Yuri Tambovtsev
Novosibirsk Pedagogical University, Russia

Labial consonant distribution in Niger-Congo: A study in typological distance

Abstract. Labial consonants exhibit peculiarities in the Niger-Congo (including Bantu) languages when juxtaposed with the functioning of labial consonants in subgroups, groups, families, and other language taxa of world languages. In this study the frequency of occurrence of the labial consonants is given in per cent to the frequency of occurrence of all the elements in the speech sound chain of the language. The analysis is made with the help of such statistical methods as the coefficient of variance and t-test. The linguistic conclusions about the similarity of functioning of labial consonants are made on the basis of these statistical criteria. It is possible to establish the typological distances between the Niger-Congo language taxon and some other language taxa (Afro-Asiatic, Turkic, Finno-Ugric, Tungus-Manchurian, Iranian, Slavonic, Germanic, etc.), based on the values of the t-test. The comparison shows that the labial consonants are used much more frequently in the Niger-Congo languages than in the other world languages.

Introduction

The goal of the article is twofold: 1) to consider the peculiarities of functioning of labial consonants in the Niger-Congo languages; 2) to compare the peculiarities of functioning of labial consonants in the Niger-Congo taxon to the subgroups, groups, families and other language taxa of world languages. It is possible to establish the typological similarities which may be represented as typological distances between some language taxa (Turkic, Finno-Ugric, Slavonic, etc.) on the values of the t-test. Lindsay J. Whaley is correct to observe that a typological study even focused on a single feature of language may help to understand some

basic facts about the phonology of this or that language taxon (**Whaley, 1997**: 10 – 11). In this case it is the frequency of occurrence of labial consonants in the speech chain.

Niger-Congo languages may be classified into several groups. Usually by the Bantu languages linguists refer to a large group of languages spoken by the Bantu peoples throughout central and southern Africa. David Crystal expresses the common view that the Bantu language taxon may include between 300 and 500 languages. This Bantu taxon has often been treated as a separate language family but nowadays they are usually classified as part of the Benue-Congo group of Niger-Congo languages (**Crystal, 1992**: 39). N.V. Ohotina believes the Bantu languages to have such common labial consonants as [p, b, m]. These may be treated as the basis for reconstructing the common proto-Bantu language (**Ohotina, 1982**: 282). I. N. Toporova gives [w, f, v] as typical for the Bantu languages (**Toporova, 1975**: 11; 16; 22; 58).

Usually, genetically close languages are also typologically close, i.e. similar. In this study they have the least typological distances between them. However, the reverse is not always correct, i.e. typologically close languages may be or may not be genetically close. Nevertheless, in the majority of cases typologically close languages are genetically close. We can find the phonostatistical closeness, which can give a good clue for genetic relatedness, which can later be established by the comparative method (**Tambovtsev, 2001-d; 2001-e; 2002-a; 2002-b; 2002-c; 2002-d; 2003-a; 2003-b; 2004**).

Why should one use quantitative methods in studying languages? A great philosopher and scientist Immanuel Kant (1724 - 1804) in his well-known works explaining the structure of the world stated that everything in this world possesses quantity and quality. Actually, quantity may go over into quality when it is great enough. Therefore, it is important to take into account not only quality, but quantity (**FS, 1980**: 144). It is also important in linguistics.

It is rather strange but in linguistics the qualitative studies are preferred and quantitative ones are neglected.

One can not but agree with Christopher Butler, who requires a quantitative treatment in any linguistic research because it is difficult otherwise to understand and evaluate how relevant the linguistic results may be (**Butler, 1998**: 255 - 264).

In establishing genetic language families, linguists compare every language with some other language or a group of languages. In fact, one can establish a typology of languages based on the quantitative data received only after comparing languages. The quantitative data give a clearer vision of the differences and similarities between languages. The quantitative load of particular language phenomena is different in different languages. One can notice 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. So, in linguistics qualitative changes are asserted with the help of quantitative factors (**Tambovtsev, 1977; 1994-a; 1994-b; 1998; 1999; 2001-c; 2001-d; Tambovtsev et al., 2007**).

In our studies of the frequency of occurrence of labial consonants we follow the principles of Zipf's dynamic philology. We compare the speech sound chains which are viewed as a succession of articulatory gestures (**Zipf, 1935**: 14). He was one of the first to calculate and compare the frequency of occurrence of different speech sounds in different languages (75; 318). Unfortunately, G.K. Zipf didn't compare the languages by the phonological classes or groups across language families. In this paper we attempt to make comparisons on the basis of the frequency of occurrence of labial consonants across language groups and families.

There are two types of labial consonants: bi-labial and labia-dental (**Zinder, 1979**: 153 – 156). However, for our study it is better to include them into one group because not every

world language has both types (**Tambovtsev, 2001-a; 2001-b; 2001-c**). It is done to maintain the principle of commensurability which allows us to compare only commensurable data (**Tambovtsev et al., 2007**).

It is rather easy to detect the labial consonants in world languages as the majority have labial consonants (**Shirokov, 1985**: 30 – 34; **Zinder, 1979**: 153).

Labial consonants resonator longer. Thus, labial consonants have some special acoustical colouring (**Tambovtsev, 1998, 1999**). In accordance with the table of L.V. Shcherba, which registers all possible labial consonants which man can pronounce in principle, there can occur only 12 types of labial consonants in a human language. It predicts some possible labial consonants which so far have not been found in any world language. However, the usual set of the labial consonants is much more limited. The most common labial consonants across the world languages are: [b, p, m, w, f, v]. These labial consonants are quite universal (**Tambovtsev, 2001-a; 2001-b; 2001-c; Zinder, 1979**: 151 - 152). A comprehensive list of labial consonants may be found in the book of Ian Maddieson who collected and compared the data of the phonological systems of 317 world languages (**Maddieson, 1981**). Unfortunately, he didn't count the frequency of occurrence of sounds in the sound chains in texts; therefore his books do not provide the frequency of occurrence of sounds in the texts in these 317 world languages.

We were able compute the frequency of occurrence of sounds in 258 world languages. We learned that the most widely spread in these languages are the same six labial consonants [p, b, m, w, f, v]. We can call them the basic labial consonants since they occur in most world languages. Maddieson calls them modal (**Maddieson, 1980**). Our research showed that the frequency of these consonants is different in different languages (**Tambovtsev, 1977; 1991; 2001-a; 2001-b; 2001-c; Tambovtsev et al., 2007**). The frequency of occurrence of the labial

consonants is given in per cent to the frequency of occurrence of all the elements in the speech sound chain of the language.

Our data on the frequency of occurrence allows us to detect which of them are marked and which unmarked. Unlike N.S. Trubetzkoy or R. Jakobson, V. A. Nikonov interpreted this opposition as frequent, i.e. unmarked, and infrequent, i.e. marked (**Nikonov, 1963**). Unfortunately, we cannot compare Nikonov's data with ours directly, since his sample volumes are too small. They may provide false results if they happen to be at the beginning or the end of the confidence interval. Our data are much more reliable than Nikonov's but it is possible to watch the common tendencies developing in the Finno-Ugric, Turkic, Caucasian, Indo-European and other language taxa. Our data in every language have greater sample volumes which make the confident interval narrower, thus increasing the reliability of the linguistic conclusions (**Tambovtsev, 1984, 1992, 1992a; 1993, 1998, 1999**).

The data on the frequency of occurrence of the labia consonants were achieved by computing the texts of different language. In order to make number of occurrences commensurable we calculated the percentage of the frequency of the occurrence of the labial consonants to all the phonemes in the speech chain. In this way we obtsained the sound picture of every language under research. In this paper we consider only the frequency of occurrence of the group of labial consonants. The data are provided in the tables (Tab. 1- 24).

