

The Indo-European Family of Languages



Indo-European languages

- The **Indo-European languages** include 150 languages spoken by about three billion people, including most of the major language families of Europe and western Asia.

Indo-European languages

- The hypothesis that this was so was first proposed by Sir William Jones, who noticed similarities between four of the oldest languages known in his time, Latin, Greek, and Sanskrit and Persian.
- Systematic comparison of these and other old languages conducted by Franz Bopp supported this theory.

Indo-European languages

- In the 19th century, scholars used to call the group "Indo-Germanic languages".
- However when it became apparent that the connection is relevant to most of Europe's languages, the name was expanded to Indo-European.
- An example of this was the strong similarity discovered between Sanskrit and olden spoken dialects of Lithuanian.

Proto-Indo-European

- The common ancestral (reconstructed) language is called Proto-Indo-European (PIE).
- There is disagreement as to the geographic location where it originated from, with Armenia and the area to the north or west of the Black Sea being prime examples of proposed candidates.

Indo-European family

The various subgroups of the Indo-European family include:

- Indo-Iranian languages
- Italic languages (including Latin and its descendants, the Romance languages)
- Germanic languages
- Celtic languages
- Baltic languages
- Slavic languages
- Illyrian languages (extinct)
- Albanian language (and extinct cousins)
- Anatolian languages (extinct, most notable was the language of the Hittites)
- Tocharian languages (extinct tongues of Tocharians)
- Greek language
- Armenian language

Introduction

The most widely studied language family in the world is the **Indo-European**.

There are a number of reasons for this:

- Many of the most important languages of the world are Indo-European. These languages are official or co-official in many countries and are important in academic, technical and world organisations.
Examples: **English, Spanish, French, German, Russian.**
- Indeed, more than half the world's population speak one or more of these languages either as a mother tongue or as a business language.
- Languages that are essential in multinational contexts or with large numbers of speakers.
Examples: **Portuguese, Hindi, German, Bengali.**
- Some of the great classical languages of religion, culture and philosophy were Indo-European.
Examples: **Latin, Greek, Persian, Sanskrit, Pali.**

-
- Languages that are scattered around the world as their speakers are part of diasporas.

Examples: **Greek, Yiddish, Polish, Armenian, Romany, Kurdish, Italian, Punjabi, Gujarati.**

- The Indo-European languages tend to be inflected (ie verbs and nouns have different endings depending on their part in a sentence). Some languages (e.g. English) have lost many of the inflections during their evolution.

- **The Indo-European Family is divided into twelve branches, ten of which contain existing languages.** I will describe each of these branches separately.

The Celtic Branch

- This is now the smallest branch. The languages originated in Central Europe and once dominated Western Europe (around 400BC). The people migrated across to the British Isles over 2000 years ago. Later, when the Germanic speaking Anglo Saxons arrived, the Celtic speakers were pushed into Wales (**Welsh**), Ireland (**Irish Gaelic**) and Scotland (**Scottish Gaelic**).
- One group of Celts moved back to France. Their language became **Breton** spoken in the Brittany region of France. Breton is closer to Welsh than to French.
- Other Celtic languages have become extinct. These include **Cornish** (Cornwall in England - now being revived), **Gaulish** (France), **Cumbrian** (Cumbria), **Manx** (Isle of Man - another language being revived), **Pictish** (Scotland) and **Galatian** (spoken in Anatolia by the Galatians mentioned in the Christian New Testament).
- Welsh has the word order **Verb-Subject-Object** in a sentence. Irish has the third oldest literature in Europe (after Greek and Latin).

The Germanic Branch

- The **Germanic languages** make one of the branches of the Indo-European (IE) group of tongues, spoken by the Germanic peoples who dwelled north and east along the borders of the Roman Empire.
- These tongues share many markers which they have in common, and which no other tongue has; of these the best known is the sound shift known as Grimm's law.

