Yuri Tambovtsev.
Novosibirsk Pedagogical University
Novosibirsk, Russia

Typological Proximity of Esperanto to the Classical Language Taxa of the Indo-European Family

Abstract. The article discusses the possibility of classifying Esperanto into a language group on the basis of the typological distances between languages. The distances are derived by the method of measuring distances in multi-dimensional Euclidean space. An antique language — Latin — has been taken as a reference point for Esperanto. The new method is based on both phonostatistics and metric analysis. A new outlook on language classification is proposed here. It is founded on the structure of the frequency of occurrence of consonants in the speech sound chain. It is a good clue for defining the typological closeness of languages. It allows a linguist to find the typological distances between a language (in this case - Esperanto) and the other languages of different genetic groups of a language family (in this case - the Indo-European language family). This method can put any language in a language taxon, i.e. a sub-group, a group or a family. The minimum distance may be a good clue for placing a language (in this case – an artificial language, Esperanto) in this or that language taxon. The method of calculating Euclidean distances between an artificial language and natural languages is used for the first time.

Key words: consonants, phonological, similarity and distance, articulatory, features, typology, frequency of occurrence, speech sound chain, statistics, artificial, natural, language similarity, comparative method, Euclidean distances, closeness, language taxon, taxa of languages, classification.

Introduction

The aim of this article is to analyse the typological distances between Esperanto and some other natural languages, mainly Indo-European. The distances are derived by the method of measuring language similarity in multi-dimensional Euclidean space on the basis of the frequency of phonemic occurrence in the language sound chain. The frequency of linguistic units is a reliable indicator of similarity or difference between languages as linguistic objects comprised of texts (Tambovtsev et al., 2007: 14). The distances between the languages are studied to show which language taxon they must be included in. The total of the distances between all the languages in a language group shows how compact this or that language taxon is (Tambovtsev, 2005-b: 5).

Some linguists may say it is not possible to compare an artificial language with a natural language. Let us consider if Esperanto is really an artificial language on the phonetic level. It is artificial on the lexical and syntax level, because different word stocks are mixed, different affixes from different languages are taken and so on. However, on the phonetic level Esperanto is in no way different from any natural language. If on the lexical level it has hybrid words, then on the phonetic level there are no hybrids. Esperanto uses natural phonemes. Thus, in this respect, it is quite a natural language. Really, every linguist remembers that Esperanto is an artificial language. However, if a linguist analyses Esperanto phonemes, then he comes to the conclusion that the constructions of its phonemes are not hybrids. Its speech sounds are the same as the speech sounds in the natural languages. Thus, from the point of view of phonetics Esperanto is the same sort of object as any other natural language. It has the same phonetic (articulatory) features as a natural language. Esperanto's speech sounds may be classified in the same way as the speech sounds of any natural language according to the work of the organs of speech, the manner of articulations, the work of the vocal cords, etc. (Tambovtsev, 2001-a: 98 - 102). Nevertheless, Esperanto, Glossa, Ido, Interlingua, Novial and Volapuek are called artificial

languages (Crystal, 1992: 28-29) because they are constructed as hybrids on the lexical level. Thus, the word structure plays the predominant role in their classification. However, from the point of view of phonetics, all these artificial languages may be compared to natural languages since they can be pronounced. If they can be pronounced, then they have the same sets of phonetic (articulatory) features.

World languages may be classified in language groups by different means. Genetic classification of world languages by the classical comparative method is widely accepted. Nevertheless, April and Robert McMahon are quite correct in saying that there are various methods for classifying languages, which yield different results, i.e., different classifications. All of them have some drawbacks and some are extremely controversial. They give the Greenberg method of mass comparison as an example of the most controversial method, i.e., which gives the most controversial result of classification (McMahon et al., 2005: 5). At the same time, they point to the fact that the well-known comparative method fails to reconstruct Latin on the basis of the Romance languages, which are the daughter languages of Latin (McMahon et al., 2005:14-19). One should agree with them that it is high time to improve the comparative method or to invent a new, more reliable method. Our method may be another good method which can help to classify languages into language taxa. It does so according to their sound similarity. It gives formal and repeatable results, which the comparative method cannot give. Stating that the comparative method is rather subjective, the McMahons believe that this brings the comparative method uncomfortably close to Greenberg's method of mass comparison, where accepting results means believing implicitly in the linguistic intuitions of the method inventor. Both methods rely on an individual linguist's knowledge of a particular language group, but this makes both inevitably subject to inference from the individual linguist's opinion. One cannot help agreeing to them that if we cannot guarantee getting the same results from the same data considered by

different linguists, we jeopardize the essential scientific criterion of repeatability (McMahon et al., 2005: 27).

An artificial language, Esperanto, has been taken as a reference point. The reason for which Esperanto has been chosen is that we know the source from which it obtained its basic stock of words. This artificial language may help us to find out how well our method works. L. L. Zamenhof, a doctor from Poland, created Esperanto in 1887. He had a nickname "Esperanto" which means "Hoping" in Latin. His nickname became the official name of this artificial language. The language of Esperanto comprises the roots of some European languages: Latin, Old Greek, Greek, Italian, Spanish, French, English, Russian, and Polish. These roots serve as the base to create comprehensive speech with the help of some affixes, which transfer the grammatical meanings (BSE, Vol. 30, 1978: 251; Tambovtsev, 2001-a: 98).

It seemed to us really interesting to see into which language group of the Indo-European languages Esperanto would go. One should also take into account that the sound chain of Esperanto may be unique because it uses several fixed affixes. Therefore, its sound chain in speech may be quite different from any other natural language.

The new method is based on both phonostatistics and metric analysis in order to put it in this or that language taxon. The new method of phonostatistics and metric measurements developed by the author is proposed here (Tambovtsev, 1977; 2002-c; 2002-d; 2003-b; 2004-a; 2004-b). It allows a linguist to find the typological distances between the languages under study (Tambovtsev, 1994-b; 2001-d; 2002-a). The obtained distances indicate to which language taxon a language belongs. In fact, the received language distances show similarity between the languages in question, the less the distance - the more similar the languages (Tambovtsev, 2001-e; 2002-b; 2004-a).

Usually, linguists agree on the genetic classification of Indo-European languages since it has a long historical tradition. The classification of Altaic, Uralic or Caucasian language taxa is often criticised. We argued elsewhere that there are more opponents to than supporters of calling these taxa families (Tambovtsev, 2001-b).

Lyle Campbell is correct in stating that the nomenclature of the Nostratic language taxon is not stable. He is quite right to ask what exactly it is. Is it a language unity, union or community? Lyle Campbell points out that it is necessary to solve the problem of the Altaic taxon. Is it a family, a unity or a union? Nevertheless, without a solution for these problems, different linguists construct different language classifications, often without a substantial foundation (Campbell, 1998).

It is advisable to set up a strict hierarchy in the classification of languages. It is logical to define the following language taxa:

Subdialect
Dialect
Language
Branch
Subgroup
Group
Family
Unity
Union
Phyle
Community

Unfortunately, until now such taxa as language subgroup may be mixed with language group, while language group may be mixed with language family or language unity with language union, etc. (Tambovtsev, 2005: 136). It is necessary to warn that genetic relatedness

can usually be proved not higher than such a language taxon as a language family. Thus, we can demonstrate it by the example of the Mansi language, which has the following classification hierarchy:

Subdialect (Verhnelozva, Ljapin, Sosva)
Dialect (Northern - Ljapin-Sosva; Southern - Tavda; Western -Pelym-Vagil; Eastern - Konda)
Language (Mansi)
Branch (Ob-Ugric)
Subgroup (Ugric)
Group (Perm-Ugric)
Family (Finno-Ugric)
Unity (Uralic)
Union (Uralo-Altaic)
Phyle (Uralo - Paleo-Asiatic)
Community (World languages)

As far as the Mansi language is concerned, its dialects are so different that their speakers do not understand each other. Thus, they may be defined as separate languages. The same is true for Hanty. For instance, the Yug language was recently considered a dialect of Ket (Tambovtsev, 2005: 137; Werner et al., 1979: 48-49). In fact, Mansi dialects are further away from each other than those dialects of the Ket language (Tambovtsev, 2001-b).

It is a great pity but even for some well-studied languages like Russian the classification is not developed completely (Tambovtsev, 2005: 137). Russian has not been classified in a hierarchy higher than a family.

Subdialect (different regional subdialects of Russian)
Dialect (Northern, Middle and Southern dialects of Russian)
Language (Russian)
Branch (not developed)

Subgroup (Eastern subgroup of Slavonic group)
Group (Slavonic)
Family (Indo-European)
Unity (not developed)
Union (not developed)
Phyle (not developed)
Community (world language)

The same is true for English, German, French and other Indo-European languages. Nevertheless, some linguists propose such a language taxon as Nostratic, but they do not define if it is a language unity, union, phyle or community. By the sample of languages the Nostratists choose, it looks more like "language community". However, as Lyle Campbell correctly points out, the nomenclature of Nostratic languages of V. M. Illich-Svitych is different to that of T. Markey, V.V. Shevoroshkin, A. R. Bomhard, J.D. Bengtson, M. Ruhlen and others (Campbell, 1998: 109).

