An Essay on the Origin and Classification of Languages

Fernando Magno Quintão Pereira

May 27, 2005

Abstract

This essay discuss the classification of languages. Some of the mechanisms that allow linguists to group languages into families are introduced. The paper presents two paradigms used for language classification: the genetic model, and the typologic model. In order to illustrate the genetic paradigm, several families of languages are described. This document also tries to emphasize the importance of the classification of languages by describing some applications that follows from such classification.

1 Introduction

There are almost 7,000 languages currently been used in the world [9]. Although the existing languages present great grammatical, phonological and morphological differences, linguists believe that it is possible to determine a common origin to some groups of languages. For instance, the so called Romance languages, whose best known examples are French, Italian, Portuguese, Rumanian and Spanish, share the same mother tongue: the vulgar Latin, which was spoken by merchants and soldiers inside the borders of the Roman empire.

Languages are dynamic entities that evolve along the time. The most recent is a language, more are the similarities between it and its ancestor and sibling languages. It is remarkable, for instance, the correlation between the Romance languages and Latin. Not so cogent, however, is the similitude between Latin and Sanskrit, the mother tongue of most of the actual languages spoken in India. Even though, linguists believe that both languages share a common ancestor; because of it, Latin and Sanskrit are classified as members of the same family of languages.

The classification of languages is important for several reasons. It helps to understand how languages evolve, it allows to reconstruct dead languages, and it permits to trace the ancestors of the modern languages. But, despite such importance, there are some discordance about how languages should be related. Even though a large number of linguists has devoted time and effort in the classification of languages, there are some open questions in this area. For instance, which factors lead to the differentiation of languages? Other interesting questions are: is there a common ancestor of all the modern languages? For how long have human societies used language as we know it presently, to communicate? And, also, how many families of languages can be catalogued?

This document address these questions, and explain some methods used by linguists in order to classify languages. The paper contains six more sections. Section 2 discusses some models related to the origin and evolution of languages. The objective of this section is to present some attempts to explain the plethora of languages on Earth. Section 3 describes some techniques used by linguists in order to classify languages. In particular, it gives some details about genetic and typological classification. Examples of families of languages are presented in Section 4. Section 5 put forward some applications that follows from the classification of languages. Finally, Section 6 concludes the paper.

2 Origin and Diversification of Languages

An essential question that challenges all the linguists is what make it possible for human beings to start communicating by means of words. Communication, in this case, is the ability to express ideas in a generative way. One of the factors that distinguishes the human languages from the languages used by other animals is the apparently unlimited number of ideas, or propositions, that a person can convey by means of words. There are several different theories that aims to explain the origin of languages. Two of the most acceptable models are the vocalization, and the gestural origin.

Vocalization is the ability of some animals to emit signals by means of sounds produced by the flow of air across their throats. Such capacity has being observed in several different families of mammals, and, most notably in primates. Some social monkeys can alert their whole community about the presence of predators, or the discovery of food, by means of different types of cries and shouts. There are even registers of monkeys that were able to lie by means of such signs: after finding food, the primate issued the alarm call, in order to avoid sharing its discovery with the other members of the community. Some scientists believe that these signals were the ultimate ancestor of the modern human languages.

Although vocalization is the most accepted model, there are scientists who are reluctant to adopt it as the true explanation for the origin of languages. One of the main arguments contrary to vocalization is the fact that animals use a very limited set of signals, that exist by themselves, and cannot be combined in order to produce more complex ideas. Thus, the vocalization ability lacks the generative component of human languages. According to some linguists, the primitive vocal calls persist in the modern human being as emotional responses, such as cries, laugh and scream, but not as speech [4].

Another theoretic model that aims to explain the origin of languages is based on gestural expression. Humans tends to express by means of gestures, and there are fully functional languages based on signs. This model was proposed in the eighteenth century. Research on this area has studied human sign-language, and the capacity of great apes to communicate by means of gestures. The defendants of the gestural origin of languages say that regular tool-using in hominids is likely to have evolved before speech capacities. In addition, they argue that, gestures suit better the necessities posed by the wild environment than sounds.

