Yuri Tambovtsev
Novosibirsk Pedagogical University
Novosibirsk, Russia

The Altaic Language Taxon: Language Family or Language Union?

Abstract. This article considers the basis for regarding the Altaic language taxon a family or a union. It is necessary to determine if the taxon of Altaic languages is natural enough to call it a family, rather than a union. If it may it is found to be a good classification, that is the classification which may be called natural. The more compact a taxon, the more natural it is. The data considered in this article deal with the peculiarities of labial consonants in the languages of the Altaic language unity and the peculiarities of functioning of labial consonants in subgroups, groups, families, and other language taxa of world languages. The analysis is made with the help of such statistical methods as the coefficient of variance, the confidence interval, Chi-square, and t-test. The linguistic conclusions on 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 some language taxa (Turkic, Finno-Ugric, Tungus-Manchurian, Slavonic, etc.) based on values determined by the t-test.

Introduction

It is accepted in linguistics that a language family is a set of languages deriving from a common ancestor or parent (**Crystal, 1992: 113**). Genetically close languages usually are typologically similar. On the other hand, a loose set of languages is called a language union. They are not genetically close. A good example of this is the Balkan language union, which consists of different language groups.

The Altaic languages comprise the languages of the three linguistic families: Turkic, Mongolic, and Tungus-Manchurian. Some linguists put them together into one family and call it

"the Altaic family" (**Crystal, 1992: 16**). We think it is not advisable to call it a family, since its parts are also families. Jaklin Kornfilt calls the set of Turkic languages a family because in terms of linguistic structure, the Turkic languages are very close to one another according to many features **(Kornfilt, 1990: 619)**. According to the theory of classification, it is wrong to call the parts of the hierarchical classification by the same names as the comprehensive classification. In fact, the Turkic, Mongolic, and Tungus-Manchurian language taxa are well established language families. The details of the discussion about whether they are families can be found elsewhere (**Tambovtsev, 2001a; 2001b; 2001c**). Therefore, a family cannot be included in a family. A family can be included only in a higher taxon, in this case a language unity **(Tambovtsev, 2003-a: 5).** One can see that Jaklin Kornfilt regards the Altaic set of languages not to be a family but a phylum (**Kornfilt, 1990: 620).** In its turn, Manchur-Tungusic languages are surely a family **(Sunic, 1968: 53).** There is no doubt, as well, that the Mongolic languages are close enough to constitute a family **(Bertagaev, 1968: 7)**.

Thus the goal of the article is twofold: 1) to consider the similarities and peculiarities of functioning of labial consonants in the Altaic languages; 2) to consider the tendencies of functioning of labial consonants in the three families which enter the Altaic language unity to compare them to the tendencies in 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.) using the values of the t-test. Lindsay J. Whaley is correct to observe that a typological study focused even on a single feature of language may help to understand some basic facts about phonology of this or that language taxon (**Whaley, 1997: 10 – 11**).

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. Their sound closeness is reflected in the frequency of occurrence. The general rule is: the more similar the language taxa, the more similar the frequency of occurrence of their sounds. This is vividly seen on the data of the Slavonic **(Tab.8)** and other genetically related languages which are indeed very close typologically. We can find the phonostatistical closeness, which can give a good clue for the genetic relatedness which can later be established by the comparative method (**Tambovtsev, 2001-d; 2001-e; 2003-a; 2003-b; 2004**).

Generally, it is assumed that the languages which enter the same language group are closer to each other than the languages which enter different language groups. However, one should agree with S. E. Jahontov that there are no classifications which can give us the exact degree of closeness judged from stages (i.e. degrees) of the hierarchy of the classification (**Jahontov, 1980: 148**). Our classification can give the estimation of such closeness.

Why should one use quantitative methods in studying languages? The 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 also quantity (**FS, 1980: 144**). It is also important to study quantity in linguistics. It is rather strange but in linguistics the qualitative studies are preferred and quantitative ones are neglected.

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 the linguistic results are (**Butler, 1998: 255 - 264**).