Labial Consonants in Niger-Congo and Other Language Taxa

It was noticed long ago that different speech sounds and their groups occur in the speech sound chain with different frequency. The frequency of occurrence of speech sounds can characterize the language. However, until now it is not fully explained why some languages use many speech sounds of a particular sort, for instance, labial consonants, while some other languages hardly use them.

George Kingsley Zipf was one of the first to study this phenomenon on the material of different languages. He explained it by the influence of biology and psychology. The fact that the occurrence of phonemes in the speech chain has its own dynamics allowed him to call this new branch of linguistic investigations by the term "Dynamic Philology" (**Zipf, 1935: XIV**). He was one of the first linguists who inquired about the phenomenon of occurrence of particular speech sounds in the speech chain of the world languages in general.

Some of his data still hold, but the problem with his studies in general is that his samples were too small, thus statistically unstable. Nevertheless, his approach showed some interesting results. One can see from our data that the counts of the frequency of occurrence of speech sounds may be different on small and large samples. In investigating world languages with the help of the methods of dynamic philology, one should bear in mind the simple rule of mathematical statistics: the greater the sample, the more reliable the results (**Tambovtsev, 2003**). We were able to observe it for the first time on different sample volumes of the Mansi (Vogul) language (**Tambovtsev, 1977**).

Let us consider the value of occurrence of all the labial consonants as one group in every language taxon (Tab. 1 – 22).

We can take first any language family. For instance, we can begin with the data of the Niger-Congo family (Tab.1). It is possible to see that the Niger-Congo languages taken for the investigation have a great concentration of labial consonants in their speech chain. They range from 13.00% in Tsewana to 18.23% in Bemba. The mean value in the Niger-Congo is quite high, 15.26%. It is very important to calculate the mean of the frequency of occurrence of the labial consonants since later we'll compare the means of different language taxa with the help of the T-criterion, often called the T-test. We will discuss this in greater detail further (see: Method of the research). However, even without the use of the criteria of mathematical statistics, we can consider the values of the means in different language taxa (Tab. 25). It

shows that the distribution of the labial consonants in the language taxa is quite different. Some taxa have a small value of the mean, while others, like the Niger-Congo taxon, have the great value.

We can compare the characteristics of the the Niger-Congo taxon to any other taxon of world languages since our data are commensurable because the frequency of occurrence of the labial consonants is measured in per cent to all the speech sounds in the text (**Tambovtsev, 2003**). Thus, first of all, we can compare the characteristics of the Niger-Congo to the languages spoken on the same continent of Africa, i.e. the Afro-Asiatic languages.

The Finno-Ugric language family is considered one taxon, although it may be divided into several subgroups:1) Ugric; 2)Permian; 3)Volga and 4)Finnic. We computed 20 languages and dialects of the Finno-Ugric family (Tab. 3). It is very important to compare the whole Finno-Ugric family as one taxon in order to take into account all the differences of distribution of labial consonant in all its languages. There are different points of view on the languages and dialects of the Finno-Ugric family. It is not our task to go into the details of dialects and languages of the Finno-Ugric family. We'll consider the usual set of languages called Finno-Ugric as it is accepted now (**LWUL, 1993**). However, some of their dialects can be called separate languages since their differences on the phonetic and grammar levels are too great. For instance, the Konda dialect and the Sos'va (Northern) dialect of the Mansi (Vogul) language should be rather considered separate languages (**Tambovtsev, 2003**). The Saami (Lopari) language is, in fact, not a united language but a set of different dialects. G. M. Kert finds at least 3 sharply different sets of dialect, while E. Lagercrantz defines 29 dialects (**JNSFUS, 1966**: 155). The least concentration of labial consonants, 8.66%, is found in the L'udikov dialect of Karelian, while the maximum, 14.44%, in Saami (Lopari). The mean for all the 20 Finno-Ugric languages is equal to 11.22% (Tab. 3).

As we can see, this value is much greater than that for labial consonants in the 26 Turkic languages (cf. 11.22% and 8.71%). The minimum, 5.98%, is found in the Altai-Kizhi language, and maximum, 12.80%, in Karakalpak, as one can see from Tab. 3. The data of this table permit us to state that the Turkic languages have a lesser concentration of the labial consonants than the Finno-Ugric languages (Tab. 5).

This value (8.71%) is less than the mean occurrence in the world languages (10.51%). It allows us to speak about the depression of the labial consonants in the Turkic language taxon. From the point of view of markedness labials in the Turkic languages must be considered more marked than in the Finno-Ugric, Samoyedic, Slavonic and some other language taxa.

We can also consider the data in some other language taxa. For instance, the Paleo-Asiatic language family Itelmen has the least frequency of the occurrence of the labial consonants in the sound chain – 6.43%. Kor'akian has the maximum – 10.00%. The mean is 7.93% (Tab. 6).

When we want to compare the means of labial consonants in different language taxa, we must be sure that they are not too dispersed. The degree of dispersion, i.e. the degree of stability, is a very important feature of a language taxon. We can hardly talk of a set of languages as a language taxon, if its stability is poor, i.e. the dispersion is too great. We can measure the degree of dispersion by the confidence intervals, the coefficient of variation and the Chi-square test.

The lower their values, the more stable their distribution in the speech chain. In the other words, the more similar the distribution in the languages under investigation, the less the value of these two statistical criteria is.

The confidence interval (under the significance level of 0.05 or 5%) in the Finno-Ugric language family is 0.67, but in the Turkic taxon it is greater - 0.98. The values of the confidence interval are correlated with the values of the coefficient of variance: 15.04% in

Finno-Ugric and 18.94% in Turkic. So, one can see that the coefficient of variance just verifies the figures of the confidence coefficient. This is why, to comprehend the dispersion of any language taxon, it is quite sufficient to consider either the confidence interval or the coefficient of variance. Perhaps it is easier and more convenient to calculate just the coefficient of variance. Thus, further, we will provide the data just for the coefficient of variance. It can indicate the fluctuation of the values of the dispersion of the labials in different language taxa (see Tab. 25 – 26).

In fact, V. A. Nikonov was one of the first researchers who dealt with the languages of Asia and Africa. He discovered that the labials function differently in the languages in different geographical parts of the world. He claimed that some languages in some parts of Africa exploit labial consonants too much (i.e. overexploit them), while some languages in Asia exploit the labials too little (i.e. underexploit them). V. A. Nikonov called this phenomenon a depression of the labial consonants. It spreads from Central Asia to the West. Nikonov pointed out that the maximum of the frequency of occurrence of the labial consonants is found in the languages of Africa, especially the Bantu languages where they may comprise up to 17% - 18% of the sound chain (**Nikonov, 1976**: 42). Unfortunately, he studied only two Bantu languages.

Our data also showed the tendency of over-exploiting labials in the 12 Niger-Congo languages taken for this study. In fact, according to our computations of the Niger-Congo languages of Africa Nikonov's data are not true in details but verify the general tendency that Niger-Congo languages may be called labial since they use too many labial consonants.

In the other African languages, e.g. Afro-Asiatic family, labials comprise 13.02% of all the phonemes in the speech sound chain. In Hebrew it is a little bit more – 13.69%, though it is an Afro-Asiatic (Semitic) language; in Arabic which is also an Afro-Asiatic (Hamitic) language the labials comprise 13.42%. It is less but still great enough in Hausa – 10.79%.

Thus, we can see that Nikonov's estimation for the Niger-Congo languages and some other language taxa is either incorrect or not exact, i.e. too high or too low. In fact, we provide a more reliable data on the labial consonants (see Tab.1-23).