2.1 The Evolution of Languages

Languages are dynamic entities. They evolve by incorporating new words, concepts and sounds. Even their syntax is subject of modifications given the proper amount of time. Although every linguist agree that languages are in constant process of evolution, how this process happens is yet matter of discussion. Essentially, languages evolve due to two basic mechanisms. The first, is the contact between different groups of people. The second, in contrast, is due to the separation between human communities.

Normally, languages evolve faster in primitive societies. For instance, the greatest density of languages on Earth is found in the jungles of Papua-New Guinea, and South America. In these region, human societies are mostly constituted by small settlements. In a society with not more than 500 individuals, intermarriage with neighboring groups is common, and this phenomenon is source of some linguistic diversity. For instance, members of different tribes are likely to use different terms in daily communication. The contact between the tribes makes it possible the sharing of such expressions, and contributes to change the original tongue adopted in each society. There are even cases of small societies that are multilingual [16].

The separation of human communities is another mechanism that makes possible the evolution of languages, according to a process called branching. Two groups that are exposed to different contexts tend to adopt different constructs in word to communicate. In this sense, migrations are major source of language modification, because they isolate groups of people, and expose organized societies to new environments. For instance, anthropologists believe that the American continent has been colonized by successive migrations from Asia. If this is true, even though the Amerindians and the societies of East Asia share a common, and relatively recent, origin, they speak completely different languages, which is probably due to a long period of separation.

3 Mechanisms of Language Classification

Languages can be classified according to different criteria. Three of the most used classification systems are: genetic classification (Section 3.3), typologic classification (Section 3.2) and areal classification. The first system is based on the hypothesis that languages derive from common ancestors. The methods used in this system are mostly archeological: written remains, genetic markers, cognate words, etc. The second system is based on morphological similarities between languages. Even if two languages do not share a common ancestor, they can present similar characteristics, such as phonemes or syntax. Finally, the areal classification is based on geographic criteria. In this case, languages are grouped together according to the area in which they were first observed.

Sections 3.2 discusses the typologic classification of languages, and Section 3.3 introduces genetic classification.

3.1 Universal versus Specific

There are characteristics of languages that are universal, that is, can be observed in all of them. Five universal properties of languages are listed below:

- *The generative aspect*: in principle, sentences containing an infinite number of ideas can be built in any language. This is due to the recursive property present in every language. This generative aspect is one of the main blocks of the syntactical system proposed by Noan Chomsky.
- *Independence between form and meaning*: in any language, the relation between most of the words and the concepts that they represent is arbitrary. Apparently there is no reason that justify the form of words such as *love* and *light*.
- *Linguistic competence*: there is no immediate relation between what a person says, and the reality her perceives. Lies, metaphors, and any other unpredictable figures are common to any human communication system.
- *Flexibility*: any language is flexible enough to deal with new concepts and unforeseen situations.
- *Dynamic aspect*: languages are not statical entities. Every language can absorb new words and sounds as the result of contact between different peoples, or as the result of progresses and discoveries. For instance, words such as radio and stem cell have being incorporated to standard English due to recent scientific achievements.

In order to classify languages, it is necessary to delineate linguistic characteristics that are not universal. These properties constitute basic blocks on which the linguist relies in order to group languages. The main units that distinguish one language of others are:

- *Phonemes*: these are the sounds found in the language. They are classified in vowels or in consonantal sounds. For example, in Finish there are eight vowel sounds, whereas in Portuguese and Japanese there are five.
- *Morphemes*: these are the basic units of language that are able to convey meaning. Root words, as *land* in *island*, are examples of morphemes. Other examples are the prefixes, and suffixes. Individual words are composed by the combination of morphemes.