In fact, in the case of the taxon of the Altaic languages the only remedy is to use statistical methods, since some linguists approve of this taxon and even call it a family, while the other think it not so. The view of the latter was vividly expressed by Boris A. Serebrennikov, who

stated that the relatedness of neither the Altaic languages, nor Caucasian or Nostratic languages has been solidly proved (**Serebrennikov, 1982: 6 -8**). Lindsay Whaley correctly points out that some of the language taxa are generally accepted (e.g., Indo-European) whereas others (e.g., Altaic and Amerind) are highly controversial (**Whaley, 1997: XX**).

When 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 some 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**).

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**). This is done to keep to 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 the world languages, and the majority of the world languages have labial consonants (**Maddieson, 1981; Shirokov, 1985: 30 – 34; Zinder, 1979: 153**). Labial consonants make the mouth resonator longer. Therefore 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 can be pronounced in principle, there can be only 12 types of labial consonants in a human language. It predicts some of the 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 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 these 317 world languages. We were able to compute the frequency of occurrence of sounds in 258 world languages. We observed 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 exist in most world languages; Ian Maddieson calls them "modal" (**Maddieson, 1980**). Our research, though, 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**).

Our data on the frequency of occurrence allowed 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. Our data are much more reliable than Nikonov's but it is possible to watch the common tendencies developing in the Turkic, Finno-Ugric, Tungus-Manchurian, 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; 1998, 1999**).

Research Material

The data on the frequency of occurrence of the labial consonants were received by computing the texts of different languages. In order to make the number of occurrences commensurable we

calculated the percentage of the frequency of occurrence of the labial consonants to all the phonemes in the speech chain. In this way we received 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)**.

Peculiarities of Function of the Labial Consonants in Different 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 has not been 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 using 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 "Dynamic Philology" (**Zipf, 1935: XIV**). He was one of the first linguists who investigated the phenomenon of occurrence of particular speech sounds in the speech chain of the world languages in general.

Some of Zipf's 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. When 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. We could 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. In Tab. 1 the frequency of occurrence of the labial consonants in the sound chain of the taxon of the Turkic Languages are shown. The frequency of the labial consonants of every language of the Turkic language family is calculated as per cent of all phonemes in the sound speech chain. We computed 26 Turkic languages. They can be divided into separate groups: 1) Bulgar; 2) Oghuz; 3) Kypchak; 4) Karluk; 5) Urjanhay; 6) Altai-Kirgiz; 7) Jakut. These groups were devised on the basis of the classifications of both N.A. Baskakov and A.M. Sherbak (Baskakov, 1969; Sherbak, 1994).

However, it is not advisable to split them into groups since their features penetrate into each other so much that every Oguz language has some Kypchak features and vice versa. Ninel Z. Gadjieva points out that the older Turkic languages must have had both Oguz and Kypchak features. Thus, she finds Oguz features in the most Kypchak languages like Kazakh or Tatar. At the same time, the Oguz languages have a lot of Kypchak features (Gadjieva, 1979: 204 – 206).

Let us compare the data from the Turkic family to some other families. Let us take, for instance, the Finno-Ugric language family. We computed 20 languages and dialects of the Finno-Ugric family (Tab. 6). 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, also called the t-test. We will discuss this in greater detail further (see: Method of the research).

It is not our task to go into detail in discussing dialects and languages of the Turkic, Finno-Ugric family or any other family. We'll consider the usual set of languages in every language family accepted by the majority of linguists.

Everything is understood in comparison. In order to understand the tendencies in the Turkic and the other Altaic languages, one has to consider some other language families. If we take a family called Finno-Ugric family, we can see that some of their dialects can be called separate

languages, since their differences at the phonetic and grammar levels are too great (**LWUL, 1993**). 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 dialects, while E. Lagercrantz defines 29 dialects (**JNSFUS, 1966: 155**). The least concentration of the labial consonants was 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% **(Table 4).** As we can see, its value is much greater than the use of labial consonants in the 26 Turkic languages (cf. 11.22% and 8.71%). The minimum is in the Altai-Kizhi language (5.98%) and maximum – in Karakalpak (12.80%) as one can see from Table 1. The data of this table lead us to state that the Turkic languages have a lesser concentration of the labial consonants than the Finno-Ugric languages.

