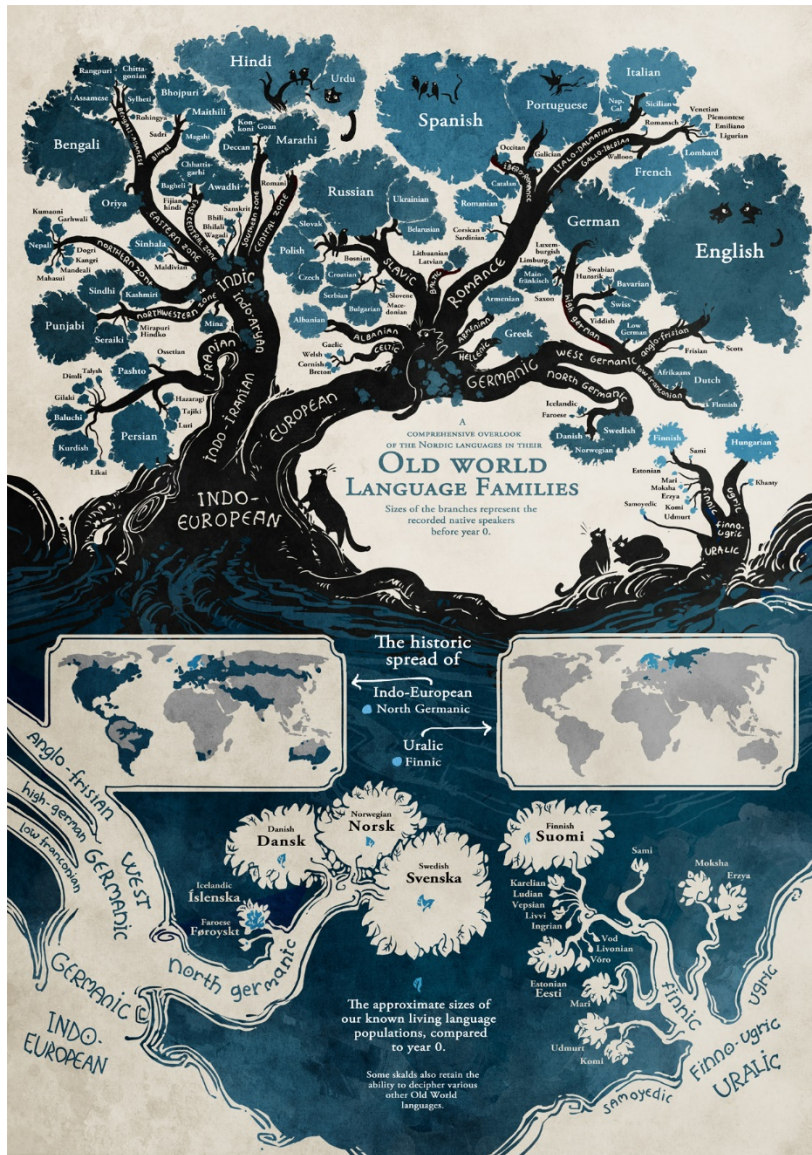# Genetic classification of languages (development of languages: tree model and wave model)

Genetic classification of languages groups together languages which developed and evolved throughout history together. Languages which descended from a common ancestor are grouped into one language family.

These families can be represented by a **tree model.**

Some features can't be explained by a tree model though. The tree model can explain why languages with a common ancestor share a feature. But what about features shared by languages with no common ancestor? These are usually caused by some areal sharing and **wave model** tries to capture that.
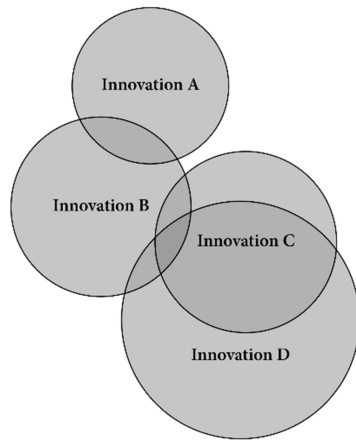
What are protolanguages?

Protolanguages are postulated lost parent languages, from which other languages have developed. A language family is formed from the daughter languages of a protolanguage (the mother language). If some or all of the daughter languages of a protolanguage are attested (meaning we have evidence of the language surviving to this day), we can attempt to partially reconstruct the protolanguage itself.
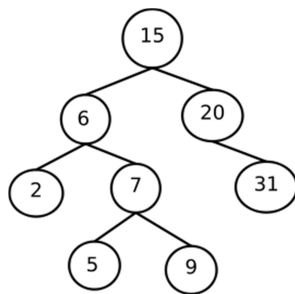
In the graph above (which uses the 'tree model' of historical linguistics), the protolanguages are 15, 6, 20, 7.