- *Syntax*: the syntax of a language is a set of rules that describes how the words can be combined in order to determine more complex sentences. Different languages may, or may not use similar rules. For instance, in English the adjective precede the nouns; just the opposite of Portuguese. However, in both languages phrases are constituted by a subject, a verb and an object, in this order.

3.2 Linguistic Typology

Linguist typology uses common features of languages in order to classify them into types. Examples of such features are the sounds present in the language, or the order in which words appear. For example, every language contains the concepts of *subject*, *verb* and *object*. The verb is a description of a state or action that applies on the subject of the phrase. The object complements the verb, by receiving, for instance, the action executed by the subject. In the phrase *John loves Mary*, John is the subject, loves is the verb, and Mary is the object. This three basic constituents can appear in six different orders, depending on the language, and can be used as a parameter of language classification. This ordering, however, is not strict. There are languages in which the verb is split into two or more pieces. For instance, in German it is possible to say *Im Wald habe ich einen Fuchs gesehen*, meaning *In-the wood have I a fox seen*. The typologic analysis simplifies such cases, by only considering phrases in which the verb is defined by a single word. Thus, German is classified as subject/verb/object, or verb/subject/object. There are languages in which any order of subject, verb and object is possible. In such languages, generally the role of each syntactical construct is defined by means of particles, affixes or suffixes. Examples of this types of languages include Polish and Latin.

There are two main types of languages, according to typologists [11]. The first type is known as *analytic languages*. The second is formed by the so called *synthetic languages*. In analytic languages, words are not inflected, and the grammatical role of each construct in the phrase is determined by the word order, or by the use of particles. For example, in Mandarin, one would say: *Wo (I) mai (buy) juzi (orange) chi (eat)*. In this case, the position of subject, verb and object is fixed, and determines the meaning of the phrase.

In the group of syntactic languages, the meaning of each word in the phrase is determined by means of morphological markers that are connected to the word stem. The three main types of syntactic languages are:

- *inflectional languages*: the meaning of the word is determined by the use of inflections. These are suffixes that are connected to the word stem. For example, in Portuguese, the verbal tense is determined by such suffixes: *Eu ando* (I walk), *Eu andava* (I used to walk), and *Eu andaria* (I would walk).

- *agglutinative languages*: in this type of languages, words are built up from a long sequence of morphemes. Each of them express a different grammatical meaning. A typical example is swahili:

```
mimi  ni  -na      -ku  -penda  wewe
me    I   PRESENT  you  love    you
```

- *incorporating languages*: in such languages, it is possible to build a whole phrase out of a single word. These words may be formed by inflection, or agglutination. Tiwi is an example of this language type:

```
ngi  -rru  -unthing       -apu  -kani
I    PAST  for some time  eat   repeatedly
```

3.3 Genetic Classification

The genetic model aims to classify languages into families that share a common ancestor. One of the objectives of genetic classification is to reconstruct the so called *proto-languages*. These are old, extinct languages that gave origin to the modern set of spoken languages. Linguist geneticists argue that languages evolve according to a tree pattern. Old languages constitute trunks from which new languages branch, or detach. Several reasons can influence the branching of languages. New languages can be the result of social segmentation, or human migration. Also new languages can arise from the contact between different peoples. Examples of this last case are the creole languages.

3.3.1 Classification by Cognate Words

One of the mechanisms used by linguists in order to trace relations between languages is based on *lists of cognates*. Cognates are words that have a common origin. According to Wikipedia [5], examples of cognates are the words *night* (English), *Nacht* (German), *noc* (Czech), *nox* (Latin), and *nakti* (Sanskrit), all meaning night and all deriving from a common Indo-European origin. Obviously, related languages tend to share more cognates than unrelated languages; therefore, the number of cognate words in the intersection between the vocabulary of different languages gives a quantitative measure of the degree of proximity between them.