This value (8.78%) 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, the incidence of 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. The Paleo-Asiatic language family Itel'men has the lowest frequency of 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 lower is the value of these two statistical criteria.

The confidence interval (at 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, in order 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).

V. A. Nikonov was one of the first researchers who dealt with the frequency of occurrence of phonemes in 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 it the *depression of the labial consonants*. This phenomenon spreads from Middle Asia to the West. 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).**

Our data also showed this tendency. In fact, according to our computations in the languages of Africa (Bantu) the frequency of occurrence had its maximum in Bemba (18.23%). The concentration of the labial consonants is also high in Swahili (16.61%). In the other African

Bantu language – Xhosa it is 13.60%, Wolof (Niger-Congo family) labials comprise 13.02%, in Tswana – 13.00% of all the phonemes in the speech sound chain. One can clearly see that the use of the labials in the Bantu speech chain is overexploited. We can find the overexploitation of the labial consonants in the other language family of Africa – the Semito-Hamitic family, now called the Afro-Asiatic family. In Hebrew the frequency of occurrence of labials is a little bit more – 13.69%, though it is not a Bantu, but 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 Bantu languages is correct, i.e., the concentration of labials is too high. His statement that only Bantu languages have a great concentration of labials is not correct. Not all of them have an extremely high concentration of labials. In the other languages of Africa, i.e., the Afro-Asiatic family, it is also high enough.

After computing some American Indian languages we found another pole of depression of the labial consonants. 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 less labial depression, i.e. only 5% in Aleut and 6% in Itel'men. Neither of the Turkic, Mongolian or Sino-Tibetan languages, according to Nikonov, has a frequency greater than 10% (Nikonov, 1976: 42). This does not seem to be quite so. We found 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 found in Altai-Kizhi which has only 5.98% of labial consonants in its speech chain. The mean frequency of occurrence for these 26 languages is 8.71%. Later we'll see if Turkic languages differ in their 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%). B.A. Serebrennikov explains the depression of the labials by the following historical facts: the proto-Mongolian [p] became [h] in Middle Mongolian. The proto-Turkic [p] became [h] in some Turkic languages and in some other Turkic languages it dropped (Serebrennikov, 1982: 31).

Our data on the Paleo-Asiatic languages more or less coincide with those of Nikonov. So, in Itel'men we received 6.43% (Nikonov – 6.00%), which is rather close. We have to point out that not only Bantu languages but Polish has a great concentration of labials – 16.66%, though it is a Slavonic, not a Bantu 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 labials in the Tungus-Manchurian languages do not go over the value of 12.46% or under 8.53%, which are close to the limits indicated by V. A. Nikonov.

Research Methods

We tried to compare the functioning of the labial consonants as a group in different language taxa. It is important to choose a criterion of mathematical statistics for this. In discussing the comparison of two language samples, Gustav Herdan proposes the simplest statistical criteria, like the standard error test or Chi-square test (**Herdan, 1966: 35 – 36**). However, the standard error test may be too rough. The Chi-square test may be no good in this case, since it requires the same number of members in a group or the language taxa. In this study we have a different number of languages in different language taxa. Therefore, the most suitable in this case may be the t-test because it does not give a rough estimation, and the number of languages in language taxa is different. As was mentioned earlier, every language taxa has its own mean of 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.

The 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). One can see from the formula of the t-test why it is so:

$$T = \frac{M1 - M2}{\text{Sqr}\{SI\ 1 / n1 + SI\ 2 / n2\}}$$

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;

SI 1 – the value of the standard squared in the first language taxon;

SI 2 – the value of the 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 the t-test 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. Using the confidence interval and 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. 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 in the use of the labial consonants. We measure the confidence interval at the significance level of 5% (**Tambovtsev, 2003: 20**). The coefficient of variance is 15.40%. This is less than in Turkic, 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 changes of different sorts. In fact, the coefficient of variation is the mean of the dispersion in per cent. It shows the degree of 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 found elsewhere (Tambovtsev, 2001-b). 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 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 (null 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 Finno-Ugric labials is 11.19; S^2= 2.82

(Tab.1). The mean for the Turkic family is 8.71, S^2=2.72. Now we must divide every S^2 by the number of the languages in the family. We obtain for the Finno-Ugric family: 2.82 / 20= 0.141 and for the Turkic family: 2.72 / 26 = 0.105. Putting these data in the formula we obtain:

$$T = \frac{11.19 - 8.71}{\sqrt{0.141 + 0.105}} = \frac{2.48}{\sqrt{0.246}} = \frac{2.48}{0.496} = 5.00$$

Now we must calculate the number of the degrees of freedom 20 + 26 – 2 = 44.

