

TUTORIAL

5.20

Language Families

This tutorial looks at the various ways that separate languages might be connected to one another, and why some languages are said to be members of the same language family.

Introduction

There are a variety of ways that languages can be connected to one another. One important way is by being members of the same *language family*.

Change is a natural process in language, and all languages are always changing. No matter how old you are, during your lifetime you have probably noticed new words and phrases being introduced, or existing ones changing their meaning, or and some falling out of use all together. Over time, because of things like geographical, cultural or religious factors, a language can change differently in different places, and diverge into separate dialects, then eventually into separate languages. Those separate languages are related to each other in a single language family. We say they are in the same family if they come from the same original source language. For example, the Romance languages are all derived from Latin. This is a linguistic result of the wide influence of the Roman Empire.

Often there is no direct information about the ancestor language because most ancient languages were never written down. But the family relationship between languages can be established by identifying relationships between their words and sounds. Linguists can then reconstruct aspects of the common ancestral languages that they came from.

There are written records of Latin, the ancestor of the Romance languages, so the family relationship is easy to establish, but there are no written records of the ancestors of the Germanic or Celtic languages. Linguists have worked to analyse these languages and have established that they are in fact descendants of an original Germanic language and Celtic language.

Language families

Today several hundred unrelated language families in the world have been recognised, but the exact relationships between languages and subgroups are regularly questioned and reassessed.

Most of the world's languages can be grouped into 40 or 50 major language families. These 'major' families are ones that have a large number of languages and/or a large number of speakers. There are only a few dozen languages that are 'isolates' - languages that can't be shown to be related to any other language, e.g. Basque (South-western France and North-eastern Spain).

The top four major language families are:

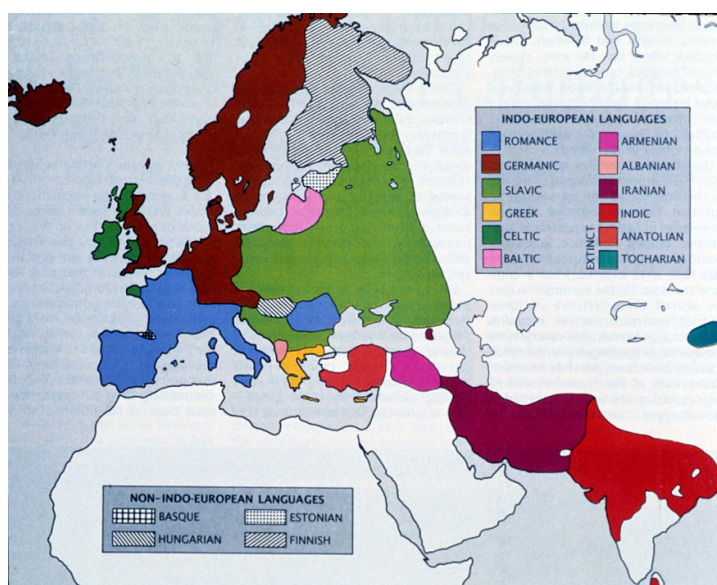
- *Indo-European* - which has well over 2 billion speakers
- *Sino-Tibetan* - with well over 1 billion speakers
- *Niger-Congo* - which has half a billion speakers
- *Afro-Asiatic* - with a quarter of a billion speakers

Indo-European

The Indo-European language family is very large, and includes most of the languages of Europe, but it also extends beyond Europe.

Many of the languages of western Asia (like Persian in Iran, Pashto in Afghanistan, and Kurdish) and of the northern half of the Indian subcontinent (such as Hindi/Urdu, Punjabi and Bengali) are Indo-European languages.

European colonial expansion in the last few centuries has also taken Indo-European languages around the world, and in many regions outside Europe they are spoken as the first language by most of the population, like in North and South America and Australia.



Language families in the region of Australia

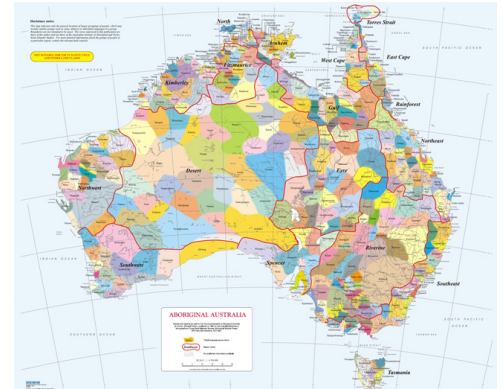
The three main linguistic groupings in this region are Australian (i.e. Aboriginal languages), Austronesian, and Papuan. The region has some of the greatest

linguistic diversity in the world, containing about 1,600 languages - one quarter of all languages.