Nevertheless, it is possible to see that Nokonov was correct to point out that some languages of Asia depress the use of the labial consonants. We found out that Mongolic and Paleo-Asiatic languages depress the use of labials (Tab.24). The mean frequency of occurrence of labials in their speech sound chains is only 7.28% (Mongolic)and 7.93% (Paleo-Asiatic).

As a matter of fact, after computing some American Indian languages we found another pole of depression of the labial consonants. The use of labials in some American Indian languages is much less than in Mongolic or Paleo-Asiatic languages. So, in Haida the frequency of occurrence is 1.70%, in Oneida – 2.40% in Wichita – 2.67%, in Owekeno – 4.30%, in Tonkawa – 4.66%. We can conclude that the labial depression in these American Indian languages is several times greater than the Asiatic depression. Nikonov's data on labial depression depict a lesser labial depression, i.e., only 5% in Aleut and 6% in Itel'men. Neither of the Turkic, Mongolian, nor Sino-Tibetan language according to Nikonov has the frequency greater than 10% (Nikonov, 1976: 42). It does not seem to be quite so. We found out that in Chuvash it is 10.10%; in Turkmen – 10.11%; in Turkish – 10.41% and the maximum is in Karakalpak – 12.80%. The minimum for the 26 Turkic languages is in Altai-Kizhi, in which labial consonants make up only 5.98% of its speech chain. The mean frequency of occurrence for these 26 languages is 8.71%. Later we'll see if Turkic languages differ by its mean from the other taxa of the world languages.

Nikonov's statement holds for the Mongolic languages (c.f. 6,65%, 7,52% and 7,67%) (Tab. 6). Our data on the Paleo-Asiatic languages do not coincide with those of Nikonov. So, in Itel'men we received 6.43% (Nikonov – 6.00%), which is not close enough (Tab. 8).

We have to point out that Polish has a great concentration of labials – 16.66%, though it is a Slavonic, but not a Niger-Congo language (Tab. 12).

We can see that the conclusions of Nikonov are verified in principle. Our data certified this tendency in the sound chains of Turkic and Mongolic languages. The data on the Tungus-Manchurian languages do not go over the value of 12.46% and do not go under 8.53% (Tab. 7), which are close to the limits indicated by V. A. Nikonov.

Study Methods

We have sought to compare functioning of the labial consonants as a group in different language taxa. It is important to choose a mathematical statistics criterion. Discussing the comparison of two language samples Gustav Herdan proposes to use the simplest statistical criteria like standard error test or Chi-square test (**Herdan, 1966**: 35 – 36). However, the standard error test may be too rough. Chi-square test may be no good in this case either since it usually requires the same number of members in a group or in the language taxa. In this study we have different number of languages in different language taxa. Therefore, the most suitable in this case may be t-test. First of all, it does not give the rough but exact estimation. Secondly, the number of languages in language taxa may be different. As it was mentioned earlier, every language taxa has its own mean of the occurrence of the labial consonants in speech. It is possible to state with the help of the t-test if two means are statistically the same or different (**Tambovtsev, 2003**: 22 - 23). In our case, t-test can show if the labial consonants are functioning in different language in the same way or differently.

T-test is also recommended for its robustness. If a statistical test is robust, then it means that it is fairly tolerant of all but rather large deviations from normality and equality of variance. However, we agree with Christopher Butler who points out that before using the t-test a rough check should be made to ensure that the variation of the data of a language taxa is not too great (**Butler, 1985**: 84). We check it by the coefficient of variance. Usually, it is

possible to use the t-test if the value of the coefficient of variance is less than 33%. One can see from the formula of the t-test why it is so important that the dispersion is not too great:

$$T = \frac{M1 - M2}{\sqrt{SI1 / n1 + SI2 / n2}}$$

Where: M1 – the mean of the frequency of occurrence of labial consonants in the first language taxon

M2 – the mean of the frequency of occurrence of labial consonants in the second language taxon

S^2 1 – the value of the first dispersion, i.e. standard squared in the first language taxon

S^2 2 – the value of the second dispersion, i.e. standard squared in the second language taxon

n1 – sample volume of the first language taxon

n2 – sample volume of the second language taxon

Therefore, if the variability in one or both language taxa is too great, then the value of t-text may be small enough to show no difference between the two language taxa in question. So, it is advisable to consider the confidence interval (**Tambovtsev, 2003**: 19 – 21). It is also possible to understand if the variability is too great with the help of the value of the coefficient of variance which should not be greater than 33% (**Tambovtsev, 2003**: 11 - 16). We provide the coefficient of variance in every table (Tab. 1 – 22).

Let us consider the confidence interval for its mean. By the confidence interval as well as by the coefficient of variance we measure the stability of the frequency of occurrence of labial consonants in the sound chain. If in one language it is greater than in the other, then we must say that its stability is less. We showed in detail how to calculate the confidence interval (Tambovtsev, 1994: 37-39). In the case of Niger-Congo languages we must multiply the standard deviation by the confidence value of the table which depends of the sample volume. It is here equal to 0.70456. Thus, 2.18 x 0.70456 = 1.53. Let us compare it to the confidence intervals of the other language taxa in order to see if it is great or small. In the case of the Tungus-Manchurian languages it is 1.47, which is greater than in the taxon of Finno-Ugric languages (0.67) or Turkic languages (0.98). This means that Tungus-Manchurian languages

are more dispersed by the use of the labial consonants. We measure the confidence interval at the significance level of 5% (**Tambovtsev, 2003**: 20).

However, to calculate the coefficient of variance is easier. We must multiply the value of the standard deviation by 100% and divide it by the mean. Therefore, here it is equal to (1.82 x 100%) divided by 15.26. The Niger-Congo taxon is 11.92%. It is less than in the taxon of the Afro-Asiatic taxon which is 18.89%. It is also less than in the Finno-Ugric languages. The coefficient of variance of the Finno-Ugric taxon is 14.97%. It is less than in Turkic (18.94%), but greater than in the Finno-Ugric taxon.

The coefficient of variance helps us to keep to the principle of commensurability because it allows us to compare the changes of different sorts. In fact, the coefficient of variation is the mean of the dispersion in per cent. It shows the variability: the greater the variability, the greater the coefficient of variance. If the value of this coefficient is greater than 33%, then the variation may be called critical (**Tambovtsev, 2003**: 11 - 14).

It is very important to know the number of the degrees of freedom. In this case, it is equal to N1 + N2 – 2, where N is the number of the languages in the first group and N2 – in the second group. If the calculated value of t is greater than or equal to the critical value as determined from the table, then we must reject the hypothesis that these two means are statistically the same (**Tambovtsev, 2003; Tambovtsev et al., 2007**). The critical values can be found in any book on statistics (e.g. **Butler, 1985**: 172).

It is advisable to provide the example of the calculation of the t-test for the family of the Finno-Ugric and the family of Turkic languages. The actual data on the frequency of occurrence of the labial consonants and the other phonemes may be taken elsewhere (Tambovtsev, 2001-b). So, we take 20 Finno-Ugric languages (Tab.1) and 26 Turkic languages (Tab. 3). During their historical development many Finno-Ugric languages were in contact with the Turkic languages. Now we would like to know if they influenced each other

so much that their data on labials are statistically the same. In the other words, we are trying to check if the typology of the distribution of labials is similar enough. We can put forward the hypothesis that the difference between their means is not statistically significant. We must put the data that we received in the formula provided above. The mean for the Niger-Congo labial consonants is 15.26; $S^2 = 3.31$ (Tab.1). The mean for the Afro-Asiatic family is 11.54, $S^2 = 4.75$. Now we must divide every S squared by the number of the languages in the family. For the Niger-Congo family we obtain: 3.31 / 12 = 0.28 and for the Afro-Asiatic: 4.75 / 7 = 0.68. Putting these data in the formula we obtain:

$$T = \frac{15.26 - 11.54}{\sqrt{0.28 + 0.68}} = \frac{3.72}{\sqrt{0.96}} = \frac{3.72}{0.98} = 3.80$$

Now we must calculate the number of the degrees of freedom, 12 + 7 – 2 = 17.

