominate in all South American countries.

According to one classification, the traditional languages belong to some 118 genetic units, 48 of which are families of two or more languages, the remainder being language isolates. One widely distributed group is the Arawakan family, which is scattered over much of the northern part of the continent. It consists of some 60 languages, most of which belong to the Maipuran group. Some other large families are: Tupi-Guaraní, consisting of some 70 languages in the southern Amazonian region, including Paraguayan Guaraní, spoken by 95% of the population of the country, and one of its national languages; Macro-Gê, consisting of 32 languages spoken in the highlands of Brazil; Carib, consisting of some 30 languages from northern Brazil, Venezuela and the Guianas; Tucanoan, made up of some 26 languages centred in western Amazonia; and Quechuan, a family of some 47 languages widely spoken in the Andean region, and the largest family of indigenous languages in the Americas in terms of numbers of speakers (over ten million). Smaller families include: Tacanan, a group of 6 languages spoken in Bolivia; Chon, consisting of two languages spoken in Argentina and Patagonia; Aymaran, consisting of three language spoken by some two million people in Bolivia and Peru; and Jivaroan, a family of four languages spoken in Peru and Ecuador.

Guide to further reading

A good overview of the linguistic diversity of the world is Comrie (2001), which provides basic information on many language families. There are a number of more comprehensive book-length treatments of the world's linguistic diversity. My recommendation is Pereltsvaig (2012), which provides much intriguing information on a range of languages and their speakers. Anderson (2012) is an accessible overview. Garry and Rubino (2001) contains basic information on 191 languages from all over the world, including the viability of the language, use in education, genetic classification and basic grammar. Each description is by a linguist with some knowledge of the language, and it is a very useful resource. An older, but still useful survey, is Lyovin (1997), which gives grammatical information on a selection of languages. Ruhlen (1987, revised version 1991), is a highly speculative classification of the world's languages, and few of the proposed groupings are accepted by experts.

References

- Anderson, S. R. (2012), *Languages: a very short introduction*. Oxford: Oxford University Press.
- Anthony, D. W. (2007), *The horse, the wheel, and language: how bronze-age riders from the Eurasian Steppes shaped the modern world*. Princeton and Oxford: Oxford University Press.
- Bellwood, P. and Sanchez-Mazas, A. (2005), 'Human migrations in continental east Asia and Taiwan: genetic, linguistic, and archaeological evidence'. *Current Anthropology*, 42, 480–484.
- Blench, R. (1999), 'The languages of Africa: macrophyla proposals and implications for archaeological interpretation', in R. Blench and M. Spriggs (eds), *Archaeology and language IV: language change and cultural transformation*. London: Routledge, pp.29–47.
- Campbell, L. (1997), *American Indian linguistics: the linguistic history of Native America*. New York: Oxford University Press.
- Childs, G. T. (2003), *An introduction to African languages*. Amsterdam and Philadelphia: John Benjamins.

- Comrie, B. (2001), 'Languages of the world', in M. Aronoff and J. Rees-Miller (eds), *The handbook of linguistics*. Oxford: Blackwell Publishers, pp.19–42.
- Foley, W. A. (2000), 'The languages of New Guinea'. *Annual Review of Anthropology*, 29, 357–404.
- Garry, J. and Rubino, C. (eds), (2001), *Facts about the world's languages: an encyclopedia of the world's major languages, past and present*. New York and Dublin: The HW Wilson Company.
- Greenberg, J. H. (1963), *The languages of Africa*. Bloomington, IA: Indiana University, and The Hague: Mouton.
- Greenberg, J. H. (1971), 'The Indo-Pacific hypothesis', in T. A. Sebeok (ed.), *Current trends in linguistics, 8: Linguistics in Oceania*. The Hague: Mouton, pp.808–871.
- Greenberg, J. H. (1987), *Language in the Americas*. Stanford: Stanford University Press.
- Güldemann, T. (2004), 'Reconstruction through de-construction: the making of person, gender and number in the Khoe family and Kwadi'. *Diachronica*, 21, 251–306.
- Güldemann, T. and Vossen, R. (2000), 'Khoisan', in B. Heine and D. Nurse (eds), *African languages: an introduction*. Cambridge: Cambridge University Press, pp.99–122.
- Lyovin, A. V. (1997), *An introduction to the languages of the world*. New York and Oxford: Oxford University Press.
- Mithun, M. (1999), *The languages of native North America*. Cambridge: Cambridge University Press.
- Pereltsvaig, A. (2012), *Languages of the world: an introduction*. Cambridge: Cambridge University Press.
- Renfrew, C. (1987), *Archaeology and language: the puzzle of Indo-European origins*. London: Jonathan Cape.
- Renfrew, C. (1989), 'The origins of the Indo-European languages'. *Scientific American*, October, 1989, 82–90.
- Ross, M. D. (2005), 'Pronouns as a preliminary diagnostic for grouping Papuan languages', in A. Pawley, R. Attenborough, R. Hide and J. Golson (eds), *Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*. Canberra: Pacific Linguistics, pp.15–66.

24 Survey of the world's languages

- Ruhlen, M. (1991), *A guide to the world's language. Volume 1: Classification (with a postscript on recent developments)*. London: Edward Arnold.
- Sagart, L. (2005), 'Sino-Tibetan–Austronesian: an updated and improved argument', in L. Sagart, R. Blench and A. Sanchez-Mazas (eds), *The peopling of East Asia: putting together archaeology, linguistics and genetics*. London: RoutledgeCurzon, pp.161–176.
- Vossen, R. ed. (2013), *The Khoesan languages*. London and New York: Routledge.
- Wurm, S. A. (ed.) (1975), *New Guinea area languages and language study. Volume 1: Papuan languages and the New Guinea linguistic scene*. Canberra: Pacific Linguistics.