Examples which are universally accepted as having existed are

- Proto-Indo-European - the common ancestor of the Indo-European language family (of western and southern Eurasia) and the most researched protolanguage, and suggested to have been spoken between 4,500 BC and 2,500 BC
- Proto-Uralic - the common ancestor of the Uralic language family (mostly in northern Eurasia) which is suggested to have been spoken between 7,000 BC and 2,000 BC
- Proto-Dravidian - the common ancestor of the Dravidian languages (mostly in southern India and north-western Sri Lanka) which is suggested to have been spoken approximately between 4,000 BC and 3,000 BC

Protolanguages can be partially attested like which means that we have some evidence of what the language looks like. Vulgar Latin (the colloquial Latin of the Late Roman Republic) is the protolanguage

from which the Romance languages (Spanish, Italian, French, etc.) developed. In this case the language is attested in surviving texts, and Ecclesiastical Latin is still the liturgical language of the Catholic Church.

Otherwise, we can have completely unattested protolanguages like Proto-Indo-European for which we have no written evidence. In this case the only choice is to attempt to reconstruct the language using the daughter language(s) of this parent protolanguage for which we do have written evidence. The most common and successful method is the comparative method, which was developed in the study of Proto-Indo-European).

An important question is how far can we go in the reconstruction of a protolanguage? Of course, for unattested protolanguages, we cannot ever know if we are 'correct', thus there are no objective criteria we can use to evaluate the reconstructions of a protolanguage. This process of reconstruction has been described as an 'intuitive undertaking' - essentially we are making an educated, but error-prone, guess.

Here is an example which suggests that Italian and English came from the same language

- In Italian: *padre, piede, pesce*
- In English: *father, food, fish*

By analyzing examples like these we can begin to construct the phonetic and phonological changes the took place between the ancestral protolanguage and its daughter languages, and from there can work backwards to construct the words of the protolanguage from the vocabulary of its daughter languages.

Reconstruction of a protolanguage

Linguistic reconstruction is the practice of establishing the features of an unattested ancestor language (protolanguage) of one or more given languages.

The most used reconstruction is the comparative reconstruction which establishes features of the ancestor of two or more related languages, belonging to the same language family, by means of the comparative method.

Method

Firstly, it is needed to group together languages that are thought to have arisen from a common protolanguage. They must meet certain criteria in order to be grouped together. This is a process called subgrouping. Since this grouping is based purely on linguistics, manuscripts and other historical documentation should be analyzed to accomplish this step.

One of the criteria is that the grouped languages usually exemplify shared innovation, which means that the languages must show common changes made throughout history.

In addition, most grouped languages have shared retention. That means that there are features that stayed the same in all of the languages in the group.

Then the comparative method is applied to a group of languages featuring similar characteristics to reconstruct protolanguage for this group.

Principles of the comparative method:

- every reconstruction works with changes which are probable, systematic
- as few changes between the protolanguage and the descendants as possible
- reconstructions should fill in the gaps in the systems rather than create unbalanced systems:

Present-day Slavonic data:

Cz.   **h**lava (head)   what is original, g or h?

Pol.   **gł**owa

Rus.   **g**olova

Upper Sorb.   **h**łowa

Answer: build a system: Reconstruction of Old Slavonic:

	Voiceless	voiced
Labial	p	b
Dental	t	d
Velar	k	**? (g!))**

Other approaches

There were other approaches used for reconstruction than comparative method, because some linguists thought that comparative method can only reach back a few thousand years before the evidence fades out, so something else must be tried.

One example of this is multilateral comparison, which is looking at many languages across a few words instead of looking at many words across a few languages.

But all of the other approaches turned out to be flawed and afford no new reliable insights, so they are no longer used.

Comparative Method of the Reconstruction of a Protolanguage

- this method compares words and expressions in different languages or dialects that derived from the same ancestor language
- the comparison may be done on different levels of language (phonological, semantic, ...)
- languages reconstructed using this method are called proto-languages (e.g. Proto-Indo-European, Proto-Uralic)

Internal Reconstruction of Former Phases of a Language

- as opposed to the comparative method the internal reconstruction studies only a single language. The assumptions about the earlier versions of the language are made based on irregularities within the language.
- languages reconstructed by this method have prefix "pre" (e.g. Pre-Old Japanese)

Example of internal reconstruction:

new Czech data: *dům – domek* a house – a small house

Question: how can the difference between the vowels *ů – o* be explained (vowel quantity, quality)?