We can see from the table of the critical values that at the significance level of 0.05 the critical value is 2.1098 (**Butler, 1985**: 172). This critical value is much less than the obtained value. It means that the means are too different. We'd like to devise a sort of distance between these two means. So, we divide the obtained value by the critical value. We call it the TMB coefficient which can show us how much the language taxa are different by some chosen parameter.

In this way we can calculate the distance between the Niger-Congo taxon and that of the Finno-Ugric languages. The TMB distance between them (3.05) is greater than the distance between the Niger-Congo and the Afro-Asiatic languages (1.80).

We can see that the Niger-Congo mean is different from the Turkic mean. Here, TMB = 5.21. In the same way we can calculate the distance between the Niger-Congo taxon and any other language taxon. In order to find out if the value TMB is too great or too small, we can compare its values to those between the other language taxa. Let us calculate the distance between the Finno-Ugric and the Samoyedic taxa by the use of the labial consonants (Tab.3

and Tab.4). After the calculations by the same formula, we receive the distance between the Finno-Ugric and the Samoyedic families, at TMB = 0.35. It is much less than one unit, which indicates that there is no statistical difference between the distributions of the labials in both language families. Slavonic languages (Tab. 10) are typologically much more far away from the Finno-Ugric languages than the Samoyedic ones with the TMB = 1.954. The Mongolic language family shows a greater distance than that, with TMB = 3.827. At the same time the Mongolic languages show that they are closer to the Turkic languages (TMB = 1.540) than to the Finno-Ugric ones by the distribution of labials. In this way, one can calculate the typological distances between different language taxa: subgroups, groups, etc. Later we will discuss the distances further in more detail. Here, we just demonstrated the method of calculations of the similarity between the language taxa in principle.

However, before discussing the results obtained by the t-test, we must again place our attention on the fact that the dispersion of every language taxon must not be too great. Let us compare these dispersions across the language taxa.

Density and Dispersion in Distribution of Labial Consonants

One can notice that different language taxa have different dispersion of the labial consonants (Tab. 25 -26). The occurrence of the labial consonants can characterize this or that language taxon. On the other hand, the dispersion of the labials in a taxon, can characterize if this taxon is a natural classification of typologically close language or a mere conglomeration of languages constructed by some other criteria, for instance by the geographical principle. It may also unite the languages which are genetically or typologically close.

The compactness of the Niger-Congo language taxon is great. The coefficient of variance is only 11.96. One must remember that the coefficient of variance shows the dispersion of a language taxon. Therefore, the less it is, the more compact the taxon. Niger-Congo taxon takes the second place after the most compact Mongolic taxon (Tab.25).

If we take the Indo-European language family, then we obtain the following statistical characteristics: the mean – 11.84%, the confidence interval – 0.49. The value of the coefficient of variance (14.66%) indicates to the stable distribution of labials. At least, the labial distribution in this case is more stable than in the Finno-Ugric (15.04%), Tungus-Manchurian (17.59%), Paleo-Asiatic (18.61%) or Turkic (18.94%) family. On the other hand, Indo-European family is more dispersed than the Mongolic (7.55%) family (Tab. 25).

Now let us consider the dispersion of different groups of the Indo-European family (Tab. 26). The most stable (i.e. compact) Indo-European group is Indic (6.85%), the least compact – Baltic (16.00%). Therefore, we can unite all the Indic languages into one group since they are all situated in one geographical region. The typology of the distribution in Germanic (9.65%) and Slavonic (10.34%) groups is rather stable.

In the 128 languages which we took for our studies, the frequencies of occurrence of the labial consonants are spread in the range from 1.70% to 16.66%. The distribution of the labials are homogeneous (TMB = 0.41). It is far from one unit.

The form of the distribution is in good accordance with the theoretical normal distribution: at the 0.05 level of significance with the 6 degrees of freedom TMB = 0.28. It means that there are few languages which greatly underexploit or overexploit the use of labials in the speech chain.

All things are known in comparison. Thus it is necessary to analyse the behaviour of the labial coefficients of variance in an ordinary text in a language. Let us calculate the values of the coefficient of variance in several languages in a coherent text to see the typology. We took the text of the languages of different families: English (12.08%); Japanese (12.91%); Finnish (13.18%); Russian (14.59%); Gypsy (14.95%); Mangarayi (18.32%).

One can see that the values of the coefficient of variance of the labial consonants in the coherent text is more or less the same as across the languages (Tab. 25 – 26).

Measuring Similarity across Language Taxa

After calculating the similarity between the languages in different language taxa, we obtained the following results for the Turkic family (Tab. 27). It turns out to be close to the taxon of the American Indian languages. It may be merely by chance since our error level is 5%. So, our results may happen to fall into the error gap. However, it may not be by chance since our reliability is 95%. Thus, we are apt to conclude that it is not by chance: there is some basic linguistic fundamental for it. The similarity between the languages in question may be caused by genetic relatedness. However, there is the other possibility. It may be, of course, purely typological, i.e. different unrelated languages developed some most convenient articulatory trends. Then, there arises a question: why for these languages are these articulatory trends most convenient? It may mean that their articulatory habits are rather similar. Why are their articulatory habits similar if they are not genetically related? So, common articulatory trends may give rise to important questions, which are usually easily answered if the languages are genetically related. In fact, nothing prevents a language from constructing words which consist only of the labial consonants in combination with different vowels. Let us take only the most common vowels which occur in most languages (**Tambovtsev, 2001a**; 2001b): [a, o, u, e, i]. In this hypothetical language there may be only such words as “ba, bo, be, bi, bu, baba, bibi, bebe, bobo, bubu, papa, pepe, pipi, papu, muma, mama, meme, mimi, wawa, wowo, wewe, wiwi, etc”. It is possible to construct many words with the labials and vowels, especially if the words get longer: “babobibebu”, “bobababebi” or more complex like “bamopefi, popamamobabo, etc”. However, it is not possible to find a natural human language which resorts only to the use of labial consonants.

The application of t-test allowed us to measure the distances between the Niger-Congo taxon and the other language taxa:

1) 0.67 (Iranian group of Indo-European family);
2) 1.24 (Samoedic);

3) 1.37 (Slavonic group of Indo-European family);
4) 1.80 (Afro-Asiatic);
5) 2.10 (Sino-Tibetan);
6) 2.20 (Caucasian);
7) 2.41 (Australian Aboriginal);
8) 2.43 (Romance group of Indo-European family);
9) 2.80 (Germanic group of Indo-European family);
10) 2.98 (Manch-Tungussic);
11) 3.05 (Finno-Ugric);
12) 3.15 (Austronesian);
13) 3.75 (Indic group of Indo-European family);
14) 3.94 (American Indian);
15) 4.05 (Paleo-Asiatic).
16) 5.21 (Turkic)
17) 5.96 (Mongolic).