The Germanic Branch

- These languages originate from **Old Norse** and **Saxon**. Due to the influence of early Christian missionaries, the vast majority of the Celtic and Germanic languages use the Latin Alphabet.
- They include **English**, the second most spoken language in the world, the most widespread, the language of technology, and the language with the largest vocabulary. A useful language to have as your mother tongue.
- **Dutch** and **German** are the closest major languages related to English. An even closer relative is **Frisian**.
- **Flemish** and **Afrikaans** are varieties of Dutch while **Yiddish** is a variety of German. **Yiddish** is written using the Hebrew script.
- Three of the four (mainland) Scandinavian languages belong to this branch: (**Danish**, **Norwegian**, and **Swedish**). Swedish has tones, unusual in European languages. The fourth Scandinavian language, Finnish, belongs to a different family.
- **Icelandic** is the least changed of the Germanic Languages - being close to Old Norse. Another old language is **Faroese**.

- **Gothic** (Central Europe), **Frankish** (France), **Lombardo** (Danube region), **Visigoth** (Iberian Peninsula) and **Vandal** (North Africa) are extinct languages from this branch.
- German has a system of four **cases** and three **genders** for its nouns. Case is the property where a noun takes a different ending depending on its role in a sentence. An example in English would be the forms: **lady, lady's, ladies** and **ladies'**. The genders are masculine, feminine and neuter. German has three dialects spoken in northern Germany, southern Germany and Austria, and a very different form spoken in Switzerland.
- English has lost gender and case. Only a few words form their plurals like German (**ox, oxen** and **child, children**). Most now add an **s**, having been influenced by Norman French.

The Latin Branch

- Also called the **Italic** or **Romance Languages**.
- These languages are all derived from **Latin**. Latin is one of the most important classical languages. Its alphabet (derived from the Greek alphabet) is used by many languages of the world. Latin was long used by the scientific establishment and the Catholic Church as their means of communication.
- **Italian** and **Portuguese** are the closest modern major languages to Latin. **Spanish** has been influenced by Arabic and Basque. **French** has moved farthest from Latin in pronunciation, only its spelling gives a clue to its origins. French has many Germanic and Celtic influences. **Romanian** has picked up Slavic influences because it is a Latin Language surrounded by a sea of Slavic speakers.
- Portuguese and Spanish have been separate for over 1000 years. The most widely spoken of these languages is Spanish. Apart from Spain, it is spoken in most of Latin America (apart from Portuguese speaking Brazil, and a few small countries like Belize and Guyana).

- **Romansh** is a minority language in Switzerland. **Ladino** was the language spoken by Spain's Jewish population when they were expelled in 1492. Most of them now live in Turkey and Israel. **Provincial** and **Catalan** are closely related languages spoken in the south of France and the north-east of Spain, respectively.
- Note that Basque (spoken in parts of Spain and France) is not an Indo-European language - in fact it is totally unrelated to any other language of the world.
- **Galician** is a Portuguese dialect with Celtic influences spoken in the north west of Spain. Finally, **Moldavian** is a dialect of Romanian spoken in the Moldova. Under the Soviets the Moldavians had to use the Cyrillic alphabet. Now they have reverted back to the Latin alphabet.
- Apart from Latin, other extinct languages include **Dalmatian**, **Oscan**, **Faliscan**, **Sabine** and **Umbrian**.
- Latin had three genders and at least six cases for its nouns and a **Subject-Object-Verb** sentence structure. Most modern Romance languages have only two genders, no cases and a **Subject-Verb-Object** structure.

The Slavic Branch

- These languages are confined to Eastern Europe.
- In general, the Catholic peoples use the Latin alphabet while the Orthodox use the Cyrillic alphabet which is derived from the Greek. Indeed some of the languages are very similar differing only in the script used (**Croatian** and **Serbian** are virtually the same language).
- One of the oldest of these languages is **Bulgarian**. The most important is **Russian**. Others include **Polish**, **Kashubian** (spoken in parts of Poland), **Sorbian** (spoken in parts of eastern Germany), **Czech**, **Slovak**, **Slovene**, **Macedonian**, **Bosnian**, **Ukrainian** and **Byelorussian**.
- The Slavic languages are famed for their consonant clusters and large number of cases for nouns (up to seven). Many of the languages have three **numbers** for verbs: **singular**, **dual** and **plural**. Macedonian has three definite articles indicating distance; all are suffixes: **VOL (ox)**, **VOLOT (the ox)**, **VOLOV (the ox here)**, **VOLON (the ox there)**.