Although generally accepted as one of the most important vehicles of language classification, the use of lists of cognates present some drawbacks. Firstly, it has limitations in terms of time, and is not effective to trace relations between languages that have been detached from each other since long ago [15]. Languages are dynamic entities: they evolve, and, during such evolution, they incorporate new words, while losing old ones. Some estimates point that given any language, and a list of words currently in its lexicon, approximately 20% of them will have been lost in one millennium. After 6,000 years of separation, two languages are not expected to bear more than 7% of cognates. A smaller number can be the result of chance; therefore, 6,000 years is the period of time generally accepted as the maximum limit of analyzes based on cognate words.

A second disadvantage in the use of cognates is the existence of the so called *false cognates*. Two words of different languages are false cognates if, at a first glance, they seem to have a common root, but linguistic analyzes shows that they have distinct origins. Examples of false cognates are the Latin verb *habere*, and the German verb *haben* [5]. The former word means to receive, and the latter, to grasp. Although they have similar sound and spelling, they evolved from different roots in the Proto-Indo-European language.

There are words that do not have any know relation at all, and, nevertheless, seem to be really similar. This happens by chance, mainly in words that are part of the core vocabulary of a language, that is, that are used often by ordinary speakers. Classical examples are *tu* (Korean for "two"), and the English *two*. Another striking example is the word *dog*, that in the Australian Aboriginal language Mbabaram happens to be pronounced in the same way, in spite of these languages having no common ancestor.

Thirdly, even though two words are cognates, it does not means that the languages to which they belong are related. The contact between different peoples permits the exchange of vocabulary. For instance, the verb *deletar*, used in Portuguese, was incorporated after the English word *to delete*. *Delete* comes from the old Latin word *deletus*. Although Portugues is a Romance language, and has its origins in the vulgar Latin, the verb *deletar* was incorporated into its lexicon less than thirty years ago.

3.3.2 Glottochronology

The same dynamic aspect that does not allow to use lists of cognates to trace relations between languages that have been detached since long ago permits to determine the date of separation between languages. This technique is know as *glottochronology* [17]. It is possible to establish an analogy between glottochronology and the dating technique based on the decay of carbon 14. It is accepted that languages lose words at a rate of approximately one word out of five at each one thousand years. Therefore, given a list of cognates from two different languages, it is possible to determine the approximate period in time in which they were so closely related that could be considered the same.

The *glottolinguist* uses a list of universal words in order to compare languages. Actually, this is a list of cognates. In this way, the words *cow*, and the German *Kuh* are considered the same. The most common cognates two languages present, the most related they are, and the earlier their period of detachment. Because languages evolve slowly, the list of universal cognates does not include words that have been coined recently, such as *television* and *Internet*. The words are mostly related to body parts, numbers, colors, animals and universal verbs. Some of these words, as presented in [13], are listed in Table 1.

3.3.3 Genetic Markers

Genetic markers are characteristics shared by related languages that hardly would happen by chance. In order to constitute a proof of relatedness, a genetic marker must attend two properties: (i) the chance of independent occurrence must be close to zero. (ii) The likelihood of diffusion must be low. The first

all	ashes	bark	belly	big	bird	bite	black
blood	bone	breast	burn	claw	cloud	cold	come
die	dog	drink	dry	ear	earth	eat	egg
eye	fat	feather	fire	fish	fly	foot	full
give	good	green	hair	hand	head	hear	heart
horn	I	kill	knee	know	leaf	lie	liver
long	louse	man	many	meat	moon	mountain	
mouth	name	neck	new	night	nose	not	one
person	rain	red	road	root	round	sand	say
see	seed	sit	skin	sleep	small	smoke	stand
star	stone	sun	swim	tail	that	this	thou
tongue	tooth	tree	two	walk	warm	water	we
what	white	who	woman	yellow			

Table 1: List of universal words used in glottoanalyzis.

property avoids the arising of the genetic marker by chance, whereas the second avoids the transmission of genetic markers among families of non-related languages. Therefore, if two languages contain a common genetic marker, the most plausible explanation is the existence of a common ancestor for them.