Until now, Esperanto has not been put into any group of natural languages. Before classifying Esperanto, let us consider the classification of Latin since Esperanto was lexically based on the stock of Latin root words. However, it is necessary to draw the attention of the reader to the fact that our classification is based on the phonetic features, but not lexical ones. The other very important note is that the nomenclatures of Esperanto and Latin speech sounds are rather different. The phonemic nomenclature of Esperanto is more similar to Italian. Some linguists may object to the classification of an artificial language Esperanto, when some natural languages have not been classified completely. Thus, there is a classification problem with such a well-studied language as Latin.

Some natural languages may change the group into which they are classified. Let us take Latin as an example. Now Latin is classified into the Italic group of the Indo-European language

family (Crystal, 1992: 199; JaDM, 1982: 19). However, not so long ago Latin was placed into one group with the Romance languages because Latin is the parent language for all the Romance languages; therefore all the Romance languages are the daughter languages of Latin (Chikobava, 1953: 207-208). Maybe it is more logical for the parent language to be in the same group with its offspring. It would be very strange if we put Old Slavonic in some separate group, but not in the Slavonic group. Our method shows the typological distances which may shed light on the closeness of Latin to the Romance languages since it is not possible to find enough long and reliable texts in the true Italic languages: Faliscan, Oscan, Umbrian and Venetic, which are dead by now. Therefore, Latin may have been placed in this language group for a lack of information. Though the number of texts in the Italic languages is limited and they are short, there are some linguists who claim that Latin belongs to the group of Italic languages. Rex E. Wallace goes even further than that. He claims without much evidence that Latin enters the Latino-Faliscan group of the Italic branch of the Indo-European language family (Wallace, 2001: 412). One must pay attention to the fact that this opens a new group and a new branch. More logical is to call his new group the Latino-Faliscan subgroup, even though his new branch is nothing else but the commonly accepted Italic group within the Indo-European language family. Though the information on the other Italic languages is scarce and unreliable, Rex E. Wallace insists that Oscan, Umbrian, South Picene, Vestinian, Marrucinian, Paelignian, Marsian, Volscian, Aequian and Hernican are more distant from Latin that Faliscan (Wallace, 2001: 412). However, it is quite possible that all the Italic languages mentioned above are just the sub-dialects and dialects of Latin. Though usually Latin is a term for the Classical Latin language, which was used only by the educated classes of Rome. Rex E. Wallace correctly points out that there were numerous different sub-dialects and dialects of Latin. He is also right to state that there were variants of

Latin for different social levels, e.g. Vulgar Latin was the speech of the common folk (Wallace, 2001: 412).

One should keep in mind that our typological classification of Esperanto is quite preliminary. It is possible to agree that meanwhile it is advisable to place Latin into the Italic group of the Indo-European language family until more solid and reliable information is received. At the same time one cannot agree to the fact that this group is called a language family. A fair representative of the linguists who believe that there could be a family inside a family is David Crystal (Crystal, 1992: 199). Unfortunately, he is not the only one who makes a logical mistake like this. April McMahon and Robert McMahon also speak about the Germanic family, which is embraced into the Indo-European language family (McMahon et al., 2005: 3-4). However, if one takes into consideration all the reasoning of their book, one may realise that the abundance of data leads them to the conclusion that Indo-European family looks like a sort of a super-family, called here a language unity, i.e. the next level of classification. Usually, the languages as the objects at this higher level are not so similar as at the lower levels. If a classification is correct, i.e. natural, then the languages at the lower levels are more similar (Tambovtsev, 2004-a: 201 - 210; Tambovtsev, 2004-b: 147 - 151).

It is high time to reconsider all the established language families and other language taxa. If it is done so, then it may be discovered that Italic and Romance groups must be merged together into one group called Romano-Italic with two subgroups: Romance and Italic. There are some arguments, which allow us to do it. One of the arguments may be the distance between Latin and the Romance languages (Tambovtsev, 2001-a). If Latin is closer to the languages of the Romance group of languages, then it surely belongs to them, rather than to any other set of languages. Our results show the shortest mean distance of Latin is to the languages of the Romance group. The same sort of reasoning we apply to Esperanto. We'll put it to the group of

languages with which it has the smallest total distance. In this case it is more convenient to use the shortest mean distance (c.f. Tab 1 - 13).

We can suppose that Esperanto may have the closest mean distance with some group or subgroup of the Indo-European language family. Let us consider how well Indo-European classification structure is constructed from the point of view of logics. It is good to see that the logical mistake of classification described above is not made by other classifiers. Thus, Kenneth Katzner calls Italic a subgroup of the Indo-European language family (Katzner, 1986:2). However, strictly speaking he also makes a sort of a logical mistake, since his subgroup does not enter a group, but a family. Thus, he omits one classification step. A logical classification of languages must incorporate subgroups into a group, groups into a family, families into a unity, unities into a phylum, phyla into a union, unions into a language community (Tambovtsev, 2004: 145).

It is high time to establish a universal and strict logical hierarchy of language taxa. All the linguists in the world should keep to one and the same order of language taxa (Tambovtsev, 2003-a: 3). The ordered series of the taxa of the world languages should include old and dead languages like Latin, Old Greek, Old Russian, Old Turkic, etc (Tambovtsev, 2001-b; Tambovtsev, 2001-b; Tambovtsev, 2001-c). While reconsidering and building new language taxa linguists should take into account the special rules. First of all it is the idea that they must separate all world languages into sets in such a way that the distances between languages in a language taxon must be less than the distances of these languages to the other world languages (Tambovtsev, 2003-a). The structure of a taxon is more dense (tight), that is compact, if the languages selected for it are more similar (Tambovtsev, 2002-b). In our studies it is usually the total of the distances between the ideal language in this or that set of language, which is expressed by the mean of a set (Tambovtsev, 2001-e). In a compact set the distances between the

mean and the other values are minimal. First we developed this idea of compact and sparse sets of languages on the data of the frequency of occurrence of phonemes in the speech chain (Tambovtsev, 1977). Then, we went on applying the idea of the measure of compactness on the basis of the consonantal coefficient, which is the ratio of the frequency of occurrence (Tambovtsev, 1986)

We have nothing against placing Latin in the group of Italic languages of the Indo-European language family. Nevertheless, it is necessary to point out that in physics, chemistry, biology and other natural sciences old classifications are often reconsidered (Kuhn, 1977; Rozova, 1986). We must also point out that based on the same known Indo-European isoglosses, Tomas V. Gamkrelidze and Vjacheslav Vs. Ivanov do not construct the group of Romance languages and the Italic group of the Indo-European language family. Instead, they define only one group of languages, i.e. the Italic group. Presumably, their Italic group embraces both Italic and Romance languages, since they do not provide a separate Romance group (Gamkrelidze et al.,1984: 415). It is fruitful that they also include not only the phonetic but the lexical and grammatical isoglosses, which allows them to obtain a more complete and reliable scheme. We have analysed this scheme in detail elsewhere and come to conclusion that their scheme is different from the usual traditional one in this aspect (Tambovtsev, 1989134- 137).

When comparing the distances between Esperanto and Old Greek or Modern Greek one must bear in mind that Old Greek and Modern Greek are considered genetically isolated languages (Crystal, 1992: 11; JaDM, 1982: 23). In a way Esperanto can also be considered a sort of an isolated language. There are some other languages, which have not been placed into any language family: Basque, Japanese, Korean, Ainu, Nivhi, Yukaghir and Ket (Yug). However, for the latter, a new language family - Yenissey has been invented. So, now Ket with all its dialects is the only memeber of the Yenissey family. Nevertheless, it is not a solution to the problem. If

we follow this way, then we must also establish separate language families for Ainu, Basque, Japanese and the other isolated languages.

The new data, which we received for Esperanto (or Latin, or any other language) may allow it to enter this or that group of languages. It is the first attempt to establish the phonostatistical measures for the typological closeness of Esperanto with the language groups, to which it may be supposed to enter. Usually, genetically close languages are also typologically close. However, typologically close languages may or may not be genetically close. Nevertheless, in the majority of cases typologically close languages are genetically close. We can find phonostatistical closeness, which can give a good clue for genetic relatedness. It was found for some Finno-Ugric, Turkic, Mongolic, Tungus-Manchurian and Paleo-Asiatic languages (Tambovtsev, 2001-d; 2001-e; 2002-a; 2002-b; 2002-c; 2002-d; 2003-a; 2003-b; 2004). There is, then, a good reason to believe this method should also work for Esperanto as it works for Mansi, Hanty, Jakut, Latin or any other language.