Australian

Australian Aboriginal languages belong to a single language family, because research has concluded that they are descended from a common ancestor language. At the time of colonisation by Europeans in 1788 there were around 260 Aboriginal languages. Today about 145 are still spoken. An interactive map of Australian indigenous languages is available here - take a look at it:

<http://www.abc.net.au/indigenous/map/default.htm>



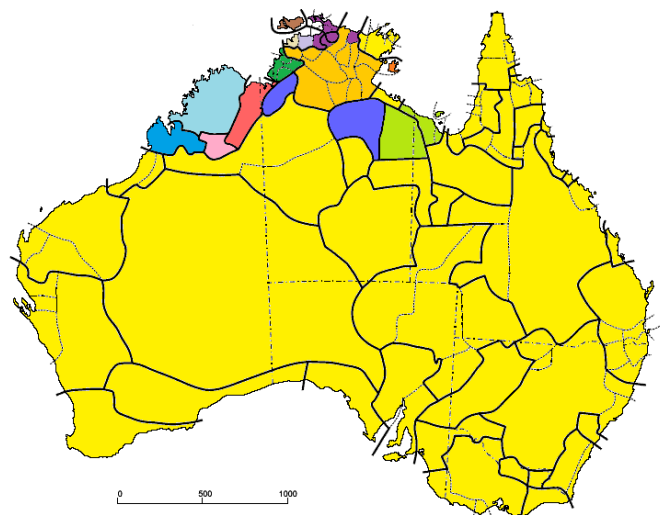
Very few of these remaining languages are spoken by more than a few hundred people, and many by only a handful of elderly people. Half have fewer than 10 speakers. Only 18 Australian languages are being learned by children today, which means that the rest will be extinct when the current generation dies out.

A small handful of Aboriginal languages are spoken by over a thousand people. The largest Aboriginal languages, with more than 2,000 speakers (each with several dialects) are:

- *Western Desert* (Central Australia, 5,800 speakers),
- *Kala Lagaw Ya* (Western Torres Strait islands, 3,000-4,000 speakers),
- *Warlpiri* (Central Australia, 3,000 speakers),
- *Eastern Arrente* (Central Australia, 2,175 speakers).

The Australian language family has several branches. Across most of Australia the languages belong to one branch, called *Pama-Nyungan*. This includes all of Australia except the Top End (the far north of the Northern Territory and Western Australia) and is shown in yellow on the map at right.

The other languages belong to about a dozen branches that get grouped together as *Non-Pama-Nyungan*, but these branches are not more closely related to each other. So, there are about a dozen branches of the language family 'Australia', of which Pama-Nyungan is one.



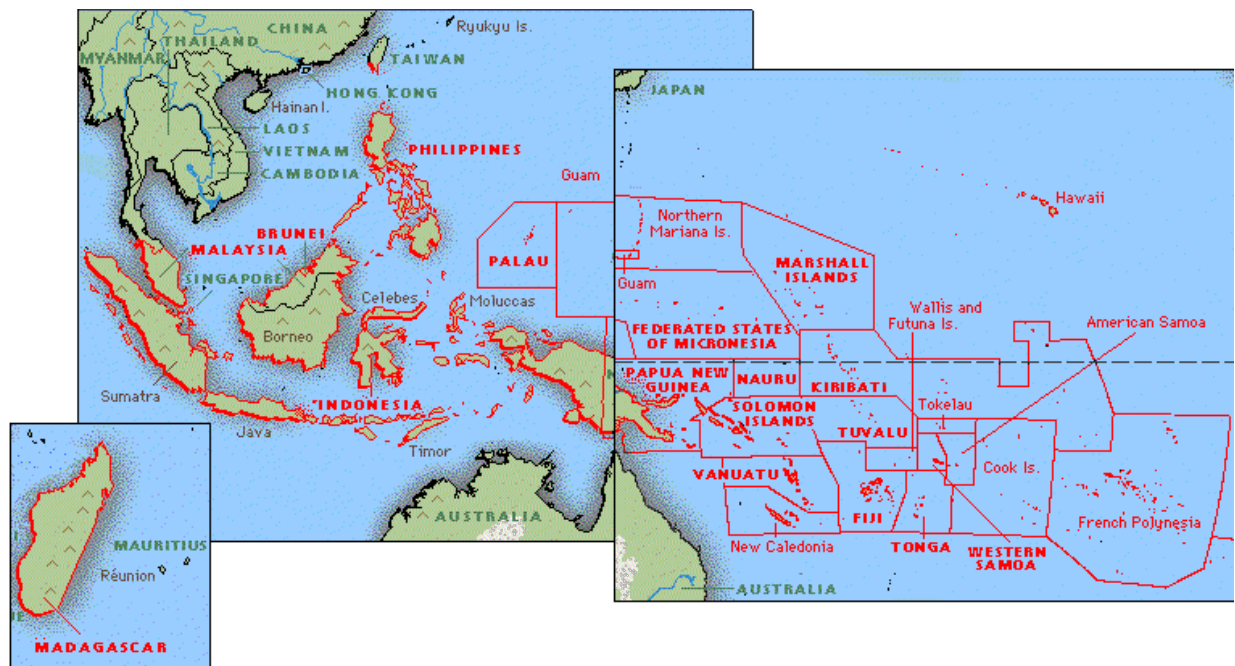
In Tasmania the indigenous language-speaking population was destroyed so quickly that very little is known about the languages there. It is thought that there were between 8 and 12 languages spoken in Tasmania. But so little is known about these languages, that it is not clear how they were related to each other, and there is no evidence that they were related to the mainland languages. Because there is no evidence that they are not related to other Australian languages, they are assumed to be members of the Australian family.