An example of genetic marker was used to determine the parenthood between the Algonquian family of languages, and two languages spoken in the cost of California: Yurok and Wiyot. The Algonquian family is formed by several languages that were spoken in the plains of North America before the European colonization. Algonquian is not similar to Yurok or Wiyot, neither these latter languages present any resemblance. However, it was recognized that all these languages present similar consonants as prefixes of verbs and nouns when used with different pronominal persons [7]. In this case, the genetic markers are the consonants used in the first person (**n-*), in the second person (**k-*), in the third person (**w-*), and in the indefinite person (**m-*). Remarkably, the Algonquian languages, Yurok and Wiyot present the same prefixes. The only exception in the Yurok first person (**d-*) and indefinite person (**b-*), which can be shown to be evolved from **n-* and **m-*.

4 Family of Languages

As discussed in Section 3 there are a number of techniques that allow to establish families of related languages. A group of languages that can be securely demonstrated to have a common origin is called a *stock*. In general, languages in a stock have detached from each other in relatively recent time, after all, most of the linguistic analyses lose effectiveness when applied to languages more than 6,000 years old. Stocks are generally represented by an structure called family

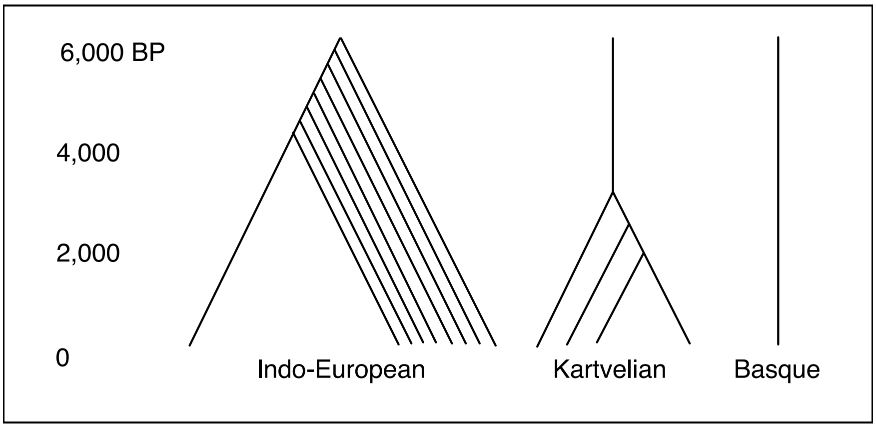

Figure 1: Examples of family trees.

tree. The family tree outlines the period of time when languages start becoming independent of each other. For instance, the diagram in Figure 1 shows three examples of linguistic stocks. These diagrams only represent major branches, but not individual languages.

The first tree represents the family of Indo-European languages. This family comprises the most widely spoken languages in Europe and Americas. Examples include English, German, Spanish, French, Greek, Albanian, and Armenian. Descents of the Proto-Indo-European language start dispersing as early as 6,000 years ago. The second stock represent the *Kartvelian* family of languages. The Kartvelian family includes the Georgian, Megrelian, Laz and Svan languages. These languages are most used in a large territory to the south of the Transcaucasian chain of mountains. Linguists agree that the first split from the original Kartvelian took place around 4,500 years ago [15]. In the diagram, the original language is shown by the vertical line. There are stocks with relatively few branches, and there are stocks with no branch whatsoever. Languages with no branch, or just a few recent dialects, are called *isolates*. An example of isolate is the Basque language, spoken by a small population living in the borders of Spain and France.

There is no consensus about the number of stocks on Earth. The number may not be greater than 300, and most linguistics agree that this number can be further reduced in more than one hundred, by finding relations among different stocks. While some stocks are almost unknown in the linguistic community, others have been deeply studied. This section describes some of the better know stocks, according to a list proposed by several authors [9]. The numbers following each family name describes the quantity of different languages in the family.

- **Niger-Congo** (*1514*) This family was proposed by Joseph Greenberg in 1966 [8]. It is the largest African family in terms of area and number of speakers, and may constitute the world's largest family in terms of number of languages. Examples of languages include *Swahili*, spoken in central Africa, and *Dogon*, spoken in Mali.