Why should one use quantitative methods in studying languages? A great philosopher and scientist Immanuil Kant (1724 - 1804) in his well-known works explaining the structure of the world stated that everything in this world possesses quantity and quality. Quantitative data characterise an object sometimes better, especially when the objects are very similar. Languages are similar in their qualitative characteristics. Thus, one should rely more on quantitative characteristics. Actually, quantity may go over into quality when it is great enough (FS, 1980: 144). In this case, English is a fair example. Must it be considered a Germanic or a Romance language? Many words of its stock are of Romance origin as the result of the Norman Conquest in 1066. It is believed that quantitative characteristics work better in the cases when qualitative characteristics fail to distinguish two linguistic objects.

Long ago, in 1935, George Kingsley Zipf stated that it was necessary to introduce so-called "Dynamic Philology" to achieve fruitful results in studying the structure and entity of Language (Zipf. 1935:XII). As George A. Miller correctly put in the introduction to Zipf's book, one who wishes to study a rose should count its petals, not just enjoy it. G. K. Zipf believed that it is necessary to study the massive statistical regularity of every linguistic unit or phenomenon (Zipf, 1935:V-VI).

Quantitative research needs the use of mathematical statistics. One can't help agreeing with Christopher Butler, who requires a quantitative treatment in any linguistic research because it is difficult otherwise to understand and evaluate how relevant are the linguistic results (Butler, 1998: 255 - 264).

When establishing genetic language families, linguists compare every language with some other language or a group of languages. Jiri Kramsky is correct to remark that one can establish a typology of languages based on the quantitative data received after comparing languages. The quantitative data gives a clearer vision of the differences and similarities between languages. The quantitative load of particular language phenomena is different in different languages. Kramsky is quite right to observe that in linguistics there is a very close relation between quality and quantity, even if the conditions of the transition of quantity into quality are not established so safely as they are in natural sciences. Nevertheless J. Kramsky assumes that in linguistics qualitative changes are asserted with the help of quantitative factors (Kramsky, 1972: 15).

Our method measures distances between languages on the phonological level. It gives a vivid picture of the typological similarity of the sound pictures of the languages under investigation and allows us to find out the archetype of this or that language family using the mean values of the frequency of the consonantal groups.

The use of quantitative data ensures that the languages are similar if the frequency of occurrence of certain linguistic units is similar. It takes into account both cases, when the units are used very frequently or very seldom. However, in classical linguistics, where the frequency is not taken into consideration, it is more often than not that the usual elements are compared with the rare elements. J. Kramsky is correct to point out that the language units which are in the centre of some language system should not be compared to those of the periphery (Kramsky, 1972: 15). The quantitative analysis shows us the units which are in the centre of a language system and those which are at the periphery of it. Therefore, the typology of languages based on the quantitative data may add much to the established language families (Tambovtsev, 2001-a; 2001-b; 2001-c; 2003).

Latin, as any other human language, has a specific structure in the speech sound chain. It can be distinguished by its structure from any other language. Every language has a unique structure of distributions of speech sounds in its phonemic chain. The distribution of Latin vowels will not be considered till the second stage of the investigation. The frequency of occurrence will be considered if and only if the frequency of occurrence of different groups of consonants will not differentiate Latin from the other world languages. Let's point out that consonants bear the semantic load in the word, not vowels. Therefore, it is more possible to understand the meaning of the message by consonants, rather by vowels. Some linguists use consonants to consider statistical models in language taxonomy.

Let us consider the way one statistical method, namely, Chi-square, which has been successfully used in linguistics to solve a problem in English. It is a well-known fact that English has undergone a considerable change under the influence of French after the Norman Conquest. Some linguists point out that about 70% of words are of French or Latin origin. One can see the details elsewhere (Tambovtsev, 2001) In this case, it is applied to place English and German in

one language group. On the basis of the frequency of fricative consonants [s] and [f] Alan Ross proved, and April and Robert McMahon proved again, that English and German are related, i.e., the use of these fricative consonants is not random (McMahon et al., 2005: 59-61). Actually, an outstanding American mathematician of Hungarian origin, G. Polya, used the same method of reasoning to establish the degree of similarity of Hungarian to English, Swedish, Danish, Dutch, German, French, Spanish, Italian and Polish. He came to the conclusion on the sample of ten numerals that Hungarian is quite different from these languages (Polya, 1975: 315 - 319).

However, if we fail to recognise and distinguish two languages, then we resort to the structure of occurrence of vowels in the speech sound chain. While comparing languages, it is necessary to keep to the principle of *commensurability*. Having that in mind, it is not possible to compare languages on the basis of the frequency of occurrence of separate phonemes, because the sets of phonemes in languages are usually different. But the articulartory features may serve as the basic features in phono-typological reasoning.

Before making computer measurements of the phonological distances, one has to choose the phonological features which are necessary and sufficient, i.e., one has to select the system of informative features. In pattern recognition such features are called basic (Zagoruiko, 1972: 54 - 75). Therefore, we have chosen all the features basic to the articulation of any speech sound. At the first stage we shall deal with consonants.

First is the classification of consonants according to the work of the active organ of speech or place of articulation (4 features). Second is the classification from the point of view of the manner of articulation or the type of the obstruction (3 features). Third is the classification according to the work of the vocal cords (1 feature). In this way, 8 basic features are obtained: 1) labial; 2) forelingual or front; 3) mediolingual or palatal; 4) guttural (back) or velar; 5) sonorant; 6) occlusive non-sonorant; 7) fricative non-sonorant; and 8) voiced non-sonorant consonants.

One should take the values of the frequency of occurrence of these 8 features in the speech chain of Latin and compare them to those of the other languages. On the basis of the "Chi-square" test and Euclidean distance, we have developed our own method of measuring the phono-typological distances between languages (Tambovtsev, 1994-a; 1994-b; 2004). It takes into account the frequency of occurrence of the 8 consonantal groups mentioned above and builds up the overwhelming mosaic of the language sound picture.

It is very important to find some typological characteristics in order to endeavour to place a language in some defined language family. Some linguists consider it impossible to put Esperanto into any of the known language families because it was never studied before as a natural language related to other natural languages. Actually, it is considered here that it is possible to put Esperanto in a language family if its phonostatistical characteristics are studied from the same articulation point of view as we studied all the other human languages from. Therefore, we undertook the study of the frequency of Esperanto phonemes on the vast sample of Esperanto texts. Fortunately, unlike the dead Italic languages mentioned above besides Latin, (Latin has an abundance of reliable texts), Esperanto has also many long and reliable texts.

The rules of reading in Esperanto are easy and stable. All the letters in Esperanto are read in the usual way as in a transcription, for example, b [b], p [p], v [v], etc. Only 5 letters may give some problem while reading them: c [ts], c^ [tʃ], g^ [dž], j^ [ž], s^ [ʃ]. Sometimes the "caps" go before these letters: ^c, ^j, ^g, ^s. Sometimes the letter <x> is used instead of the cap. It comes after these 5 letters, for example, cx, jx, gx, sx.

We fed into the computer the following texts:

1) *Master and Margarita* by M. Bulgakov, translated into Esperanto by Dr. Sergej Pokrovskij

2) B. Prus "Faraon"

3) Bible

(The author thanks Prof. Yuri M. Plusnin and Dr. Sergej B. Pokrovskij for the Esperanto texts and consultations on Esperanto.)

After I.V. Sergeev, we define the following Esperanto phonemes (Sergeev, 1961: 5 - 17):

Vowels [i, u, u^, e, o, a]
Consonants [p, b, v, f, m, t, d, s, z, ts, ʃ, tʃ, dž, ž, n, l, r, j, k, g, x, h]

The classification of the Esperanto consonants by the place of articulation:

Labial [p, b, v, f, m]
Forelingual (front) [t, ts, d, s, z, ʃ, tʃ, dž, ž, n, l, r]
Mediolingual (palatal) [j]
Guttural (velar or back) [k, g, x, h]

The classification the Esperanto phonemes by the manner of articulation (the character of the obstruction):

Sonorant [m, n, l, r, j]
Occlusive non-sonorant [p, b, t, d, ts, tʃ, dž, k, g]
Fricative non-sonorant [v, f, s, z, ʃ, ž, h]

The classification of the Esperanto phonemes by the work of the vocal cords:

Voiced non-sonorant consonants [b, v, d, z, ž, dž, g]

The Esperanto text of the book by M. Bulgakov, *Master and Margarita*, was divided into several parts in order to check the stability of the phonemic groups in the different parts of the book. The first part of the book ("Never talk to strangers") shows the following frequency of occurrence of the group of consonants in the sound chain:

	Frequency	% to all phones	% to consonants
Labial	2,024	8.80	16.47
Forelingual (front)	8,494	36.93	69.12
Palatal (mediolingual)	522	2.27	4.25
Guttural (back)	1,249	5.43	10.16

Sonorant	5,773	25.10	46.98
Occlusive non-sonorant	4,186	18.20	34.06
Fricative non-sonorant	2,330	10.13	18.96
Voiced non-sonorant	1,410	6.13	11.47

The total of consonants: 53.43%

The total of vowels: 46.57%

The value of the consonantal coefficient (i.e. the ratio of consonants to vowels): 1.15

Sample volume of the first part of *Master and Margarita*: 23,000 phonemes.