We can see from the table of the critical values that at the significance level of 0.05 the critical value is 2.021 (**Butler, 1985: 172**). One can see that 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 this two means. So, we divide the obtained value by the critical value. We call this the TTM coefficient, which can show us how much the Finno-Ugric mean is different from the Turkic mean. Here, TTM=2.47. In the same way we can calculate the distance between the Finno-Ugric mean and the mean of the Samoyedic labial consonants. The data for the Samoyedic languages are taken from Tab.2.

After the calculations by the same formula, we determine the distance between the Finno-Ugric and the Samoyedic families, TTM=0.35. This is much less than one unit, thus there is no statistical difference between the distributions of the labials in both language families. The Slavonic languages (Tab. 10) are typologically much farther away from the Finno-Ugric languages than the Samoyedic ones with the TTM= 1.954. The Mongolic language family shows a greater distance than that, with TTM=3.827. At the same time the Mongolic languages show that they are closer to the Turkic languages (TTM=1.540) than to the Finno-Ugric languages according to the distribution of labials. In this way, one can calculate the typological distances between different language taxa: subgroups, groups, etc. We'll discuss the distances further in more details. 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 put 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 of Language Taxa from the Point of the Distribution of the 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 whether this taxon is a natural classification of typologically close languages or a mere conglomeration of languages constructed by some other criteria, for instance by the principle or geographical proximity. The dispersion may also unite the languages which are genetically or typologically close. Therefore, we can unite all the Amerindian languages into one group since they are all situated in one geographical region.

When we investigate the Indo-European language family, 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 disperse than the Mongolic family (7.55%) (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%). 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 frequency of occurrence of the labial consonants are distributed in the range from 1.70% to 16.66%. The distribution of the labials are homogeneous (TMB = 0.41). This 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. This means that there are few languages which greatly underexploit or overexploit the use of labials in the speech chain.

Everything is known in comparison. This is why it is necessary to analyse the behaviour of the labial coefficients of variance in the ordinary text in a language. Let us calculate the values of the coefficient of variance in several languages using 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 coherent text are more or less the same as across the languages (Tab. 25 – 26).

Discussion of the Results of Measuring the 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 turned to be close to the taxon of the American Indian languages. It may be merely by chance, since our error level is 5%, such that 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 their 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, a common articulatory trend may produce important questions which are usually easily answered if the languages are genetically related (Tambovtsev, 2001a; 2001b). In fact, nothing interferes prevents a language from constructing words which consist only of the labial consonants in combination with different vowels. Let us take only the vowels which occur most commonly in languages: [a, o, u,

e, i]. In this hypothetic language there may be only such words 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”, “bobabubebi” 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.

In explaining the close distances between Turkic and American Indian languages, we must recall the original hypothesis put forward by some 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, who, in 1938, 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 Northern-Eastern point of Siberia. There they found the Bering ice bridge which allowed them to get 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 arrived on the shores of the Pacific Ocean (Okladnikov et al., 1976: 12 - 67). I should guess a part of these peoples moved eastward to the Japanese Islands. Perhaps the ancient Ainu were amaong their number. Then the other Neolithic tribes who were relatives of the Siberian peoples moved farther and got to South America but preserved their articulation basis. This may be why the distribution of the consonantal groups in the Turkic and American Indian languages is typologically similar. We must point to the fact that the articulatory basis is usually preserved, even when the people begin to speak 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 ice Bering bridge to the 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 occurred 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 19 thousand years. At least some 19 years ago it existed. A. P. Okladnikov believed that the Americas were inhabited by 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.27). 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 were in contact. 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 how reliable is 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 the tables (Tab. 27 - 28). 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 the 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.