Austronesian

A very large and important language family in the region of Australia is *Austronesian*. These languages all belong to a single family, and are descended from a common ancestor language. Today there are more than 1,200 Austronesian languages spoken, with about 300 million speakers.

Austronesian languages are the indigenous languages of:

- Taiwan (not of the current Sinitic-speaking Chinese majority)
- The Philippines
- A few coastal points on mainland South East Asia (mainly Vietnam)
- Malaysia
- Indonesia
- Many coastal areas of the island of New Guinea
- Much of island Papua New Guinea
- Melanesia (Bougainville, Solomon Islands, Vanuatu, New Caledonia and Fiji)
- Polynesia (including New Zealand, Samoa, Tonga, Hawai'i, Easter Island)
- Micronesia
- Madagascar



Almost all speakers of Austronesian languages speak various languages of the western Austronesian branches - mainly in Indonesia, Malaysia and the Philippines. Many languages of Indonesia and the Philippines have several million speakers - Javanese has about 75 million speakers.

But there are many Austronesian languages with only a few thousand or even a few hundred speakers. The Oceanic branch of Austronesian (Papua New Guinea, Melanesia, Polynesia and most of Micronesia) has hundreds of languages but includes less than 2 million speakers in total. For example Vanuatu has about 105 languages with a total population of about 150,000 speakers.

Papuan

The term Papuan refers to a group of languages that are in the same region and of the same type - but it is made up of at least 60 separate language families. There are more than 750 Papuan languages in the group.

Papuan languages are spoken on:

- the island of New Guinea (comprising the nation of Papua New Guinea, plus the Indonesian province of West Papua);
- parts of some of Papua New Guinea's offshore islands;
- parts of some islands of eastern Indonesia (including in parts of Timor);
- some parts of the Solomon Islands.

About 5 million people speak a Papuan language, but most are spoken by only a few hundred people, or a few thousand at the most. Only three are spoken by more than 100,000 people. The largest Papuan family is *Trans-New-Guinea*,

found across most of mainland New Guinea. It has about 500 languages, including several large languages like Engan and Chimbu. Another large family is *Sepik-Ramu*, with about 100 languages, all spoken by small numbers of people. Another large Papuan family is *Torricelli*, with about 50 languages, also spoken by small numbers of people.

Some other major language families

Uralic

Hungarian is not an Indo-European language, but belongs to the Ugric branch of the Uralic family. Finnish, Estonian, and some smaller European languages belong to the Finnic branch which is also of the Uralic family. Most Uralic languages are spoken further east in Siberia.

Altaic

Turkish is the most westerly member of the Turkic branch of the Altaic family, that also includes other languages of Central Asia like Azerbaijani and Uzbek, and some as far east as western China, such as Uighur. Other branches of Altaic include Mongolian.

Dravidian

The languages of southern India belong to the Dravidian family, including Tamil, Telugu and Malayalam.

Sino-Tibetan

The Sino-Tibetan family has two main branches. *Sinitic* contains the main languages of China, such as Mandarin, Cantonese and Hakka. *Tibeto-Burman* contains languages such as Tibetan, Burmese and Karen.

Other languages of East Asia

Japanese and Korean are not definitely known to be related to each other or any other language, but many people think they are distantly related to Altaic. Ainu is the indigenous pre-Japanese language of the Japanese islands. It is an isolate.