The second part of *Master and Margarita*, "Pontij Pilat," shows the following frequency of occurrence of the groups of consonants in the sound chain:

	Frequency	% to all phones	% to consonants
Labial	3,604	8.58	16.06
Forelingual (front)	15,607	37.16	69.59
Palatal (mediolingual)	970	2.31	4.33
Guttural (back)	2,247	5.35	10.02
Sonorant	10,706	25.49	47.73
Occlusive non-sonorant	7,791	18.55	34.74
Fricative non-sonorant	3,931	9.36	17.53
Voiced non-sonorant	2,520	6.00	11.24

The total of consonants: 53.40%

The total of vowels: 46.60%

The value of the consonantal coefficient (i.e. the ratio of consonants to vowels): 1.15

Sample volume of the second part of *Master and Margarita*: 42,000 phonemes.

The first and second part of the book “Master and Margarita” showed the following frequency of occurrence of the groups of consonants in the sound chain:

	Frequency	% to all phones	% to consonants
Labial	5,629	8.66	16.21
Forelingual (front)	24,102	37.08	69.43
Palatal (mediolingual)	1,489	2.29	4.29

Guttural (back)	3,497	5.38	10.07
Sonorant	16,478	25.35	47.46
Occlusive non-sonorant	11,979	18.43	34.51
Fricative non-sonorant	6,260	9.63	18.03
Voiced non-sonorant	3,926	6.04	11.31

The total of consonants: 53.41%

The total of vowels: 46.59%

The value of the consonantal coefficient: 1.15

Sample volume of the 2nd part of "Master and Margarita": 65,000 phonemes.

The fourth ("Chase"), the fifth ("In Griboedovo") and sixth ("Schizophrenia") parts of the book showed the following frequency of occurrence of the groups of consonants in the sound chain:

	Frequency	% to all phones	% to consonants
Labial	4,286	8.93	16.48
Forelingual (front)	17,856	37.20	68.66
Palatal (mediolingual)	1,258	2.62	4.84
Guttural (back)	2,606	5.43	10.02
Sonorant	12,062	25.13	46.38
Occlusive non-sonorant	8,990	18.73	34.57
Fricative non-sonorant	4,954	10.32	19.05
Voiced non-sonorant	2,933	6.11	11.28

The total of consonants: 54.18%

The total of vowels: 45.82%

The value of the consonantal coefficient: 1.18

Sample volume of the 4th, the 5th and the 6th parts of "Master and Margarita": 48,000 phonemes.

The Esperanto text of the novel "Pharaoh" by B. Prus was processed with the following resulting data for the frequency of occurrence of the groups of consonants in the sound chain:

	Frequency	% to all phones	% to consonants
Labial	8,254	9.07	16.45
Forelingual (front)	33,687	37.02	67.13
Palatal (mediolingual)	3,558	3.58	6.48
Guttural (back)	4,987	5.48	9.94
Sonorant	24,215	26.61	48.25
Occlusive non-sonorant	16,862	18.53	33.60
Fricative non-sonorant	9,109	10.01	18.15
Voiced non-sonorant	5,597	6.15	11.15

The total of consonants: 55.15%

The total of vowels: 44.85%

The value of the consonantal coefficient: 1.23

Sample volume of *Pharaoh*: 91,000 phonemes.

The Esperanto text of the novel *Hero of Our Times* (*Princess Mary*) by M. Yu Lermontov was processed with the following resulting data for the frequency of occurrence of the groups of consonants in the sound chain:

	Frequency	% to all phones	% to consonants
Labial	5,747	9.74	17.98
Forelingual (front)	21,435	36.33	67.08
Palatal (mediolingual)	1,723	2.92	5.39
Guttural (back)	3,050	5.17	9.55
Sonorant	14,980	25.39	46.88
Occlusive non-sonorant	10,562	17.90	33.05
Fricative non-sonorant	6,413	10.87	20.07
Voiced non-sonorant	3,334	5.65	10.43

The total of consonants: 54.16%

The total of vowels: 45.84%

The value of the consonantal coefficient: 1.18

Sample volume of *Princess Mary*: 59,000 phonemes.

The Esperanto text of the Old Testament of the Bible "The Book of Daniel" gave the following results for the frequency of occurrence of the groups of consonants in the sound chain:

	Frequency	% to all phones	% to consonants
Labial	3,423	7.78	14.35
Forelingual (front)	15,897	36.13	66.65
Palatal (mediolingual)	1,861	4.23	7.80
Guttural (back)	2,671	6.07	11.20
Sonorant	11,304	25.69	47.39
Occlusive non-sonorant	8,087	18.38	33.91
Fricative non-sonorant	4,461	10.14	18.70
Voiced non-sonorant	2,983	6.78	12.51

The total of consonants: 54.21%

The total of vowels: 45.79%

The value of the consonantal coefficient: 1.18

Sample volume of “The Book of Daniel”: 44,000 phonemes.

Proverbs and sayings in Esperanto were processed to obtain the following results for the frequency of occurrence of the groups of consonants in the sound chain:

	Frequency	% to all phones	% to consonants
Labial	5,003	9.81	18.52
Forelingual (front)	18,911	37.08	69.98
Palatal (mediolingual)	867	1.70	3.21
Guttural (back)	2,239	4.39	8.29
Sonorant	12,128	23.78	44.89
Occlusive non-sonorant	8,966	17.58	33.18
Fricative non-sonorant	5,926	11.62	21.93
Voiced non-sonorant	3,279	6.43	12.14

The total of consonants: 52.98%

The total of vowels: 47.02%

The value of the consonantal coefficient: 1.13

Sample volume of proverbs and sayings in Esperanto: 51,000 phonemes.

In order to find the comprehensive sound picture of Esperanto, it is advisable to sum up all the data provided above. The total of the texts in Esperanto, yielding 358,000 phonemes, provided the following data for the frequency of occurrence of the groups of consonants in the sound chain:

	Frequency	% to all phones	% to consonants
Labial	32,342	9.03	16.68
Forelingual (front)	131,888	36.85	68.09
Palatal (mediolingual)	10,456	2.92	5.40
Guttural (back)	19,050	5.32	9.83
Sonorant	91,167	25.47	47.06
Occlusive non-sonorant	65,446	18.28	33.78
Fricative non-sonorant	37,123	10.37	19.16
Voiced non-sonorant	22,052	6.16	11.38

The total of consonants: 54.12%

The total of vowels: 45.88%

The value of the consonantal coefficient: 1.18

Sample volume of all the texts in Esperanto: 358,000 phonemes.

As we pointed out above, Esperanto was created on the basis of several European languages, Latin among them, and the main Esperanto root word stock was taken from Latin. Let us compare the data of the frequency of occurrence of the Esperanto groups of consonants to some other languages, especially Latin. When we compare Latin words to the Esperanto words, we notice that their forms are different since the affixes are different. The frequency of affixes from the other languages, such as Greek, Italian, Polish, Russian, etc. may be great. Thus, they add different sound colouring to Esperanto. The sounds of these affixes are those which are not used in Latin, [tʃ, dž ʃ, ž]. These sounds may increase the distances between Latin and Esperanto.

After computing the Latin text by V. N. Kuprejanova, N. M. Umnova, Ja.M. Borovskij and A.V. Boldyrev, we found the following frequencies of the phonemic occurrence in the sound chain:

	Frequency	% to all phones	% to consonants
Labial	4,561	13.82	24.12
Forelingual (front)	12,248	37.12	64.77
Palatal (mediolingual)	140	0.42	0.73
Guttural (back)	1,964	5.95	10.38
Sonorant	7,463	22.62	39.47
Occlusive non-sonorant	7,297	22.11	38.58
Fricative non-sonorant	4,153	12.58	21.95
Voiced non-sonorant	2,702	8.19	14.29

The total of consonants: 18,913 phonemes - 57.31%

The total of vowels: 14,087 - 42.69%

The value of the consonantal coefficient: 1.34

Sample volume of the Latin proverbs: 33,000 phonemes.

Zipf's Latin data have the following frequency of phonemic occurrence in the sound chain (Zipf et al., 1939):

	Frequency	% to all phones	% to consonants
Labial	560	11.20	20.86
Forelingual (front)	1,705	34.10	63.50
Palatal (mediolingual)	25	0.50	0.93
Guttural (back)	395	7.90	14.71
Sonorant	1,076	21.52	40.07
Occlusive non-sonorant	1,149	22.98	42.79
Fricative non-sonorant	460	9.20	17.13
Voiced non-sonorant	260	5.20	9.68

The total of consonants: 2685 phonemes - 53.70%

The total of vowels: 2315 - 46.30%

The value of the consonantal coefficient: 1.16

Sample volume of the Zipf's Latin text: 5000 phonemes.