Explaining the close distances between Turkic and American Indian languages we must recall the original hypothesis put forward by an unknown Catholic monk and then picked up by the great mathematician Gottfried Wilhelm Leibniz (1646 - 1716). In Russia it was developed by an outstanding archaeologist A. P. Okladnikov. Actually, in 1938 he published an article in which he claimed that the people in the Americas originated from the peoples composed of Siberian tribes. According to his ideas the Neolithic people from Siberia migrated to the most Northeastern point of Siberia. There they found the Bering ice bridge which allowed them to cross to Alaska in Northern America (**Okladnikov, 1938**: 224). However, according to his theory the Neolithic peoples who used to live on the banks of the Angara and Lena Rivers and the Baikal Lake first moved towards the East and got to the shores of the Pacific Ocean (et al., 1976: 12 - 67). I should guess part of these peoples moved eastward to the Japanese Islands. Perhaps the ancient Ainu were in their number. Then the other Neolithic tribes who were related to the Siberian peoples moved farther and got to South America but preserved their articulation basis. This may account for the typological similarity in the distribution of the consonantal groups in the Turkic and American Indian languages. We must point out to the fact that the articulatory basis usually is preserved even

when the people begin to talk in the other language. This is called the effect of the substratum.

A. P. Okladnikov points out that the anthropological features of American Indians and Siberian peoples are similar. The other strong point in Okladnikov's reasoning is that in South and North America there never were any apes or monkeys from whom people may have developed. Actually, many animals from Siberia also crossed this Bering ice bridge to North America. This is why not only people but also the animals in Siberia and America are the same. In fact, the Bering ice bridge existed twice.

First, it was some 65 - 35 thousand years ago, and then some 28-25 thousand years ago. It is supposed that each period during which it existed was not less than 18 - 15 thousand years. At least some 19,000 years ago it existed. A. P. Okladnikov believed that the Americas were inhabited in two waves, i.e., in the middle and upper Palaeolithic period (**Okladnikov, 1938**; **Okladnikov et al., 1976**). Our data support this theory. From the typological point of view, some American Indian languages (cf. Tab. 4) are also very close to the Paleo-Asiatic languages. We cannot state that the Turkic language family is close to any language taxon. So, Turkic language family is not typologically close to the Iranian (TTM = 3.636) or Slavonic (TTM = 4.440) languages of the Indo-European family (cf. Tab. 28). It is close enough to the languages of the Tungus-Manchurian family. However, this may be easily explained. Probably, Tungus-Manchurian family is closer to the Turkic family because during their historical development they had contacts. Some linguists (e.g. V.M. Illich-Svitych, E.D. Polivanov, N. Poppe, G.J. Ramstedt, etc.) believe them to be so close that they comprise a taxon of the Altaic languages which include Turkic, Mongolic and Tungus-Manchurian languages. Other linguists (e.g. V. Kotvich, A.M. Shcherbak, E.A. Potseluevskij, B.A. Serebrennikov, etc.) vigorously oppose the view that these three language families are genetically related and should be united in one language family since it is impossible to prove

reliable phonetic and lexical similarity. The details of this discussion can be found elsewhere (Tambovtsev, 2001-b: 56). We support the third group of linguists who think that it is not possible to prove if some phonetic and lexical similarities are due to their genetic relatedness or arose due to the long intensive contacts between them (e.g. A. N. Kononov). Let us point out to the fact that for the typological study it does not matter much why or how this similarity arose, the main problem is if there is a statistically significant or insignificant similarity. Our study may show the reliable of this or that similarity, if any (Tambovtsev, 2001-b: 56 – 57).

One can find more details on the typological distances between Turkic family and the other world language taxa in tables (Tab. 28 - 29). Though it is possible to state a great typological closeness between Turkic and some American Indian languages, we are far from stating that genetically they are close. However, from the point of view of typology Turkic family is very similar to the American Indian languages under study. Having this typological clue, linguists may have a closer look at them from the genetic point of view.

The distances between the Slavonic group of the Indo-European family can be seen from Tab. 29. The Iranian group is the closest to the Slavonic languages (0.019). S.V. Bromley and others claim that the Slavonic tribes came into contact with the Iranian speaking tribes of the Sarmats in the 8^{th} – 9^{th} centuries to the south of the Oka river. The details of the discussion can be found elsewhere **Tambovtsev, 2001-a**: 69). The Baltic languages are the next closest (0.349) to the Slavonic group. Many linguists (S.B. Bernshtein, P.S. Kuznetsov, O.S. Shirokov, etc.) believe that there was a sort of the Balto-Slavonic language community (Tambovtsev, 2001: 70 – 71). So, one can see that long contacts between the Slavonic and Turkic peoples did not influence their articulatory basis, i.e. the articulatory habits concerning the labial consonants were not borrowed. Therefore the distribution of labials is so different in the Slavonic and Turkic languages.

It is interesting to analyse if the groups of languages which enter the Turkic language family have similar distributions of the labials. Let us consider the Oguz, Kypchak, Karluk and Siberian Turkic groups of the Turkic family defined by N.A. Baskakov. It was discussed elsewhere that Baskakov's classification is one of the 17 classifications of the Turkic languages created by now. We use it because it is the most popular (Tambovtsev, 2001b: 60 – 61).

Comparing the distances between the Ugric and the other two groups of the Finno-Ugric family one can see that the Ugric and Permic groups of the Finno-Ugric (TTM=0.041) are the closest. So, those linguists who constructed the Ugric-Permic language community (Budez, Haidu, Moor, Redei) were correct (Tambovtsev, 2001).

Conclusions

1. The frequency of occurrence of labial consonants in the 12 Niger-Congo languages make us believe that this language taxon may be called labial, as their labials function in the range of 13% – 1 18.23%, with a mean of 15.26%. In fact, the world languages taken for this study demonstrate that they are distributed in the limit from 1.70% to 18.23%. The mean is 10.51%. We can state that the languages which employ them with less frequency under-exploit the labial consonants while those which employ them with greater frequency over-exploit them in their speech chains.

2. The Niger-Congo language is quite compact. Its coefficient of variance is 11.96. The least dispersed language taxon is the Mongolic family (V = 7.55). This means that the languages of the Niger-Congo taxon are very typologically close. The American Indian languages are very dispersed (V = 44.09), which indicates that their speech sound chains are rather different by structure.

3. The use of the t-test can demonstrate the similar peculiarities of the distribution of the labial consonants in different languages taxa. It is possible to construct the typological

distances between different language taxa. For instance, the distribution of the labial consonants in the speech chain of the Niger-Congo taxon is very similar to that of the Iranian group of the Indo-European languages (0.67). The depression of labials in the taxon of Mongolic family makes the distance the greatest – 5.96. The Turkic language taxon is not similar to the Niger-Congo taxon, but is very similar to that of the American Indian languages.

4. Surely, we could not embrace all the languages of the world, but our sample of 200 languages is great enough to state that the tendencies that we found are true for any human language. The statistical investigation of the functioning of the labial consonants in the speech sound chains of world languages gives good clues to understanding how human language works.

Tables

Tab. 1 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Niger-Congo Languages (% to all phonemes).

#	Language	%		#	Language	%
1.	Tsewana	13.00		7.	Luganda	15.15
2.	Wolof	13.02		8.	Fulfulde	16.56
3.	Hanga	13.11		9.	Swahili	16.61
4.	Xhosa	13.60		10.	Tonga	17.04
5.	Chichewa	14.82		11.	Moore	17.07
6.	Kinyarwanda	14.88		12.	Bemba	18.23
	Statistics data:					
	Mean	15.26			S^2	3.31
	S	1.82			V%	11.96

Tab. 2 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Afro-Asiatic Languages, i.e. Semito-Hamitic Language Family (% to all phonemes).

#	Language	%		#	Language	%
1.	Somalian	7.62		5.	Hebrew	13.34
2.	Hausa	9.93		6.	Assirian	13.39
3.	Sokotrian	11.18		7.	Arabic	13.42
4.	Neo-Aramaic	11.92				
	Statistics data					
	Mean	11.54			S^2	4.75
	S	2.18			V %	18.89

Tab. 3 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Finno-Ugric Languages (% to all phonemes).