- **Austronesian** (*1268*) The family comprises languages spoken in the islands of Southeast Asia and the Pacific Ocean. Some elements in this family are Javanese, with more than 80 million speakers, Malay, Tagalog and Malagasy. The latter is spoken in Madagascar.

- **Trans-New Guine** (*564*) This family is formed mostly by languages spoken in the island of New-Guinea. This is a remarkable area, with one of the highest densities of languages in the world. Although the number of languages in this family is big, the number of native speakers is not. This is due, mostly, to the fact that the native speakers live in primitive settlements, which contributes to the diversity of languages.

- **Indo-European** (*449*) This is the family of languages that have the largest number of native speakers. Languages in this group include English, German, France, Spanish, Albanian, Indo-Iranian, Sanskrit and Armenian. All these languages are believe to descent from the Proto-Indo-European language, which was probably spoken during the Neolithic and Bronze Ages.

- **Sino-Tibetan** (*403*) The Sino-tibetan family is the second largest family in terms of number of speakers. Most of the Sino-Tibetan languages are tonal, that is, the pitch influences the meaning of the word. Examples of languages include all the main Chinese languages, such as Mandarin and Cantonese, as well as Tibetan and Burmese.

- **Afro-Asiatic** (*375*) The Afro-Asiatic family includes languages that are spoken in the north of Africa, and in the Middle-East. In comprises the subfamily of Semitic languages. Some of the Semitic languages are among the oldest languages with written record: Aramaic, Moabite and Phoenician. Presently, the most spoken languages of this family is Arabic, with its several dialects.

- **Australian** (*263*) The Australian family is constituted by several distinct aboriginal languages. Such languages are mostly native of Australia, and some nearby islands. There is no consensus about the exact constitution of the Australian family, because some of its members are thought to be isolates. Also, since the arrival of English colonizers, the number of native speakers of Australian aborigine languages has been decreasing.

- **Nilo-Saharan** (*204*) This is a controversial family of languages that are spoken for approximately 11 million of speakers. Although the number of native speakers is not big, the family is more diverse than the Indo-European family, and there is no agreement regarding the classification of some of its proposed elements. For instance, some linguists [1] consider the *Kadu language* to be Nilo-Saharan, while others [8] put it in the group of *Kordofanian languages*. There are even theories that classify Kadu as an small isolate [6]. The languages in this family as spoken along the upper part of the Nile, in the region of the African Lakes, and in plateaus of the south Sahara.

- **Oto-Manguean** (*174*) This is a family of languages spoken in Mexico before the arrival of the Spaniards. Most of these languages are extinct, and others have just a few thousands of speakers. But they have contributed several words to the spanish vocabulary that is used in Mexico. Examples of languages in this group include *Zapotec* and *Mixtec*.

- **Austro-Asiatic** (*169*) The languages in the Austro-Asiatic group are spoken in Southeast Asia and India. It is generally believed that this is the original family of languages spoken in the region. Other families (Indo-European, Sinto-Tibetan, etc) were incorporated as results of movements of population. Perhaps the most important language in this family is the *Khmer*, spoken in Cambodia, and also in the neighboring countries: Vietnam, Thailand, Burma and Laos.

- **Sepik-Ramu** (*100*) This is, probably, the most controversial family in this list of examples. It has been proposed in order to group several isolates and minor stocks from Papua New Guinea. This family is likely to not stand intact to a careful scrutiny, because most of the elements used in its classification, such as personal pronouns, are thought to be areal features, instead of genetic characteristics.

- **Tai-Kadai** (*76*) This family had its origin in Southern China. After prehistoric migrations, it started to be used in Thailand and Laos. Among the most spoken examples, are *Thai*, the national language of Thiland, *Laotian*, the national language of Laos, and Zhuang, a language spoken in Southern China. All these three languages are classified as *Tai languages*.