The author has also computed the epic poem *Aeneidos* by Vergilius. Publius Vergilius Maro received a good education in philosophy, poetry and rhetoric. He worked on his poem for some 11 years. It is considered to be a good sample of classical Latin. *Aeneid* has the following frequency of the phonemic occurrence in the sound chain:

	Frequency	% to all phones	% to consonants
Labia	43,514	12.15	21.19
Forelingual (front)	135,892	37.95	66.20
Palatal (mediolingual)	1,504	0.41	0.72
Guttural (back)	24,411	6.82	11.89
Sonorant	80,515	22.48	39.21
Occlusive non-sonorant	82,351	23.00	40.12
Fricative non-sonorant	42,455	11.85	20.67
Voiced non-sonorant	25,218	7.04	12.28

The total of consonants: 205,321 phonemes - 57.33%

The total of vowels: 152,800 - 42.67%

The value of the consonantal coefficient: 1.34

Sample volume of the Latin text of *Aeneid*: 358,121 phonemes.

The united data computed by the author consists of Latin proverbs and *Aeneid*. It has the following frequency of the phonemic occurrence in the sound chain:

	Frequency	% to all phones	% to consonant
Labial	48,075	12.29	20.97
Forelingual (front)	148,140	37.88	64.63
Palatal (mediolingual)	1,644	0.42	0.73
Guttural (back)	26,375	6.74	11.76
Sonorant	87,978	22.49	39.23
Occlusive non-sonorant	89,648	22.92	9.98
Fricative non-sonorant	46,608	11.92	20.79

Voiced non-sonorant	27,920	7.14	12.45

The total of consonants: 224,234 phonemes - 57.33%

The total of vowels: 166,887 - 42.67%

The value of the consonantal coefficient: 1.34

Sample volume of the Latin text of Aeneid: 358,121 phonemes.

In linguistics it is recommended to use some exact measure to place the languages more objectively. In pattern recognition such exact measures of distances between two objects are used. Nikolai G. Zagoruiko recommends using Euclidean distances when the value of the features are equal (Zagoruiko, 1999: 198 - 199). We consider all our features to be equal since we cannot claim that the frequency of occurrence of labials is more important than the frequency of occurrence of sonorants, or the frequency of occurrence of palatals is more important than the frequency of occurrence of the fricatives, and so on.

We measure here the distances by the well-known formula of measuring the distance between points in the Euclidean space:

$D = \sqrt{(x1 - x2)^2 + (y1 - y2)^2 + (z1 - z2)^2 + \text{etc.}}$

where

D - distance

x1 - the frequency of occurrence of labials in Latin

x2 - the frequency of occurrence of labials in the second language

y1 - the frequency of occurrence of front consonants in Latin

y2 - the frequency of occurrence of front consonants in the second language

z1 - the frequency of occurrence of palatals in Latin

z2 - the frequency of occurrence of palatals in the second language, etc.

The details of calculating Euclidean distances between two phonetically transcribed texts may be found elsewhere (Tambovtsev, 2003-c: 122). This method is good because it can use any number of features in any number of languages. Therefore, a linguist can take as many linguistic features as he wants. Nor is the number of languages limited. So, this method was used to

calculate the distance between Basque and Latin (10.54), although the least distances were between Basque and Kazah (5.310) or Tofalar (5.96) and the other Turkic languages (Tambovtsev, 2003-c: 125).

It is necessary to introduce some system of references when dealing with the distances between Latin and the other languages. Such a point may be the distance between two texts in some language. We calculated the distances between two texts in the Markiz language, one of the Austronesian languages at 0.505. Now let us take any other language as a point for the system of references. It can be any language, which is far away from Latin and contact with which is not probable. Such a language may be Ainu, whose native speakers live in Japan. So the influence of Latin on Ainu is not possible. For calculating the distances between Ainu and the other languages we used the same method. The language closest to Ainu is one of the Austronesian languages - Tagalog with the distance of 9.310. The closest language to Latin by this method is Moldavian (4.275), then comes Italian (5.242) and Romanian (6.913). We can see that Latin is much closer to Moldavian than Ainu to its closest language. In fact, it is closer by two times. We can see the other distances between Latin and Romance languages in Table 1 (Appendix).

The distance between Latin and Moldavian, the least calculated, means that they are the closest languages among the chosen Romance and other languages (c.f. Table 1-13). It is not surprising since Moldavian and Romanian are spoken by the descendants of Roman soldiers and settlers, who occupied the Roman province of Dacia (Carlton, 2001: 598). In my mind, Italian, Moldavian, and Romanian preserved the articulation base of Latin and thus the frequency of occurrence of sounds in Latin and in these languages is more similar than in the others. Actually, this (smallest) distance between Latin and Moldavian may speak for many more remnants in Moldavian, rather than Italian. It is always so that at the periphery there are more obsolete

features than in the centre. These distances may also point out that the articulation base of these three languages is rather similar.

As a matter of fact, the *articulation base* is the main factor in measuring the frequency of occurrence of speech sounds in any language. We can see it in the examples of other languages, e.g. Ainu. Let us remember the words of N. A. Nevskij, that Ainu is close to Paleo-Asiatic languages (Tambovtsev, 2001-b). Indeed, one of the Paleo-Asiatic languages, i.e., the Chookchi language with the distance 10.954 is rather close. The next closest language is also a Paleo-Asiatic language, Koriak, with the distance 12.781. Korean is a bit closer at 12.636. Japanese is more far away, 15.269. As we can see from the tables below, the other languages are also rather far away. The closest Tungus-Manchurian language is Ul'ch with the distance 13.464.

However, the closest to Ainu proved to be the American Indian languages of the North and South America. So, Quechua has the distance of 5.451 and Inga 7.388. They both belong to the Quechua family of American Indian languages. Quechua and Inga Indians live in South America.

Let us take some other languages as reference points. Japanese is a good choice since it is an isolated language. Having compared Japanese to some languages, we obtained the following phono-typological distances: Japanese - Ujgur (6.77); Japanese - Nanaj (8.12); Japanese - Jakut (8.26); Japanese - See Dajak (8.86); Japanese - Kazah (9.02); Japanese - Turkish (9.05); Japanese - Ket (9.52); Japanese - Baraba Tatar (9.76); Japanese - Uzbek (10.63); Japanese - Hausa (10.98); Japanese - Georgean (11.05); Japanese - Kazan Tatar (11.07) and so on. One can see, that Ujgur, Jakut, Kazah, Turkish, Baraba Tatar, Uzbek and Kazan Tatar are Turkic languages. Nanaj is a Tungus-Manchurian language. Therefore, one can notice that Japanese is closer to the so-called Altaic languages which include Turkic, Mongolian and Tungus-Manchurian languages. Many world languages were compared to Japanese, although we can't show all the distances here

for lack of space. However, the maximum distances were found for Japanese - German (22,24); Japanese - English (19.83); Japanese - Rumanian (15,08) and Japanese - Swedish (17.03).

Finally, we can also state that the speech sound picture of Japanese is rather far away from the following languages, which are geographically close: Chinese, Nivh, Itelmen, and Indonesian, which was a surprise to us. Our data state that the speech sound pattern of Japanese resembles that of Ujgur - one of the Turkic languages spoken in the Middle Asia. The Ujgur people are often linked to the Old Turkic tribes, who used to live in the steppes of Southern Russia before the Tatar-Mongols captured them in the 9th century A.D. We must point out that this is not a coincidence, since the other native Altaic people have very similar data of closeness to Japanese. Turkic and Tungus-Manchurian tribes may have had a sort of common origin with Japanese. It may verify the Altaic hypothesis of Japanese origin. This is especially vivid, when the Austro-Oceanic and other languages do not show such a great closeness.

Considering the mean distance between Latin and the other languages and sets of languages, one may notice a clear preference. The mean distance between Latin and the Romance languages is the least 6.706 (c.f. Table1). The Baltic languages (Latvian and Lithuanian) are also rather close (8.504) to Latin (c.f. Table 5). Latin is closer in general to the Eastern Slavonic languages (Russian, Old Russian, Ukrainian and Belorussian), than to the other two Slavonic subgroups. The mean distance is less (9.259) than that of Latin to Southern Slavonic (9.810) or Western Slavonic (13.008). So, it speaks again for similarity between Eastern and Southern Slavonic subgroups (c.f. Tab 1 - 4). The Iranian group is closer (10.673), than Germanic (11.160) or Indic (12.400) groups. It is possible to see that Old Greek (8.482) and Modern Greek (8.653) are not so close to Latin. However, Armenian is a bit further (8.838). Albanian is not close enough either (9.325).

Nevertheless, the Indo-European languages are closer to Latin than the Samoyedic family (15.400) or the Ob-Ugrian subgroup of the Ugric group of the Finno-Ugric family(16.333). The Northern dialect of Mansi (19.017) or the Konda dialect of Mansi (18.261) may be the champions (c.f. Table 14).