The Baltic Branch

- Three Baltic states but only two Baltic Languages (Estonian is related to Finnish).
- **Lithuanian** is one of the oldest of the Indo-European languages. Its study is important in determining the origins and evolution of the family. Lithuanian and **Latvian** both use the Latin script and have tones. Lithuanian has three numbers: singular, dual and plural.
- **Prussian** is an extinct language from this branch

The Hellenic Branch

- The only extant language in this branch is **Modern Greek**.
- Greek is one of the oldest Indo-European languages. **Mycenaean** dates from 1300BC. The Ancient Greek of **Homer** was written from around 700BC. The major forms were **Doric** (Sparta), **Ionic** (Cos), **Aeolic** (Lesbos), and **Attic** (Athens). The latter is Classical Greek.
- The New Testament of the Christian Bible was written in a form of 1st Century AD Greek called **Koine**. This developed into the Greek of the **Byzantine** Empire. Modern Greek has developed from this.

- Greek has three genders and four cases for nouns but no form of the verb infinitive. The language has its own script, derived from Phoenician with the addition of symbols for vowels. It is one of the oldest alphabets in the world and has led to the Latin and Cyrillic alphabets. The Greek Alphabet is still used in science and mathematics.
- Until the 1970s Greek was a **Diglossic** language. This means that there were two forms: **Katharevousa** used in official documents and news broadcasts and **Demotic** used in common speech.
- The Greek spoken in Cyprus includes many Turkish, Arabic and Venetian words and has a different pronunciation to the official Greek of Greece.

The Illyric Branch

- Another single language branch. Only **Albanian** (called **Shqip** by its speakers) belongs to this branch. It has been written in the Latin script since 1909; this replaced a number of writing systems including Greek and Arabic scripts. Albanian has many avoidance words. Instead of saying **wolf**, the phrase **may God close its mouth** is used. The definite article is shown by a suffix: **BUK** (**bread**) **BUKA** (**the bread**). Many noun plurals are irregular.
- There are two dialects that have been diverging for 1000 years. They are mostly mutually intelligible. **Geg** is spoken in the north of Albania and Kosovo (Kosova). **Tosk** is spoken in southern Albania and north west Greece.
- The ancient **Illyric** and **Mesapian** languages, spoken in parts of Italy, are considered by some to be an extinct member of this branch.

The Anatolian Branch

- This branch includes the language of the **Hittite** civilisation which once ruled central Anatolia, fought the Ancient Egyptians and was mentioned in the Christian Bible's Old Testament. Other languages were **Lydian** (spoken by a people who ruled the south coast of Anatolia), **Lycian** (spoken by a Hellenic culture along the western coastal regions), **Luwian** (spoken in ancient Troy) and **Palaic**.
- All languages in this branch are extinct.

The Thracian Branch

- This branch is represented by a single modern language, **Armenian**. It has its own script.
- Armenian is spoken in Armenia and Nagorno-Karabakh (an enclave in Azerbaijan). The language is rich in consonants and has borrowed much of its vocabulary from Farsi (Iranian). Nouns have 7 cases and the past tense of verbs take an E prefix like Greek.
- Three extinct languages from this branch are **Dacian** (or **Daco-Mysian** - spoken in the ancient Balkan region of Dacia), **Thracian** and **Phrygian** (spoken in ancient Troy).

ARMENIAN

Եզական զէմք մըն է հասել Մովսէս Խորենացի, որուն հանդէպ երախտիքի անասկման պարտք մը ունին իր ազգակիցները: Նպատաք զար ան բերած է մեր գրականութեան անգին է: Ան մեր առաջին պատմագիրն է: Այդ պարճառով ան կոչուած է «պատմահայր», մեր պատմաբաններուն հայրը: Ան շանացած է գրել հայոց պարմութիւնը, մեր ցեղի ծագումէն մինչեւ Տրդ դար, իր օրերը:

The Iranian Branch

- These languages are descended from **Ancient Persian**, the literary language of the Persian Empire and one of the great classical languages.
- The main language of this branch is **Farsi** (also called **Iranian**, **Dari** and **Persian**), the main language of Iran and much of Afghanistan. **Kurdish** is a close relation. Kurdish is spoken in Turkey, Syria, Iran and Iraq by the Kurds. It is the second largest of the Iranian languages after Farsi. In Turkey it was banned until recently.
- **Pashto** (also called **Pushtu** or **Pakhto**) is spoken in Afghanistan and parts of north west Pakistan. **Baluchi** is spoken in the desert regions between Iran, Afghanistan and Pakistan. These languages are written in the Nastaliq script, a derivative of Arabic writing. It is interesting that you cannot tell which family a language belongs to by the way it is written.
- **Ossetian** is found in the Caucasus mountains, north of Georgia. **Tadzhik** is a close relative of Farsi, written in Cyrillic and spoken in Tadzhikistan (of the former USSR) as well as northern Afghanistan.
- **Avestan** is the extinct language of the Zoroastrian religion. **Scythian** is an extinct language of a warrior people who once lived north of the Black Sea.

The Indic Branch

- This branch has the most languages. Most are found in North India. They are derived from **Sanskrit** (the classical language of Hinduism dating from 1000BC). This gave rise to **Pali** (the language of Buddhism), **Ardhamagadhi** (the language of Jainism) and the ancestors of the modern North Indian languages.
- Of the modern North Indian languages, **Hindi** and **Urdu** are very similar but differ in the script. The Hindi speakers are Hindus and use the Sanskrit writing system called Devanagari (writing of the Gods). Urdu is spoken by the Muslims so uses the Arabic Nastaliq script. These two languages are found in north and central India and Pakistan. **Nepali** is closely related to Hindi.

Hindi

सिन्धु सभ्यता भारत की प्राचीनतम सभ्यता है। इस सभ्यता का विकास आज से लगभग पाँच हजार वर्ष पहले हमारे देश में हुआ था। यह सभ्यता पंजाब के मोंटगुमरी जिले के हड़प्पा, और सिन्धु प्रांत के लरकाना जिले के मोहनजोदड़ो में फैली हुई थी। मोहनजोदड़ो कराची से २०० मील उर में सिन्धु नदी के तट पर है।

- In India most of the states have their own language. These languages either use Devanagari script or a derivation (if the people are Hindus) or the Arabic Nastaliq script (if the people are Muslims).
- **Bengali** (West Bengal as well as Bangladesh), **Bhili** (Central India), **Oriya** (in Orissa), **Marathi** (in Maharashtra), **Assamese** (in Assam), **Punjabi** and **Lahnda** (from the Punjab), **Maithili** and **Maghadi** (from Bihar), **Kashmiri** (Kashmir - written mainly in Nastaliq), **Sindhi** (the Pakistan province of Sindh - also written in Nastaliq), **Gujarati** (Gujarat in western India), **Konkani** (in Goa, an ex Portuguese colony, uses the Latin script), **Sinhalese** (Sri Lanka - uses its own script derived from Pali), **Maldivian** (Maldives - with its own script based on Arabic).

Bengali

ভাষাধিকারী গুরুত্বপূর্ণ সূত্রগুলি

Punjabi ਜੈ ਘਰਿ ਕੀਰਤਿ ਆਖੀਐ ਕਰਤੇ ਕਾ ਹੋਇ ਬੀਚਾਰੋ॥ ਤਿਤੁ ਘਰਿ ਗਾਵਹੁ ਸੋਹਿਲਾ ਸਿਵਰਿਹੁ ਸਿਰਜਨਹਾਰੋ॥ 1॥ ਤੁਮ ਗਾਵਹੁ ਮੇਰੇ ਨਿਰਭਉ ਕਾ ਸੋਹਿਲਾ॥ ਹਉਵਾਰੀ ਜਿਤੁ ਸੋਹਿਲੇ ਸਦਾ ਸੁਖੁ ਹੋਇ॥ 1॥