- **Tupi** (*76*) The Tupi family is constituted by a large group of languages that were spoken along the Brazilian coast at the time of the arrival of Portuguese conquerers. It is believed that the number of native speakers before the Portuguese discovery was around 5 million. Presently this number is not larger than 200 thousands, and is decreasing rapidly. Although the number of native speakers is decreasing, Tupi languages have contributed considerably to the Portuguese vocabulary spoken in Brazil.

- **Dravidian** (*73*) This family is formed by languages spoken mainly in Southern India and Sri Lanka. Presently, there are more than 200 million native speakers. According to linguists [10], the Dravidian family seems not to be related to any other of the know stocks of languages, although similarities between Dravidian and the Uralic group of languages suggest prolonged contact between these families. Examples of languages in this group are *Tamil*, *Kannada* and Telugu.

- **Mayan** (*69*) This family is formed by languages as old as 5,000 years. They are spoken in South-Eastern Mexico, and Central America by at least 3 million people. In these regions, Spanish is the official language; however, Mayan dialects are used as primary or secondary languages. There are written records of Mayan languages, which were found in pottery, buildings and in highly elaborate scripts, known as the Maya hieroglyphics.

5 Applications of Language Classification

5.1 The Age of Languages

Some estimates of the age of the genus *Homo* put it around 1,800,000 years [3]. The age of the specie *Homo Sapiens* is, probably, between 500,000 years [2] and 150,000 years [12]. Such estimates are mostly based on carbon dating of human fossils. The anatomical features that enable humans to manipulate sounds in order to communicate ideas are present in human beings since the arouse of the *Homo sapiens*. Therefore, it is safe to assume that the biological features that make human language possible have existed since at least 100,000 years. On the other hand, there is no sufficient archeological evidence to establish a period as the starting point of human languages.

One of the techniques used to estimate the age of human languages essentially statistical, as described by Johanna Nichols [14]. The basic mechanisms consists in comparing the number of languages in the present against the number of languages that existed 6,000 years ago, in order to reach a rate in which new languages are appearing. By assuming that all the existing languages came from a single ancestor, it is possible to estimate the age of languages.

According to Nichols, statistics shows that the number of distinct stocks grows by 50% at every period of 6,000 years. Thus, one way to estimate the age of languages is to divide the number of existing stocks by 1.5, until the resulting quantity is inferior than 2. The number of divisions is then multiplied by 6,000, and the resulting product is the final estimate.

As discussed in Section 4, it is believed that the number of stocks on Earth is between 200 and 300. The first value assigns languages 72,000 years; the latter, 78,000. This number is, nevertheless, less than the anatomical estimate, which makes human languages possible in periods as early as 100,000 ago. One explanation for the discrepancy is the extinction of stocks. Another, is the fact that languages the process of language diversification has being faster in recent days than in the primeval times of humankind.

5.2 Reconstruction of Languages

The classification of languages allows the linguist to know how languages change, and, ultimately, allow her to reconstruct extinct languages. There are two basic strategies that are applied in the reconstruction of languages: *Internal Reconstruction* and *external reconstruction*. In the internal reconstruction, a known language is used as the only source of information in order to reconstruct an old one. In the external reconstruction paradigm, also know as *comparative reconstruction*, systematic relationships between genetically related languages are used in order to recover aspects of a dead language such as the sound and spelling of morphemes. In order to illustrate how external reconstruction can be used, notice that the following process may occur as a consequence of language evolution:

- *Weakening*: this is a type of modification in which a lessening in the time or degree of a consonant's closure occurs. For example, in Latin, the word *maturus* (mature) has changed into *maduro* in Portuguese.
- *Metathesis*: this is a reordering in the sequence of segments that constitutes a morpheme. For example, the word dirt evolved from *drit* by metathesis.
- *Vowel reduction*: this is a process that converts a full vowel to a less strong sound. For example, *stanas*, from the old English, has changed to stones due to vowel reduction.