#	Language	%		#	Language	%
1.	Karelian-L'udik	8.66		11.	Komi-Permian	11.15
2.	Finnish	8.73		12.	Karelian-Livvik	11.16
3.	Mari-Lawn	9.47		13.	Mordovian-Moksha	11.26
4.	Karelian-Tihvin	9.66		14.	Vodian	11.95
5.	Mari-Mountain	9.99		15.	Mansi (Konda)	12.29
6.	Hungarian	10.04		16.	Hanty (Kazym)	12.60
7.	Estonian	10.21		17.	Mansi (Northern)	13.56
8.	Komi-Zyrian	10.21		18.	Udmurt	13.66
9.	Hanty (Eastern)	10.45		19.	Mordovian-Erzia	13.72
10.	Vepsian	11.11		20.	Saami	14.44
	Statistics data:					
	Mean	11.22			S^2	2.82
	S	1.68			V %	14.97

Tab. 4 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Samoyedic Languages, i.e. Samoyedic Language Family (% to all phonemes).

#	Language	%		#	Language	%
1.	Nganasan	7.71		3.	Nenets	12.14
2.	Sel'kup	11.99		4.	Kamasin	13.99
	Statistics data					
	Mean	11.46			SI	7.08
	S	2.66			V %	23.21

Tab. 5 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Turkic Languages, i.e. Turkic Language Family (% to all phonemes).

#	Language	%		#	Language	%
1.	Altai - Kizhi	5.98		14.	Karacha-Balkar	8.76
2.	Jakut	6.10		15.	Tatar-Baraba	9.04
3.	Shorian	6.33		16.	Salar	9.17
4.	Tofalar	6.50		17.	Tuvin	9.30
5.	Hakas	7.40		18.	Uzbek	9.42
6.	Sary-Ujgur	7.51		19.	Ujgur	9.65
7.	Altai -Chalkan	7.87		20.	Azeri	9.66
8.	Kazah	7.99		21.	Tatar-Krym	9.79
9.	Tatar-Kazan	8.03		22.	Chuvash	10.10
10.	Dolgan	8.43		23.	Turkmen	10.11
11.	Kirgiz	8.43		24.	Turkish	10.41
12.	Bashkir	8.54		25.	Tatar-Chulym	11.03
13.	Kumandin	8.69		26.	Karakalpak	12.80
	Statistics data:					
	Mean	8.71			S^2	2.72
	S	1.65			V %	18.94

Tab. 6 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Mongolic Languages, i.e. Mongolic Language Family (% to all phonemes).

#	Language	%		#	Language	%
1.	Kalmyk	6.65		3.	Buriat	7.67
2.	Mongolic	7.52				
	Statistics data:					
	Mean	7.28			S^2	0.30
	S	0.55			V %	7.55

Tab. 7 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Tungus-Manchurian Languages, i.e. Tungus-Manchurian Language Family (% to all phonemes).

#	Language	%		#	Language	%
1.	Even (Lamut)	8.34		5.	Orokian	10.38
2.	Negidal	8.53		6.	Orochian	10.47
3.	Evenk (Tungus)	8.73		7.	Ul'chian	12.46
4.	Udyge	8.74		8.	Manchurian	13.31

5.	Nanai	10.15				
	Statistics data:					
	Mean	10.12			S^2	3.17
	S	1.78			V %	17.59

Tab. 8 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Paleo-Asiatic Languages, i.e. Paleo-Asiatic Language Family (% to all phonemes).

#	Language	%		#	Language	%
1.	Itel'men	6.43		4.	Chookchee	8.76
2.	Eskimo-Imaklin	6.72		5.	Koriak	10.00
3.	Eskimo-Naukan	7.76				
	Statistics data					
	Mean	7.93			S^2	2.18
	S	1.48			V %	18.61

Tab. 9 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Indic Languages, i.e. Indic Group of the Indo-European Language Family (% to all phonemes).

#	Language	%		#	Language	%
1.	Marathi	9.51		4.	Gipsy	10.61
2.	Hindi	9.97		5.	Gudjarati	11.35
3.	Bengali	10.06				
	Statistics data					
	Mean	10.30			S^2	0.50
	S	0.71			V %	6.85

Tab. 10 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Iranian Languages, i.e. Iranian Group of the Indo-European Language Family (% to all phonemes).

#	Language	%		#	Language	%
1.	Iranian (Persian)	11.78		5.	Dari (Afganistan)	12.85
2.	Osetian	12.26		6.	Tadjik	13.11
3.	Talysh	12.81		7.	Gilian	15.18
4.	Pashto	12.82		8.	Kurdish	16.25
	Statistics data					
	Mean	13.38			S^2	2.33
	S	1.53			V %	11.40

Tab. 11 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Slavonic Languages, i.e. Slavonic Group of the Indo-European Language Family (% to all phonemes).

#	Language	%		#	Language	%
1.	Macedonian	11.67		7.	Ukranian	13.01
2.	Serbian	11.96		8.	Czech	13.57

3.	Slovenian	12.54		9.	Belorussian	14.45
4.	Russian	12.63		10.	Sorbian	14.83
5.	Slovak	12.79		11.	Polish	16.66
6.	Bulgarian	12.91				
	Statistics data					
	Mean	13.35			S^2	1.90
	S	1.38			V %	10.34

Tab. 12 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Baltic Languages, i.e. the Baltic Group of the Indo-European Language Family (% to all phonemes).

#	Language	%		#	Language	%
1.	Latvian	10.83		2.	Lithuanian	13.63
	Statistics data:					
	Mean	12.25			S^2	3.84
	S	1.96			V %	16.00

Tab. 13 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Romance Languages, i.e. the Romance Group of the Indo-European Family (% to all phonemes).

#	Language	%		#	Language	%
1.	Spanish	9.79		4.	Moldavian	11.06
2.	Rumanian	10.22		5.	Portuguese	11.10
3.	Italian	10.38		6.	French	13.96
	Statistics data					
	Mean	11.08			S^2	2.24
	S	1.50			V %	13.49

Tab. 14 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Germanic Languages, i.e. the Germanic Group of the Indo-European Language Family (% to all phonemes).

#	Language	%		#	Language	%
1.	German	9.88		5.	Danish	11.95
2.	Gothic	10.56		6.	Dutch	12.03
3.	Norwegian	10.60		7.	English	13.05
4.	Swedish	11.00				
	Statistics data					
	Mean	11.30			S^2	1.19
	S	1.09			V %	9.65

Tab. 15 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Caucasian Languages, i.e. Caucasian Language Family (% to all phonemes).

#	Language	%		#	Language	%
1.	Chechenian	7.51		4.	Kabardian	10.70
2.	Abhazian	9.17		5.	Adygian	12.22

3.	Avarian	9.75		6.	Georgian	13.35
	Statistics data					
	Mean	10.45			S^2	4.67
	S	2.16			V %	20.67

Tab. 16 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Sino- Tibetan Languages (% to all phonemes).

#	Language	%		#	Language	%
1.	Dungan	8.22		4.	Thai	12.63
2.	Burmanese	8.79		5.	Tibetan	12.67
3.	Chinese	9.13				
	Statistics data					
	Mean	10.29			S^2	4.75
	S	2.18			V %	21.19

Tab. 17 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Austro - Asiatic Languages (% to all phonemes).

#	Language	%				
1.	Vietnamese	10.07				

Tab. 18 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Austronesian Languages (% to all phonemes).