Oriya

ଦକ୍ଷିଣ ଦେଶରେ ସିନ୍ଧୁ ନାମକ ରାଜ୍ୟ ଥିଲା । ସେଠାରେ ବୀରବାହୁ ବୋଲି ଜଣେ ରାଜା ଥିଲେ । ତାଙ୍କର ପୁଅଟି ରାଣୀ ଥାନ୍ତି । ବଡ଼ରାଣୀର ନାମ ପ୍ରେମଶୀଳା, ସାନ ରାଣୀର ନାମ କନକମଞ୍ଜରୀ । ସାନ ରାଣୀଟି ବଡ଼ ସୁନ୍ଦରୀ । ତାଠାରେ ରାଜା ବଡ଼ ସ୍ନେହ କରୁଥାନ୍ତି । ବଡ଼ ରାଣୀଟିକୁ ଦେଖି ପାରନ୍ତି ନାହିଁ । କେତେ ଦିନ ଗଲାପରେ ଦୈବଯୋଗେ ବଡ଼ ରାଣୀର ଗୋଟିଏ ପୁଅ ଜନ୍ମ ହେଲା । ବଡ଼ ରାଣୀର ପୁଅ ହେବା ଦେଖି ସାନରାଣୀ ମନେ ମନେ ଚିନ୍ତାକରି ବିଗୁଳା, ତାର ତ ପୁଅ ହେଲାଣି ସେ ଗନ୍ଧ୍ୟ ପାଇବ, ଭଲ ମନ୍ଦ ହେଲେ ରାଜା ମୋତେ ନିଶ୍ଚେ ରାଜ୍ୟରୁ ଚଢ଼େଦେବେ । ମୁଁ ଏବେ କି ଭୟାୟ କରିବି?

Sinhalese

මල් රජකින් බୋහෝ මල් දම් ගୋනන්නේ

- The most surprising language in this branch is **Romany**, the language of the Roma (also known as Gypsies - this is a derogatory term which should not be used). The Roma migrated to Europe from India.
- Sanskrit had three genders as has Marathi; most modern Indic languages have two genders; Bengali has none.
- The fascinating point about India is that the south Indian languages (like Tamil) are not Indo-European. In other words, Hindi is related to English, Greek and French but is totally unrelated to Tamil. North Indians visiting Madras (in the south) are as baffled by Tamil as a foreigner would be.

The Tokharian Branch

- **Turfanian** and **Kuchean** are recently identified extinct languages once spoken in north west China. Very little is known about this branch as only a few manuscripts dating from 600 AD are in existence. The languages disappeared around the 8th century AD. The closest relatives of these languages are from the Celtic, Anatolian and Latin branches.

Celtic Branch

Welsh : Irish Gaelic : Scottish Gaelic : Breton
Cornish : Gaulish : Cumbrian : Manx : Galatian

Germanic Branch

English : Dutch : Flemish : Frisian : Afrikaans
German : Yiddish : Danish : Swedish :
Norwegian
Faroes : Icelandic
Anglo Saxon : Old Norse : Frankish : Gothic
Lombardo : Visigoth : Vandal

Romance (Latin) Branch

Italian : Sardinian : French : Provençal : Catalanian
Spanish : Ladino : Galician : Portuguese : Romansh
Romanian : Moldavian
Latin : Oscan : Umbrian : Faliscan : Sabine : Dalmatian

Slavic Branch

Russian : Belorussian : Ukrainian : Polish : Sorbian
Czech : Slovak : Slovene : Croatian : Serbian
Kashubian : Bulgarian : Macedonian : Bosnian
Old Church Slavic

Baltic Branch

Lithuanian : Latvian
Prussian

Hellenic Branch

Modern Greek
Mycenaean : Koine : Byzantine Greek
Classical Greek (Attic : Doric, Ionic, Aeolic)

Illyric Branch

Albanian
Illyric : Mesapian

Anatolian Branch

Hittite : Lydian : Lycian : Luwian : Palaic

Thracian Branch

Armenian
Dacian : Thracian : Phrygian

Iranian Branch
Farsi : Kurdish : Pashto : Baluchi : Ossetian : Tadjik Persian : Avestan : Scythian
Indic Branch
Hindi : Urdu : Nepali : Bengali : Assamese : Oriya Kashmiri : Punjabi : Sindhi : Marathi : Gujarati Bhili : Lahnda : Maithili : Magahi Konkani : Sinhalese : Maldivian : Romany Sanskrit : Pali : Ardhamagadhi
Tokharian Branch
Turfanian : Kuchean

Extinct languages are in lighter type.

The end!