In conclusion, it is possible to state that Esperanto is similar to Old Greek. This means that the groups of consonants are distributed in Esperanto and Old Greek in more or less the same way. One should bear in mind that this is a typological closeness, though genetically related languages are always typologically close. There is also a great typological closeness between Esperanto and some languages of the Romance group of the Indo-European family. We are far from stating that genetically Esperanto is closer to the languages of the Romance group, since it is an artificial language, but typologically it is certainly much closer than to the languages of the other language groups. It is not as close to the Italic group, represented here by Latin, on whose base it was partly created. However, this can be explained by the absence in Latin several phonemes [t͡ʃ, dž ʃ, ž], which can be found in Italian and other languages, which show the minimal distances. Although Esperanto is an artificial language, from the point of view of phono-typology it is very similar to the natural languages of the Romance languages.

Having this typological clue, linguists may have a closer look at some natural languages which are included in this or that language taxon incorrectly from the genetic point of view. Maybe it is advisable to reconsider both the Italic and Romance groups and unite them into one group, Romano-Italic, with two sub-groups, Romance and Italic.

Appendix of Tables

Tables of EUCLIDEAN DISTANCES between Esperanto and other world languages, united in different genetic groups, families, and other language taxa.

…….Table 1

Phonostatistical EUCLIDEAN DISTANCES between Esperanto and Romance language group of the Indo-European language family. The mean of the distances: 7.211

Language	Distance
	Esperanto
0. Esperanto	0.000
1. Italian	4.255
2. Moldavian	6.072
3. Spanish	8.192
4. Romanian	8.308
5. Portuguese	9.230

…….Table 2

Phonostatistical EUCLIDEAN DISTANCES between Esperanto and Italic language group of the Indo-European language family

Language	Distance
	Esperanto
0. Esperanto	0.000
Latin	7.330

……..…Table 3

Phonostatistical EUCLIDEAN DISTANCES between Esperanto and the Eastern Subgroup of the Slavonic language group of the Indo-European language family. The mean of the distances - 11.389

Language	Distance
	Esperanto
0. Esperanto	0.000
1.Russian	8.691
2.Old Russian	11.441
3.Ukranian	12.349
4.Belorussian	13.075

…….Table 4

Phonostatistical EUCLIDEAN DISTANCES between Esperanto and the Southern Subgroup of the Slavonic language group of the Indo-European language family. The mean of the distances - 10.885

Language	Distance
	Esperanto
0. Esperanto	0.000
1.Serbian	7.795
2.Slovenian	9.739
3.Macedonian	12.210
4.Bulgarian	13.797

…….Table 5

Phonostatistical EUCLIDEAN DISTANCES between Esperanto and the Western Subgroup of the Slavonic language group of the Indo-European language family. The mean of the distances - 13.500

Language	Distance
	Esperanto
0. Esperanto	0.000
1. Sorbian (Luzhik)	10.801
2.Slovak	12.295
3.Czech	13.079
4.Polish	17.828

…….Table6

Phonostatistical EUCLIDEAN DISTANCES between Esperanto and Baltic language group of the Indo-European language family. The mean of the distances - 11.344

Language	Distance
	Esperanto
0. Esperanto	0.000
1.Latvian	9.662
2. Lithuanian	13.026

…….Table 7

Phonostatistical EUCLIDEAN DISTANCES between Esperanto and Indic language group of the Indo-European language family. The mean of the distances - 9.468

Language	Distance
	Esperanto
0. Esperanto	0.000
1.Sanscrit	6.898
2. Gypsy	6.975
3.Marathi	7.611
4.Bengali	11.916
5.Hindi	13.942

…….Table 8

Phonostatistical EUCLIDEAN DISTANCES between Esperanto and Iranian language group of the Indo-European language family. The mean of the distances - 13.134

Language	Distance
	Esperanto
0. Esperanto	0.000
1.Persian (Iranian)	10.093
2.Osetian	11.917
3.Tadjik	17.391

…….Table 9

Phonostatistical EUCLIDEAN DISTANCES between Esperanto and Celtic language group of the Indo-European language family

Language	Distance
	Latin
0. Latin	0.000
1.Irish	9.137

…….Table 10

Phonostatistical EUCLIDEAN DISTANCES between Esperanto and Germanic language group of the Indo-European language family. The mean of the distances - 11.160

Language	Distance
	Esperanto
0. Esperanto	0.000
1.Old English	8.024
2.Norwegian	8.119
3.Gothic	9.008
4.English	14.574
5.German	17.469

…….Table 11

Phonostatistical EUCLIDEAN DISTANCES between Esperanto and Isolated languages of the Indo-European language family

Language	Distance
0. Latin	0.000
Old Greek	4.209
Armenian	9.021
Greek	9.615
Albanian	14.812

…….Table 12

Phonostatistical EUCLIDEAN DISTANCES between Latin and Romance language group of the Indo-European language family. The mean of the distances - 6.706

Language	Distance
	Latin
0. Latin	0.000
1. Moldavian	4.275
2. Italian	5.242
3. Rumanian	6.913
4. Spanish	7.353
5. Portuguese	9.747

……..…Table 13

Phonostatistical EUCLIDEAN DISTANCES between Latin and the Eastern Subgroup of the Slavonic language group of the Indo-European language family. The mean of the distances - 9.259

Language	Distance
	Latin
0. Latin……	0.000
1. Russian…..……......	4.275
2. Old Russian……..	9.048
3. Belorussian……	10.124
4. Ukrainian……...	10.169

…….Table 14

Phonostatistical EUCLIDEAN DISTANCES between Latin and the Southern Subgroup of the Slavonic language group of the Indo-European language family. The mean of the distances - 9.810

Language	Distance
	Latin
0. Latin……	0.000
1. Macedonian…….....	7.502
2. Slovenian……	8.582
3. Serbian…….	9.579
4. Bulgarian……	13.577

…….Table 15

Phonostatistical EUCLIDEAN DISTANCES between Latin and the Western Subgroup of the Slavonic language group of the Indo-European language family. The mean of the distances - 13.008

Language	Distance
	Latin
0. Latin……	0.000
1. Slovak……	11.653
2. Czech…….	11.743
3. Luzhits-Sorbian….	11.789
4. Polish…….	16.848

…….Table 16

Phonostatistical EUCLIDEAN DISTANCES between Latin and Baltic language group of the Indo-European language family. The mean of the distances - 8.504

Language	Distance
	Latin
0. Latin……	…0.000
1.Latvian……	7.344
2. Lithuanian……..	9.664

…….Table 17

Phonostatistical EUCLIDEAN DISTANCES between Latin and Indic language group of the Indo-European language family. The mean of the distances - 9.231

Language	Distance
	Latin
0. Latin……	0.000
1.Gypsy……...	6.939
2.Sanscrit……..…	8.074
3.Marathi…….……	8.097
4.Bengali…….…..…	10.268
5.Hindi…….……	12.779

…….Table 18

Phonostatistical EUCLIDEAN DISTANCES between Latin and Iranian language group of the Indo-European language family. The mean of the distances - 10.673

Language	Distance
	Latin
0. Latin……	0.000
1.Persian (Iranian)…...	7.877
2.Osetian……..…	9.804
3.Tadjik……..…	14.338

…….Table 19

Phonostatistical EUCLIDEAN DISTANCES between Latin and Celtic language group of the Indo-European language family

Language	Distance
	Latin
0. Latin……	0.000
1. Irish……...	13.057

…….Table 20

Phonostatistical EUCLIDEAN DISTANCES between Esperanto and Germanic language group of the Indo-European language family The mean of the distances - 11.348

Language	Distance
	Latin
0. Latin……	0.000
1.Dutch……	8.075
2.Norwegian……	8.793
3.Old English…..……	10.002
4.English……..…...	11.763
5.Gothic……..……	12.258
6.German……..…...	16.067

…….Table 21

Phonostatistical EUCLIDEAN DISTANCES between Latin and Isolated languages of the Indo-European language family

Language	Distance
	Latin
0. Latin……	0.000
1. Old Greek……	8.482
2. Modern Greek…….	8.653
3. Armenian…… ……	8.838
4. Albanian……..…….	9.325

…….Table 22

Phonostatistical EUCLIDEAN DISTANCES between Latin and Esperanto - an artificial language

Language	Distance
	Latin
0. Latin……	0.000
1.Esperanto……	7.330

…….Table 23

Phonostatistical EUCLIDEAN DISTANCES between Latin and the Ob-Ugric Subgroup of the Ugric language group of the Finno-Ugric language family. The mean of the distances - 16.333

Language	Distance
	Latin
0. Latin……	0.000
1.Eastern Hanty…….	11.823
2.Kazym Hanty…..….	16.231
3.Konda Mansi…….	18.261
4.Northern Mansi…..	19.017

…….Table 24

Phonostatistical EUCLIDEAN DISTANCES between Latin and the Samoedic language family. The mean of the distances - 15.400

Language	Distance
	Latin
0. Latin……	0.000
1.Nenets……	14.375
2.Nganasan……	15.572
3.Selkup…….	16.252