Answer: internal reconstruction:

old Czech: reconstruction of a vowel system:

short long

a   á

e   é

o   ?

u   ú

ě   ie

y   ý

i   í

ó changed to ů [u:], originally, the vowels differed just in the quantity: *dóm – domek*

Glottochronology

Glottochronology -- linguistic method to estimate a separation date between two genetically related languages. It studies the rate of change occurring in the vocabularies of languages for the purpose of calculating the length of time during which two related languages have developed independently. It rests upon statistical comparison and it is based on the assumption that the rate of vocabulary replacement is constant over sufficiently long time periods.

Often criticized, uncertain.

Kinds of linguistics changes and examples

Many different kinds of changes can occur within a language. This is a general classification of some of those linguistic changes with specific examples.

Sound changes: there are phonemic and non-phonemic changes, and among the first ones there are mergers and splits. Here are two examples of mergers in Spanish (two distinct phonemes merging into one): ʎ, j → j in most of Latin America and a big part of Spain (making *halla* and *haya* sound the same) and θ, s → s in Latin America and some dialects in Spain (making *caza* and *casa* sound the same).

Analogical changes: a change happens by analogy of how something else works in the language. For example, in English, the pattern of the verb *speak/spoke/spoken* developed by analogy with verbs of the pattern *break/broke/broken*.

Semantic and lexical changes: for example, by widening (the range of meanings of a word increases). *Salary* in English comes from *salārium* in Latin, which was a soldier's allotment of salt (based on Latin *sal* "salt"). The word *dog* originally referred to a specific powerful breed of dog.

Syntactic change: for example, by borrowing: Pipil (a language in El Salvador) borrowed the comparative expression *mas … ke* from Spanish (*más … que,* "more … than").

Cognate word forms

Cognate words are pairs of words in (usually) different languages that share a common origin.

Firstly, they are not be confused with doublets, which are words in the same language which have a common origin. While technically speaking, doublets are cognates, cognates don't have to be in the same language, i.e. doublets are a subset of cognates.

Secondly, they are not to be confused with borrowed words, as these words are simply stolen from other languages to fill in lexicon gaps. Whereas cognate words have evolved and diverged from a common origin.

Some examples of cognate words are:

brother (English) vs Bruder (German)

dish (English) vs Tisch (German; table)

noche (Spanish) vs Nacht (German)

fragile (English) vs frail (English).

Language family

The language family is a group of many languages sharing the same parental language, so called proto-language. We can imagine language family as a tree structure with the specific proto-language as a root. Then all edges there determine relations of which language has descended from which other language by evolution.

The Ethnologue publication says there are 142 different language families. On the contrary, the WALS says there are 258 of them.

Language families; examples

Both English and Czech belong to the Indo-European language family, which has spread throughout the entire world from its native Europe and parts of the Middle East and the Indian subcontinent. Wikipedia states that 46% of the world's population speak an Indo-European language as a first language, making it by far the most common language family. According to WALS, this family contains 176 languages divided into 11 genera. The largest genus in terms of the number of languages is the Indic genus with 50 languages (such as Hindi), followed by 39 Germanic languages. One of them, Afrikaans, is the only Indo-European language native to Africa, having originated from Dutch in the southern part of the continent. Other prominent genera are the Iranian, Romance and Slavic languages. But there are also some very small genera, and the Albanian language even forms a genus by itself.

Over half a billion people living in some parts of Africa and the Middle East speak a language from the Afro-Asiatic family, making it the fourth-largest family by the number of native speakers. WALS divides its 145 languages into 16 genera. The majority of speakers of this family speak a Semitic language. This genus contains not only Hebrew and all regional varieties of Arabic, but also Maltese, the only Afro-Asiatic official language of the European Union.

A considerably smaller family is that of Uralic languages, spoken by around 25 million people, predominantly in Northern Eurasia. The family consists of 27 languages divided into 7 genera. Hungarian from the Ugric genus and Finnish and Estonian from the Finnic genus are the remaining non-Indo-European official languages of the European Union.

Cognate languages. Language isolate.

Languages in a language family tree structure are in relationships that resemble the kinship relations in a family tree. Therefore, languages have ancestors and children. Languages with the common ancestral language are called cognate languages or more commonly sister languages. If more than one language has developed from the same proto-language, those languages are said to be sister languages, members of the same language family. For example, Portuguese and Spanish are sister languages, descending from the same ancestor - Vulgar Latin.