#	Language	%		#	Language	%
1.	Maori	7.11		5.	Cebuano	9.85
2.	Marquis	7.80		6.	Tagalog	10.50
3.	Hawaian	7.87		7.	Indonesian	11.96
4.	Dajak	8.77		8.	Uma	12.16
	Statistics data					
	Mean	9.50			S^2	3.72
	S	1.93			V %	20.30

Tab. 19 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Australian Aboriginal Languages (% to all phonemes).

#	Language	%		#	Language	%
1.	Ngaanyatjarara	8.42		4.	Djingili	11.35
2.	Ngandi	9.92		5.	Nunggubuyu	12.47
3.	Nyangumada	10.40		6.	Mangarayi	14.51
	Statistics data					
	Mean	11.18			SI	4.54
	S	2.13			V %	19.07

Tab. 20 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the American Indian Languages of North America (% to all phonemes).

#	Language	%		#	Language	%
1.	Haida	1.70		13.	Kawasquar	9.05
2.	Oneida	2.40		14.	Secoya	9.29
3.	Wichita	2.67		15.	Inga	9.89
4.	Navaho	4.15		16.	Cofan	10.02
5.	Owekeno	4.30		17.	Pocomchi	10.83
6.	Tonkawa	4.66		18.	Siriano	11.18
7.	Iquito	4.83		19.	Kechua	11.40
8.	Piratapuyo	6.56		20.	Nahuatl	11.73
9.	Mam	7.33		21.	Sayula populuca	12.34
10.	Totonac	7.38		22.	Kaiwa	12.75
11.	Kadiweu	7.74		23.	Guarani	12.92
12.	Capanahua	8.04				
	Statistics data					
	Mean	7.96			S^2	12.35
	S	3.51			V %	44.09

Tab. 21 The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Isolated Languages (% to all phonemes).

#	Language	%		#	Language	%
1.	Basque	5.72		6.	Armenian	10.32
2.	Japanese	6.94		7.	Greek	10.81
3.	Ket (Yug)	8.36		8.	Yukaghir	11.10
4.	Ainu	9.28		9.	Nivhian	11.34
5.	Korean	10.00		10.	Albanian	12.07
	Statistics data					
	Mean	9.60			S^2	4.14
	S	2.04			V%	21.25

Tab. 22 Minimum Values of the Frequency of Occurrence of the Labial Consonants in the Speech Sound Chain in the Studied Language Taxa, % to all phonemes.

	Language and Taxon	%
1.	Haida (American Indian)	1.70
2.	Basque (Isolated)	5.72
3.	Altai – Kizhi (Turkic)	5.98
4.	Itel'men (Paleo-Asiatic)	6.43
5.	Kalmyk (Mongolic)	6.65
6.	Maori (Austronesian)	7.11
7.	Chechenian (Caucasian)	7.51
8.	Somalian (Afro-Asian)	7.62
9.	Nganasan (Samoedic)	7.71
10.	Dungan (Sino-Tibetan)	8.22
11.	Even-Lamut (Manchu-Tungussic)	8.34
12.	Ngaanyatjarara (Australian Aboriginal)	8.42

13.	Karelian-L'udik (Finno-Ugric)	8.66
14.	Marathi (Indic group of Indo-European family)	9.51
15.	Spanish (Romance group of Indo-European family)	9.79
16.	German (Germanic group of Indo-European family)	9.88
17.	Latvian (Baltic group of Indo-European family)	10.83
18.	Macedonian (Slavonic group of Indo-European family)	11.67
19.	Iranian-Persian (Iranian group of Indo-European family)	11.78
20.	Tsewana (Niger-Congo)	13.00

Tab. 23 Maximum Values of the Frequency of Occurrence of the Labial Consonants in the Speech Sound Chain in the Studied Language Taxa, % to all phonemes.

	Language and Taxon	%
1	Buriat (Mongolic)	7.67
2	Koriak (Paleo-Asiatic)	10.00
3	Gudjarati (Indic group of Indo-European family)	11.35
4	Albanian (Isolated)	12.07
5	Uma (Austronesian)	12.16
6	Tibetan (Sino-Tibetan)	12.67
7	Karakalpak (Turkic)	12.80
8	Guarani (American Indian)	12.92
9	English (Germanic group of Indo-European family)	13.05
10.	Manchurian (Manchu-Tungussic)	13.31
11	Georgian (Caucasian)	13.35
12	Arabic (Afro-Asian)	13.42
13	Lithuanian (Baltic group of Indo-European family)	13.63
14	French (Romance group of Indo-European family)	13.96
15	Kamasin (Samoedic)	13.99
16	Saami (Finno-Ugric)	14.44
17	Mangarayi (Australian Aboriginal)	14.51
18	Kurdish (Iranian group of Indo-European family)	16.25
19	Polish (Slavonic group of Indo-European family)	16.66
20.	Bemba (Niger-Congo)	18.23

Tab. 24 Mean Values of the Frequency of Occurrence of the Labial Consonants in the Speech Sound Chain in Language Families, % to all phonemes.

#	Family	%		#	Family	%
1.	Mongolic	7.28		8.	Caucasian	10.45
2.	Paleo-Asiatic	7.93		9.	Australian aborigin.	11.18
3.	American Indian	7.96		10.	Finno-Ugric	11.19
4.	Turkic	8.71		11.	Samoyedic	11.46
5.	Austronesian	9.50		12.	Afro-Asiatic	11.54
6.	Tungus-Manchurian	10.12		13.	Indo-European	12.22
7.	Sino-Tibetan	10.29		14.	Niger-Congo	15.26

Tab. 25 Mean Values of the Frequency of Occurrence of the Labial Consonants in the Speech Sound Chain in the Language Groups of the Indo-European Family, % to all phonemes.

#	Group	%		#	Group	%
1.	Indic	10.30		4.	Baltic	12.25
2.	Romance	11.08		5.	Slavonic	13.35
3.	Germanic	11.30		6.	Iranian	13.38

Tab. 25 The Coefficient of Variance in Different Language Families (V %).

#	Family	V %		#	Family	V %
1.	Mongolic	7.55		8.	Turkic	18.94
2.	Niger-Congo	11.96		9.	Australian (aborig.)	19.07
3.	Indo-European	14.66		10.	Caucasian	20.67
4.	Finno-Ugric	15.04		11.	Sino-Tibetan	21.19
5.	Tungus-Manchurian	17.59		12.	Samoyedic	23.21
6.	Paleo-Asiatic	18.61		13.	Austronesian	20.30
7.	Afro-Asiatic	18.89		14.	American Indian	44.09

Tab. 26 The Coefficient of Variance in Different Groups of the Indo-European Language Family (V %).

#	Group	V %		#	Group	V %
1.	Indic	6.85		4.	Iranian	11.40
2.	Germanic	9.65		5.	Romanic	13.49
3.	Slavonic	10.34		6.	Baltic	16.00

Tab. 27 Typological Distances between the Turkic Language Family and the other Language Taxa Based on the TTM Coefficient.

#	Language Taxon	TTM		#	Language Taxon	TTM
1.	American Indian	0.466		6.	Indic of group of I.-E.	1.716
2.	Tungus-Manchurian	1.021		7.	Finno-Ugric	2.470
3.	Paleo-Asiatic	1.060		8.	Iranian group of I.-E.	3.636
4.	Mongolic	1.540		9.	Slavonic group of I.-E.	4.440
5.	Afro-Asiatic	1.566				

Tab. 28 Typological Distances between the Slavonic Group of the Indo-European Language Family and the other Language Taxa Based on the TTM Coefficient.