…….Table 25

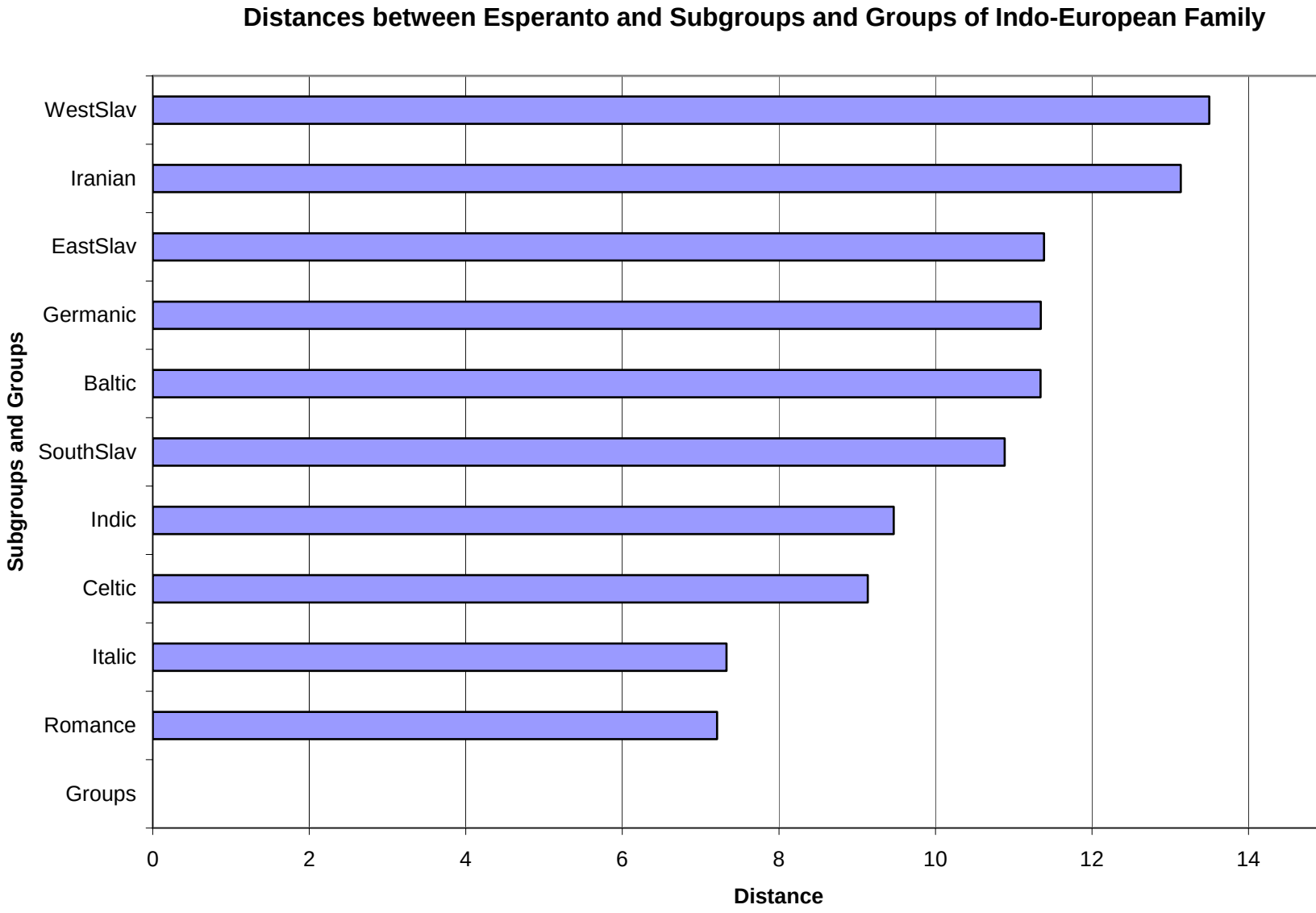
The Ordered Series of the Mean Phonostatistical EUCLIDEAN DISTANCES between Latin and Some Subgroups and Groups of the Indo-European family. The mean of the distances inside every language taxon

Language	Mean Distance
……...Latin	
0. Latin……..0.000	
1. Romance……...6.706	
2. Baltic…….8.504	
3.Eastern Slavonic…..…9.259	
4.Southern Slavonic…..9.810	
5.Iranian……..10.673	
6.Germanic……..…11.160	
7.Indic……..…12.400	
8.Western Slavonic.…13.008	

…….Table 26

The Mean Phonostatistical EUCLIDEAN DISTANCES between Esperanto and the Language Taxa (Subgroups and Groups) of the Indo-European Family

Language Taxon	Mean Distance
Romance group……..7.211	
Celtic group……9.137	
Indic group…….9.468	
Southern subgroup of the Slavonic group..…….10.885	
Baltic group…….11.344	
Germanic group……...11.348	
Eastern subgroup of the Slavonic group.……11.389	
Iranian group……...13.134	
Western subgroup of the Slavonic group.……...13.500	


Distances between Esperanto and Subgroups and Groups of Indo-European Family
Subgroups and Groups
WestSlav
Iranian
EastSlav
Germanic
Baltic
SouthSlav
Indic
Celtic
Italic
Romance
Groups
0
2
4
6
8
10
12
14
Distance

References

Borovskij et al., 1949 - Borovskij Ja. M. and Boldyrev A. V. Latinskij jazyk [Latin]. - Moskva: ILIJA, 1949.

BSE, 1978 - Bol'shaja Sovetskaja Entsyklopedija [Great Soviet Encyclopaedia]. 30 Volumes. Moskva: Sovetskaja Entsiklopedija Publishing House, 1978.

Butler, 1998 - Butler, Christopher. Statistics. - In: Projects in Linguistics. A Practical Guide to Researching Language. Alison Wray, Kate Trott and Aileen Bloomer with Shirley Reay and Chris Butler. - London-New York: Arnold-Hodder, 1998. - 303 p.

Campbell, 1998 - Campbell, Lyle. Nostratic: a personal assessment. - In: Nostratic. Sifting the Evidence (Edited by Joseph C. Salmons and Brian D. Joseph). Current Issues in Linguistic Theory. Vol. 142, 1998, p. 107 - 152.

Carlton, 2001 - Carlton, Charles M. Romanian. - In: Facts About the World's Languages: An Encyclopaedia of the World's Major Languages, Past and Present (Edited by Jane Garry and Carl Rubino). - New York and Dublin: The H. W. Wilson Company- A New England Publishing Associates Book, 2001, p. 598 - 601.

Chikobava, 1953 - Chikobava A, S. Vvedenie v jazykoznanie [Introduction into Linguistics]. - Moskva: Prosveshchenie, 1953. - 243 pages.

Crystal, 1992 - Crystal, David. An Ecyclopedic Dictionary of Language and Languages. - Oxford: Blackwell, 1992. - 428 p.

FS, 1980 - Filosofskij slovar' [Phylosophy Dictionary] (Editor I. T. Frolova). - Moskva: Politizdat, 1980. - 445 pages.

Derjugin et al., 1986 - Derjugin, Aleksandr A., Lukjanova Larisa M. Latinskij jzyk [Latin]. - Moskva: Vysshaja shola, 1986. - 295 p.

Gamkrelidze et al., 1984 - Tomas V. Gamkrelidze and Vjach. Vs. Ivanov. Indo-European and the Indo-Europeans. Volume 1 and 2. [in Russian] - Tbilisi: The Tbilisi University Press, 1984. - 1392 p.

JaDM, 1982 - Jazyki i dialekty mira. [Languages and Dialects of the World]. - Moskva: Nauka, 1982.

Katzner, 1986 - Katzner, Kenneth. The Languages of the World. - London: Routledge and Kegan Paul Publishers, 1986. - 376 pages.

Kuhn, 1977 - Kuhn T. S. The Structure of Scientific Revolutions [in Russian]. - Moskva: Progress Publishing House, 1977. - 300 pages.

McMahon et al, 2005-a - McMahon, April and McMahon, Robert. Language Classification by Numbers. - Oxford: Oxford University Press, 2005. - 263 pages.

Polya, 1975 - Polya, George. Mathematics and Plausible Reasoning. - Moskva: Nauka, 1975. - 463 pages. [Russian translation of his book published in 1954 by Princeton University Press was used].

Rozova, 1986 - Rozova S. S. Klassifikatsionnaja problema v sovremennoj nauke [Classification Problem in Modern Science]. - Novosibirsk: Nauka, 1986. - 223 pages.

Sergeev, 1961 - Sergeev I. V. Osnovy esperanto [Fundamentals of Esperanto]. - Moskva: Gosizdat, 1961. - 164 pages.

Tambovtsev, 1977 - Tambovtsev, Yuri. Nekotorye harakteristiki raspredelenija fonem mansijskogo jazyka [Some Characteristics of the Distribution of the Mansi Language]. - In: Soviet Finno-Ugric Studies, Volume XIII, # 3, 1977 (Tallin), p. 195 - 198.

Tambovtsev, 1986 - Tambovtsev, Yuri. Konsonantnyj koeffitsient v jazykah raznyh semej [Consonant Coefficient in the Languages of Different Families]. - Novosibirsk: Novosibirsk University, 1986. - 17 pages.