Another concept that is important to note is the concept of language isolates. Those are the languages that cannot be classified into larger language families. Language isolates do not demonstrate any genealogical relationships with other languages, that is - they do not appear to have common ancestors with any other language. The popular examples of such languages are Korean and Basque, which is the only language isolate in Europe. Japanese has been considered to be a language isolate for some time, but now it is actually considered to be a part of the Japonic language family. Therefore, it is important to

note that the languages may be reclassified as the linguistic data grows or as the linguistic theories change over time.

Unclassified languages also do not demonstrate any genealogical relationships with other languages, but the reason for this is usually the absence or scarcity of (reliable) linguistic data. Usually, these are some extinct languages, i.e. Gutian (spoken by the Gutian people who ruled Sumer in the 22nd century BC) and Cacán (spoken by the Diaguita and Calchaquí tribes in northern Argentina and Chile). However, not all extinct languages are unclassified - Sumerian is considered to be a language isolate because it is well documented.

Linguistic area (Sprachbund)

The term linguistic area refers to a geographical area in which, due to borrowing and language contact, languages of a region share certain structural features - not only borrowed words, but also shared elements of phonological, morphological or syntactic structure. Linguistic areas are also referred to by the terms Sprachbund, diffusion area and convergence area.

Examples of well-known linguistic areas

The Balkans

- Greek, Albanian, Serbo-Croatian, Bulgarian, Macedonian and Romanian
- some features : a central vowel /1/; syncretism of dative and genitive; absence of infinitives

The Baltic

- Lithuanian, Latvian, Finnish, Estonian, Livonian, High and Low German, Danish, Swedish, Norwegian, Russian, Belorussian, Ukrainian, Polish
- some features : first syllable stress; palatalization of consonants; tonal contrasts; SVO (subject-verb-object) word order; prepositional verbs

Other examples of linguistic areas:

Mesoamerica

- Nahua (brach of Uto-Aztecan), Mayan, Cuitlatec, Hauve, Totonacan
- some features : non-verb-final word order; vigesimal numeral systems based on combination of twenty (60 = 3*20 = uʃ-k'al)

South Asia

- Indo-Aryan, Dravidian, Munda and Tibeto-Burman families
- some features : retroflex consonants; absence of prefixes; SOV word order; absence of a verb to have

Which linguistic features can spread from one language into another one?

All kinds of linguistic features can be spread from one language into another one. It can affect the pronunciation, the vocabulary, or the sentence structure and the grammar of the languages concerned. One of the most common changes are phonological changes which cause the phonological systems to

gradually become more and more similar. One example is the change of a closed vowel to an opened one or vice versa. Another example is the loss of contrast between two categories of consonants i.e., velar and post-velar/uvular plosives become indistinguishable.

Lexicon and grammar changes include the imitation in the choice of case endings, adding reflective pronouns, copying a whole set of prefixes, or changing the way of forming certain grammatical phenomena such as plurals or numeral formation. Borrowing whole words is also common. The sentence structure changes mainly concern the word order (SVO, SOV...) or the formation of interrogative sentences.

Is your mother tongue influenced by other contemporary languages? If so, then in which way?

My mother tongue is Slovak and it is mostly influenced by Czech and English. The most common phenomenon nowadays is borrowing words from other languages, mainly English. Sometimes the spelling is changed to make the borrowed word look more Slovak (jam – džem, meeting – míting). Slovak is also influenced by Czech a lot, resulting in the usage of some unchanged Czech words.

Pidgin and creol languages

All languages are derived from some languages from the past. However, there is a category of languages that have evolved under special circumstances - they are known as pidgin and creol languages.
When speakers of different languages come into a situation where they have to communicate with each other and there is no possibility to choose some common language, they have to develop their own language - a pidgin language.
Vocabulary of the pidgin language comes mostly from dominant language, but the vocabulary is reduced in size. Grammar of the pidgin language is radically different from the grammar of dominant language and typicaly involves more regularity and less redundancy. A pidgin language is used only as second language by all of its speakers.

A creol language has some things in common with pidgin, but in other ways creol is quite different. Like pidgins, creols develop in situations where speakers of different languages are in contact and they use vocabulary mostly from the dominant language. However, creols are not only used as second language but also as primary language with full communication function, therefore they are more developed, their lexicons are richer,.....

Example of pidgin language can be found between sailors, in the marine,..
When Malasians were taken by English-speaking Europeans for example from Papua New Guinea to work on sugarcane plantages in 19th century, new pidgin language developed quickly. Later on, it became a creol language.