Niger-Congo

This is the largest language family of Africa. Most people in Sub-Saharan Africa speak a Niger-Congo language. The biggest and most important branch is *Bantu*, with languages such as Swahili, Zulu, Rwanda and Xhosa. Other branches include languages like Igbo, Yoruba and Fulani in Nigeria, and Akan in Ghana.

Afro-Asiatic

This family is found throughout North Africa, North-east Africa and the Middle East. One important branch is *Semitic*, which contains Arabic, Hebrew, Maltese

and Aramaic. Other branches include languages such as Berber, Hausa, Somali, Coptic and Ancient Egyptian.

Other families of Africa

The *Nilo-Saharan* family is located in east central Africa and contains languages such as Nubian. The *Khoisan* family is located in south-west Africa and includes the so-called 'click languages'.

South-east Asia

There are quite a few language families in South-east Asia. One is *Austro-Asiatic*, including languages such as Vietnamese and Khmer (Cambodian). Another is *Tai-Kadai*, including languages such as Thai and Laotian.

The Americas

North, Central and South America contain many indigenous language families. Some families in North and Central America include:

Eskimo-Aleut, with languages in the Arctic (including in north-east Asia) such as Greenland Inuit and Yupik.

Algonquin, with languages such as Blackfoot, Cheyenne and Cree.

Na-Dene with languages such as Navajo and Apache.

Hokan-Siouan with languages such as Mohawk and Sioux.

Aztec-Tanoan with languages such as Comanche and Nahuatl.

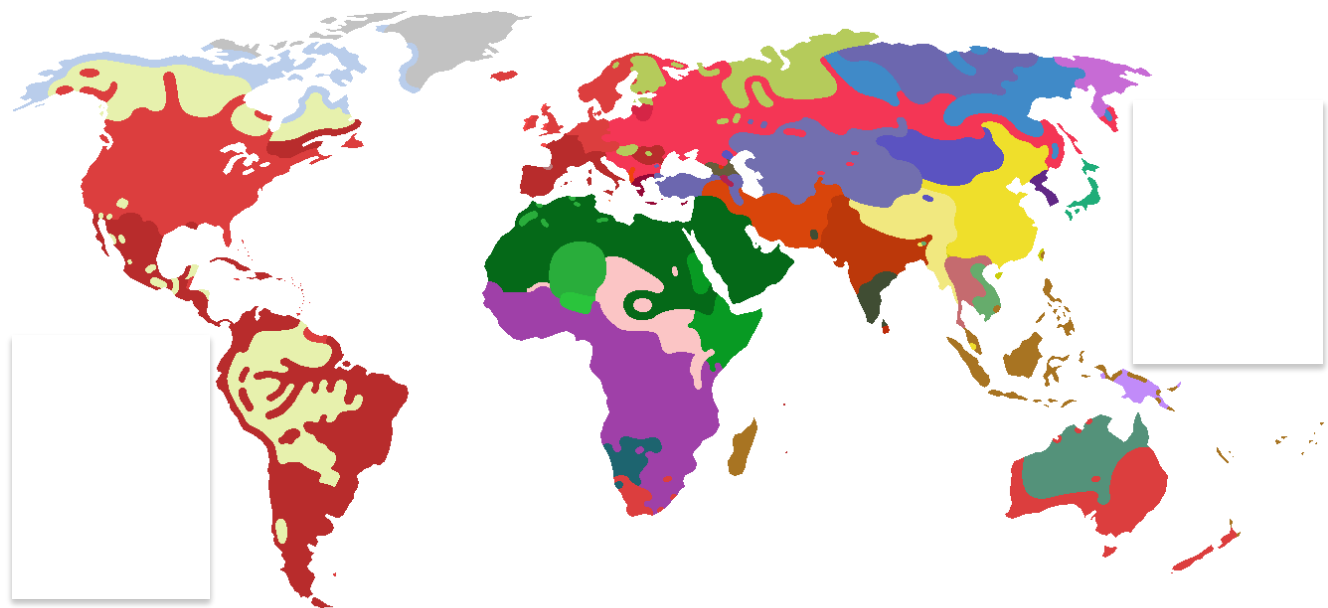
Mayan with languages such as Tzeltal and Quichean.

There are many language families in South America, and not much is known about some of them. Some of the better known families include:

Quechuan, with descendants of the languages of the Incas such as Quechua.

Carib, with the indigenous languages of the Caribbean.