The following example appears in [11], and consists in the reconstruction of the word *father* in Old Latin. The descendants of Latin are the Romance languages: Catalan, French, Italian, etc. In these languages, the word for father has the following spellings: *padre* (Italian), *pare* (Catalan), and *pere* (French). The following processes can have occurred during the evolution of Latin into its descendant languages:

- weakening: *t* may have changed into *d*;
- metathesis: *er* may have changed into *re*;
- vowel reduction: *a* in the first syllable may have changed into *e*;

Due to the previous analysis, it is possible to conclude that the original word, in Latin, could have been *pater*, *peter*, and even *p?ter*. If this process of reconstruction is applied to other words of Latin, it is possible to conclude that the correct form is *pater*.

6 Conclusion

This paper has discussed several topics related to the classification of languages. One of the conclusions is that there remains yet a long way ahead until linguists can reach an agreement about how languages should be grouped into families. Two main types of classification: genetic and typologic, have been presented. In both models of classification there exists some discordance among linguists. However, the divergences are greater when considering the genetic paradigm.

In despite of the variances between linguists, it is a fact that some languages share common ancestor languages. This can be shown clearly if the language is not too old. A number of methods to evince this fact, such as glottochronology, and the use of genetic markers have being discussed in this document.

Also, it seems to be clear that the classification of languages gives the linguists theoretical tools to reach conclusions about how languages evolve, how old languages can be, and how languages change over time. I believe that, with the advances of research in linguistic area, more and more of the open questions regarding the origin of languages will be able to be answered. This is a remarkable fact, because, in my opinion, to know the primordial language is to know the origin of human societies. In addition, to understand how language evolves is to understand how human beings evolve, both in physical and behavioral terms.

References

[1] Roger Blench. Is niger-congo simply a branch of nilo-saharan? In *Fifth Nilo-Saharan Linguistics Colloquium*, pages 36 – 49, 1992.

[2] Noel T Boaz and Alan J Almguist. *Biological Anthropology: A Synthetic Approach to Human Evolution*. Prentice Hall, 1993.

[3] Jeffrey Calister. *Brief Review in Earth Science*. Prentice Hall, 1993.

[4] Michael C Corballis. The gestural origins of language. *American Scientist*, 87(2):138, 1999.

[5] Public domain. The free encyclopedia: Cognates, 2005. http://en.wikipedia.org/wiki/Cognate - last visit: May 2005.

[6] Christopher Ehret. *A Historical-Comparative Reconstruction of Nilo-Saharan*. Koln, 2001.

[7] I Goddard. Algonquin, wiyot, and yurok: proving a distant genetic relationship. *Linguistics and Anthropology: In Honor of C. F. Voegelin*, 26(1):249 – 262, 1975.

[8] Joseph Greenberg. *The Languages of Africa*. Indiana University Press, 1966.

[9] SIL International. Ethnologue: Languages of the world, 2005. http://www.ethnologue.com - last visit: May 2005.

[10] Bhadriraju Krishnamurti. *The Dravidian Languages*. Cambridge University Press, 2003.

[11] G Krynicki. Families of languages and types of linguistic classification, 2005. http://elex.amu.edu.pl/ krynicki - last visit: May 2005.

[12] Roger Lewin. Molecular genetics and the age of man, 1997. New Scientist:5.

[13] Ohn Lienhard. Engines of our ingenuity, 2005. http://www.uh.edu/engines/epi1566.htm - last visit: May 2005.

[14] Johanna Nichols. *Linguistic Diversity in Space and Time*. Chicago Press, 1992.

[15] Johanna Nichols. *The Origin and Dispersal of Languages: Linguistic Evidence*, chapter 7th. Allen Press, 1998.

[16] Malcom Ross. *Contact-induced change and the comparative method: cases from Papua New Guinea*, chapter 7th, pages 180 – 217. Oxford University Press, 1996.

[17] W S Y Wang. Glottochronology, lexicostatistics, and other numerical methods. *The Encyclopedia of Language and Linguistics*, 3, 1994.