Tambovtsev, 1989 - Tambovtsev, Yuri. Review on the book by Tomas V. Gamkrelidze and Vjach. Vs. Ivanov. Indo-European and the Indo-Europeans. Tbilisi: The Tbilisi University Press, 1984. Volume 1 and 2. - In: Gengo Kenkyu, # 96, 1989 (Tokyo), p. 119-143.

Tambovtsev, 1994-a - Tambovtsev, Yuri. Dinamika funktsionorovanija fonem v zvukovyh tsepochkah jazykov razlichnogo stroja. - Novosibirsk : NGU, 1994-a. - 133 pages.

Tambovtsev, 1994-b - Tambovtsev, Yuri. Tipologija uporjadochennosti zvukovyh tsepej v jazyke. - Novosibirsk: NGU, 1994-b. - 199 pages.

Tambovtsev, 2001-a - Tambovtsev, Yuri. Kompendium osnovnyh statisticheskih harakteristik funktsionirovanija soglasnyh fonem v zvukovoj tsepochke anglijskogo, nemetskogo, frantsuskogo i drugih indoevropejskih jazykov. [Compendium of the basic statistical characteristics of functioning of consonants in the speech chain of English, German, French and other Indo-European languages]. - Novosibirsk: Novosibirskij klassicheskij institut, 2001-a. - 129 pages.

Tambovtsev, 2001-b - Tambovtsev, Yuri. Funktsionorovanie soglasnyh fonem v zvukovoj tsepochke uralo-altajskih jazykov. [Functioning of consonants in the sound chain of Ural-Altaic languages]. Novosibirsk: Novosibirskij klassicheskij institut, 2001-b. - 132 pages.

Tambovtsev, 2001-c - Tambovtsev, Yuri, Nekotorye teoreticheskie polozhenija tipologii upor'adochennosti fonem v zvukovoj tsepochke jazyka i kompendium statisticheskih harakteristik osnovnyh grupp soglasnyh fonem. [Some theoretical fundamentals of the typology of oderliness of phonemes in the sound chain of language and the compendium of statistical characteristics of the basic groups of consonants], - Novosibirsk: Novosibirskij klassicheskij institut, 2001-c. - 130 pages.

Tambovtsev, 2001-d - Tambovtsev, Yuri. The phonological distances between Mongolian and Turkic languages based on typological consonantal features. - In: Mongolian studies. Journal of the Mongolia Society (USA), Vol.24, 2001-d, p.41 - 84.

Tambovtsev, 2001-e - Tambovtsev, Yuri. Perednejazychnye soglasnye kak pokazatel' odnoj iz lingvisticheskih universalij jazykov mira [Front consonants as an indicator for one of the universals of world languages]. - In: Sibirskij Lingvisticheskij Seminar, # 2 (2), Novosibirsk, 2001-e, p.18 - 27.

Tambovtsev, 2002-a - Tambovtsev, Yuri. Comparative typological study of language distances based on the consonants in sound chains of various languages. - In: The 5th National Colloquium for Computational Linguistics in the UK. Proceedings of the Conference. (Edited by John Elliot). 8-9 January, 2002. University of Leeds, UK. - Leeds: University of Leeds, 2002-a, p.77-80.

Tambovtsev, 2002-b - Tambovtsev, Yuri. Is Kumandin a Turkic language? - In: Dilbilim Arashtirmalari (Istanbul), 2002-b, p. 63 - 104.

Tambovtsev, 2002-c - Tambovtsev, Yuri. Korean and Japanese as Members of the Altaic Language Family. - In: Abstracts. Permanent International Altaistic Conference 45th Meeting, Budapest June 23-28, 2002. Budapest: Research Group for Altaic Studies. Hungarian Academy of Sciences, 2002-c, p. 13-14.

Tambovtsev, 2002-d - Tambovtsev, Yuri. Structure of the frequency of occurrence of consonants in the speech sound chain as an indicator of the phono-typological closeness of languages. - In: ALL - ACH 2002. New Directions in Humanities Computing. The 14th Joint International Conference, University of Tuebingen, 24 - 28 July, 2002. Conference Abstracts. - Tuebingen: Universitaet Tuebingen, 2002-d, p. 138 - 139.

Tambovtsev, 2003-a - Tambovtsev, Yuri. Tipologija Funktsionorovanija fonem v zvukovoj tsepochke indoevropejskih, paleoaziatskih, uralo-altajskih i drugih jazykov mira: kompaktnost' podgrupp, grupp, semej i drugih jazykovyh taksonov. [Typology of functioning of phonemes in sound chain of Indo-European, Paleo-Asiatic, Ural-Altaic and other world languages: compactness of subgroups, groups, families and other language taxins]. - Novosibirsk: Sibirskij nezavisimyj institut, 2003-a. - 143 pages. [in Russian].

Tambovtsev, 2003-b - Tambovtsev, Yuri. The phonological similarity between Turkic languages based on some phonological features of consonants. - In: Linguistic and Oriental Studies from Poznan', Vol. 5, Poznan', (Poland), 2003-b, p. 85 - 118.

Tambovtsev, 2003-c - Tambovtsev, Yuri. Phonological similarity between Basque and other world languages based on the frequency of occurrence of certain typological consonantal features. - In: The Mathematical Bulletin of Mathematical Linguistics # 79 - 80, 2003 (Praha), p. 121 - 126.

Tambovtsev, 2004-a - Tambovtsev, Yuri. Uralic Language Taxon: Natural or Artificial? (Typological Compactness of Uralic Languages and other Language Taxons: Branches, Subgroups, Groups, Families and Superfamilies). - In: Fenno-Ugristica, # 26, 2004 (Tartu), p. 200 - 246.

Tambovtsev, 2004-b - Tambovtsev, Yuri. Phonostatistical Distance Totals as an Indicator of the Compactness of Language Taxons: a Typologo - Metrical Approach. - In: Lingua Posnaniensis. Vol. XLVI, Poznan' (Poland), p. 145 - 172.

Tambovtsev, 2005-a - Tambovtsev, Yuri. Klassifikatsija jazykov na osnove novyh metodov opredelenija mezhjazykovyh rasstojanij (na primere jazykov samodijskoj i finno-ugorskoj semej [Classification of languages on the basis of some new methods of measuring distances between languages (on the example of languages of the Samoedic and Finno-Ugric families)]. - In: Fenno-Ugristica # 27, Tartu (Estonia), 2005, p. 122 - 140.

Tambovtsev, 2005-b - Tambovtsev, Yuri. Language taxons and the naturalness of their classification. - In: California Linguistic Notes. Vol. XXX, # 2, Fall, 2005, p. 1 - 46.

Tambovtsev et al., 2007 - Tambovtsev, Yuri, Ludmila Tambovtseva and Juliana Tambovtseva. The frequency of linguistic units as an indicator of typological similarity between texts. - In: California Linguistic Notes, Vol. XXXII, # 2, Spring 2007, p. 1- 28.

Wallace, 2001 - Wallace, Rex E. Latin. - In: Facts About the World's Languages: An Encyclopaedia of the World's Major Languages, Past and Present (Edited by Jane Garry and Carl Rubino). - New York and Dublin: The H. W. Wilson Company- A New England Publishing Associates Book, 2001, p.201 - 416.

Werner et al., 1979 - Werner H. K. and Yuri Tambovtsev. Nekotorye rezul'taty fonoststisticheskogo analiza zvukovogo sostava jugskogo jazyka [Some results of the phonostatistical analysis of the sound structure of the Yug language]. - In: Teoreticheskie voprosy fonetiki i grammatiki jazykov narodov SSSR [Theoretical Problems of Phonetics and Grammar of the Languages of the USSR]. - Novosibirsk: Novosibirsk University Press, 1979, p.47 - 52.

Zagoruiko, 1972 - Zagoruiko, Nikolai G. Metody raspoznavanija i ih primenenie. [The Methods of Pattern Recognition and Their Application] (in Russian). - Moskva: Sovetskoe Radio, 1972. - 206 pages.

Zagoruiko, 1999 - Prikladnye metody analiza dannyh i znanij. [Applied Methods of Data and Knowledge Analysis](in Russian). - Novosibirsk: Institute of Mathematics SORAN, 1999. - 269 pages.

Zipf, 1935 - Zipf, George Kingsley. The Psycho-Biology of Language. An Introduction to Psycho-Biology of Language. - Cambridge: Massachusetts Institute of Technology Prtess, 1935. - 336 pages.

Zipf et al., 1939 - Zipf G.K. and Rogers F. M. Phonemes and variphones in four present-day Romance languages and classical Latin from the viewpoint of dynamic philology. - In: Archives Neederlandaises de Phonetique Experimentale, 15, 1939, p. 112 -147.

Prof. Dr. Yuri Tambovtsev
Address: Russia, 630123, Novosibirsk- 123, P.O. Box 104, Department of English and Linguistics of NPU
Home phone:(7-383) - 200 - 14 - 55;
E-mail: yutamb@mail.ru