Language Families of the World



- Indo-European
 - Albanian
 - Armenian
 - Baltic
 - Celtic
 - Germanic
 - Greek
 - Indic
 - Iranian
 - Romance
 - Slavic
- Sino-Tibetan
 - Chinese
 - Tibeto-Burman
- Afro-Asiatic
 - Berber
 - Chadic
 - Cushitic
 - Semitic
- Altaic
 - Mongolian
 - Tungusic
 - Turkic
- American Indian (several families)
- Australian (several families)
- Austro-Asiatic
- Austronesian
- Basque
- Caucasian
- Dravidian
- Eskimo-Aleut
- Japanese (possibly Altaic)
- Khoisan
- Korean (possibly Altaic)
- Niger-Congo
- Nilo-Saharan
- Paleo-Siberian (several families)
- Papuan (several families)
- Tai-Kadai
- Uralic

Relationships between languages

Languages that belong to the same language family often have similarities in their grammar, which makes sense because they have inherited a lot of their grammar from their original ancestor language.

But languages that are not of the same family can also resemble each other for other reasons. Languages in the same geographical area can resemble each other even though they may not be members of the same language family. This happens because languages that are in contact with each other can influence each other over time.

For example in some coastal regions of Vietnam, some small Austronesian languages (belonging to the Chamic subgroup of Austronesian) are

interspersed among Vietnamese, which is an unrelated Austro-Asiatic language. Many speakers of Chamic languages are bilingual with Vietnamese. Most Chamic languages have many grammatical characteristics that look more like Vietnamese than their Austronesian relatives elsewhere.

On the island of Bougainville, communities living side by side speak languages belonging to the Austronesian and two separate Papuan families, and many people are bilingual. Many Papuan languages have the word order Subject + Object + Verb (SOV). Most Austronesian languages are either VSO or SVO, but some Austronesian languages have the same order - SOV - as their Papuan neighbours. Some Bantu languages are spoken in areas of southern Africa close to where the Khoisan languages are spoken and they have developed clicks.

Similarities like these are due to mutual borrowing and influence and do not necessarily indicate that the languages are in the same family. This kind of influence of one language on another can lead to *convergence* of languages in a particular area. One example of convergence is the Balkans - Greek, Bulgarian, Albanian and Serbo-Croat are only distantly related, but share many features. Also in central India, where there are two completely separate language families (Indo-European and Dravidian), the languages share many features in common.



ACTIVITIES

Language Families

1. Find out more about one of the indigenous Australian languages that are still spoken today. How many speakers does it have, is the language written, is it taught to children, what written materials are available, are the speakers multi-lingual and in what other languages?
2. Determining if languages are related can be difficult - apparent similarities could be the result of common origins or might just be because of language contact. One way to see if languages are truly related is to look at basic vocabulary such as terms for body parts, close kinship terms, common features of the natural world (e.g. river, sun, cloud) and words for the low numbers. These words are usually resistant to being 'borrowed' from other languages and so similarities in these words would show that languages are related. Consider the following data from five

European languages. Which do you think are related and why? Try to work it out before looking at the answer and explanation below.

Finnish	Hungarian	Basque	Estonian	Spanish	gloss (definition)
kala	hal	arrain	kala	pez	<i>fish</i>
lintu	madár	xori	lind	ave	<i>bird</i>
sarvi	szarv	adar	sarv	cuerno	<i>horn</i>
pea	fej	buru	pää	cabeza	<i>head</i>
silmä	szem	begi	silm	ojo	<i>eye</i>
käsi	kéz	esku	käsi	mano	<i>hand</i>
jalka	láb	oin	jalg	pie	<i>foot</i>
tähti	csillag	izar	täht	estrella	<i>star</i>
pilvi	felhő	hodei	pilv	nube	<i>cloud</i>
puu	fa	zuhaitz	puu	árbol	<i>tree</i>
vihreä	zöld	berde	roheline	verde	<i>green</i>
kaksi	kettő	bi	kaks	dos	<i>two</i>

(Answers on next page)

Answer:

Finnish and Estonian are the most closely related, and both are more distantly related to Hungarian. Basque and Spanish are not related to any of the other languages - the similarity in the words for 'green' is due to a case of borrowing. To support the conclusion that Finnish and Estonian are related we can point to a number of words that are identical in both languages (*kala, puu*), or that vary just in the final sound (*sarvi-sarv, kaksi-kaks*). There is a consistent pattern that word-final *-i* in Finnish has been lost in Estonian, as well as a less-consistent loss of word-final *-u* and *-a*. These two Finno-Ugric languages are

also related to Hungarian, but more distantly. We can see a consistent $p \rightarrow$ sound shift between Finnish and Hungarian, and also word-initial $s \rightarrow sz$. k and l in the Finnish data is often retained in the Hungarian, but can be inconsistent.