#	Language Taxon	TTM		#	Language Taxon	TTM
1.	Iranian group of I.-E.	0.019		7.	Austronesian	2.353
2.	Baltic group of I.-E.	0.349		8.	Paleo-Asiatic	3.299
3.	Romance group of I. E.	1.467		9.	Turkic	4.440
4.	Germanic group of I.-E.	1.697		10.	Mongolic	5.531
5.	Finno-Ugric	1.954		11.	American Indian	7.505
6.	Tungus-Manchurian	2.161				

Tab. 29 Typological Distances between the Oguz Group of the Turkic Language Family and the other Language Taxa Based on the TTM Coefficient.

#	Language Taxon	TTM		#	Language Taxon	TTM
1.	Karluk group of Turkic	0.68		3.	Siberian group of Turkic	2.496
2.	Kypchak group of Turkic	2.091				

Tab. 30 Typological Distances between the Ugric Group of the Finno-Ugric Language Family and the other Language Taxa Based on the TTM Coefficient.

#	Language Taxon	TTM		#	Language Taxon	TTM
1.	Permic group of F.-U.	0.041		3.	Volgaic group of F.-U.	0.250
2.	Finnic group of F.-U.	0.103				

References

Brainerd, 1974 – Brainerd B. Weighing Evidence in Language and Literature: A Statistical Approach. – Toronto: University of Toronto Press, 1974.

Butler, 1985 – Butler Christopher. Statistics in Linguistics. – Oxford: Basil Blackwell, 1985.

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.

Crystal, 1992 – Crystal, David. An Encyclopedic Dictionary of Language and Languages. – Oxford: Blackwell, 1992.

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

Herdan, 1966 – Herdan G. The Advanced Theory of Language as Choice and Chance. – Berlin: Springer-Verlag.

JNSFUS, 1966 – Jasyki Narodov SSSR. T. 3. Finno-Ugorskie i Samodijskie Jazyki. [The languages of the peoples of the USSR. T. 3. Finno-Ugric and Samoyedic languages]. – Moskva: Nauka Publishing House, 1966.

LWUL, 1993 – Jazyki mira: Ural'skie jazyki. [Languages of the World: Uralic Languages] – Moskva: Nauka, 1993.

Maddieson, 1980 - Ian Maddison. UPSID: UCLA Phonological Segment Database. UCLA Working Papers in Phonetics, 50, August, 1980, p. 4 56.

Maddieson, 1981 - Ian Maddison. UPSID: UCLA Phonological Segment Database. Data and Index. UCLA Working Papers in Phonetics, # 53, November, 1981.

Nikonov, 1963 – Nikonov V. A. Interpretatsija foneticheskih chastot. [The interpretation of phonetical frequencies]. - In: Uchonye zapiski instituta slav'anovedenija. T. 27, 1963, p. 259 - 270.

Nikonov, 1976 – Nikonov V. A. Glottogenez Sibiri I Dal'nego Vostoka v svete fonostatistiki. [Glottogenesis of Siberia and the Far East from the point of view of phonostatistics]. In: Proishozhdenie aborigenov Sibiri I ih jazykov. – Tomsk: Tomsk State Pedagogical University Press, 1976, p. 41 – 46.

Ohotina, 1982 – Ohotina N. V. Jazyki Bantu. [Bantu languages]. – In: Sravnitel'no-Istoricheskoe Izuchenie Jazykov Raznyh Semej. [Comparative-Historical Study of the Languages of Different Families]. – Moskva: Nauka, 1982, p. 274 – 290.

Okladnikov, 1938 - Okladnikov, Aleksej Pavlovich. Archeologicheskie dannye o drevneishej istorii Pribaikal'ja. [Archeological data on the ancient history of the Lake Baikal area]. - in: Vestnik Drevnej Istorii, # 1 (2), 1938, p. 224 - 260.

Okladnikov et al., 1976 - Okladnikov, Aleksej Pavlovich and Vasil'evskij, Ruslan Sergeevich. Po Al'askr i Aleutskim ostrovam. [Travels in the Alaska and the Aleut Islands]. - Novosibirsk: Nauka, 1976. - 168 pages.

Shirokov, 1985 – Shirokov O.S. Vvedenie v jazykoznanie. [Introduction into Linguistics]. – Moskva: Moscow University Press, 1985.

Tambovtsev, 1977 – Tambovtsev Yu. A. Nekotorye harakteristiki raspredelenija fonem mansijskogo jazyka. [Some characteristics of phonemic distribution in Mansi]. – In: Sovetskoe finno-ugrovedenie. [Soviet Finno-Ugric studies]. T.XXXIII, # 3, 1977, p. 195 – 198.

Tambovtsev, 1984 - Yuri A. Tambovtsev. Phoneme Frequency and Closeness Quotient: Establishing Genetic Relationship Degrees by Phonostatistics. In: UralAltaic Yearbook, 56, 1984, p. 103 119.

Tambovtsev, 1992 - Yuri A. Tambovtsev. Phonostatistical Characteristics of Different Dialects of Eskimo. In: The Bulletin of the Phonetic Society of Japan, #200, August, 1992, p.19 - 20.

Tambovtsev, 1992a - Yuri A. Tambovtsev. The Measurement of the Distance between Languages within Language Families Based on the Value of the Consonantal Coefficient. (Part 3). In: The Bulletin of the Phonetic Society of Japan, # 199, April, 1992, p. 27 - 43.

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, 1998 - Yuri Tambovtsev. Some Phonological Features as Measure of Closeness of Dialects. – In: Philologia FennoUgrica, #4, 1998, p.1 - 19.

Tambovtsev, 1999 - Yuri Tambovtsev. Language Propinquity from the Point of View of Phonological Features. – In: Linguistica Uralica, XXXV, #1, 1999, p.49 58.

Tambovtsev, 2001a – Yuri Tambovtsev. Kompendium osnovnyh statisticheskih harakteristik funktsionirovanija soglasnyh fonem v zvukovoj tsepochke anglijskogo, nemetsekogo, frantsuzskogo I drugih indoevropejskih jazykov. [Compendium of the basic statistical characteristics of functioning of consonantal phonemes in the sound chain of English, German, French and other Indo-European languages]. – Novosibirsk: Novosibirsk klassicheskij institute Press, 2001a.

Tambovtsev, 2001b – Yuri Tambovtsev. Funktsionirovanie soglasnyh fonem v zvukovoj tsepochke uralo-altajskih jazykov. [Functioning of the consonantal phonemes of the Ural-Altaic languages]. – Novosibirsk: Novosibirsk: Novosibirsk klassicheskij institute Press, 2001b.

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.

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, 2003 – Yuri Tambovtsev. Tipologija funktsionirovanija fonem v zvukovoj tsepochke indoevropejskih, paleoaziatskih I drugih jazykov mira: kompaktnost' podgrupp, grupp, semej I drugih jazykovyh taksonov. [Typology of functioning of phonemes in the sound chain of Indo-European, Paleo-Asiatic, Uralo-Altaic and other world languages: compactness of the sub-groups, groups, families and other language taxa]. - – Novosibirsk: Novosibirsk: Novosibirsk klassicheskij institute Press, 2003.

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

Toporova, 1975 – Toporova I. N. Tipologija fonologicheskih system jazykov Bantu. [Typology of the phonological systems of the Bantu languages]. – Moskva: Nauka, 1975.

Whaley, 1997 – Whaley, Lindsay J. Introduction to Typology: The Unity and Diversity of Language. – London: SAGE Publications, 1997.

Zinder, 1979 – Zinder L.R. Obshchaja fonetika. [General phonetics]. – Moskva: Vyshsaja shkola, 1979.

Zipf, 1935 – Zipf, George Kingsley. The Psycho-Biology of Language: An Introduction to Dynamic Philology. – Cambridge: Massachusetts Institute of Technology Press, 1935.