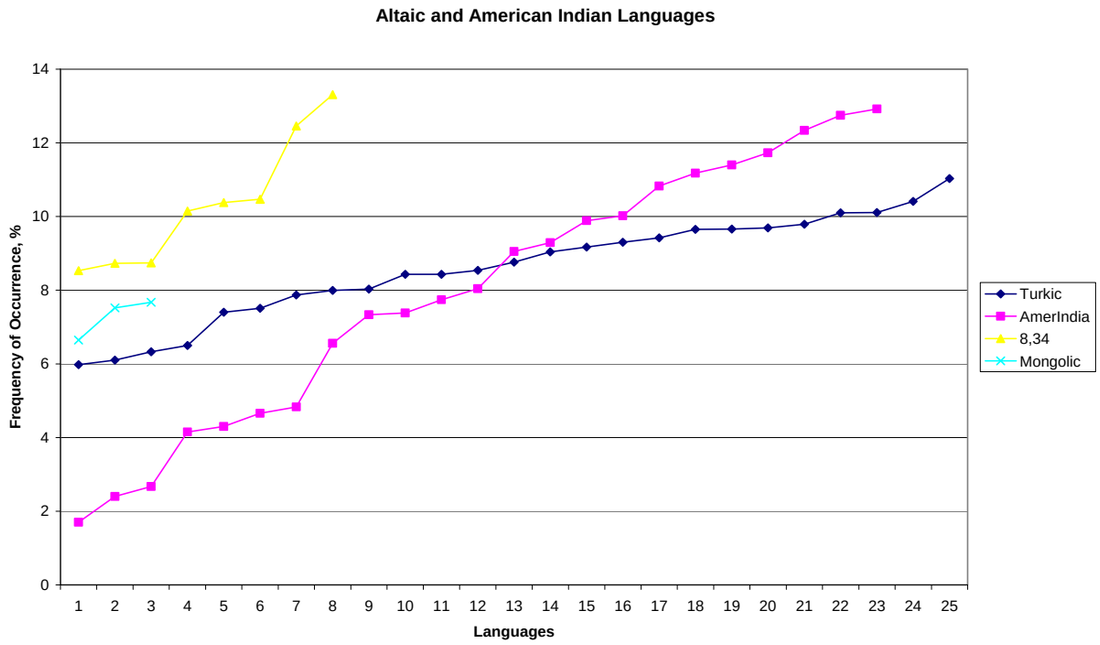
The distances between the Slavonic group of the Indo-European family can be seen from Tab. 28. The Iranian group is the closest to the Slavonic languages (0.019). S.V. Bromley and others claim that the Slavonic tribes came into contact with the Iranian speaking tribes of the Sarmats in the 8^{th}- 9^{th} 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 close (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 Balto-Slavonic language community (Tambovtsev, 2001: 70 – 71). So, one can see that long contact between the Slavonic and Turkic peoples did not influence their articulatory basis, i.e., the articulatory habits concerning the labial consonants were not borrowed. This is why the distribution of the 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).

The tendencies may be seen in Figure 1:

Figure 1



Having analysed the Altaic languages by the common word stock and similar morphology V. L. Kotvich, who was the strong proponent of the relatedness of the genetic relatedness of these languages, came to the conclusion that they are very similar. Now let us consider what V.L. Kotvich thought to be similar. He found a 50% similarity of elements in morphology and 25% in the stock of words of Mongolic and Turkic languages. The similarity between all the three Altaic languages (Tungus-Manchurian on the one hand, and Mongolic and Turkic - on the other) is much less: common word stock is 10% and 5% - in morphology (**Kotvich, 1962: 351)**. As we have proved elsewhere, logical reasoning and mathematical criteria allow us to believe that minimum of 75% of common elements can safely prove similarity. At least in biology, geology

and other natural sciences, two objects are considered similar if and only if they have not less than 75% in common (**Tambovtsev, 2004: 220 - 227**). Comparing this number to those of Kotvich, one can see that Kotvich's claim has no solid foundation. Indeed, how is it possible to speak of any sort of similarity if 90% of the word stock is different? It is even worse for the morphological elements which comprise only 5%. It means that 95% are not similar. In everyday life two objects, 95% of whose elements are different, can hardly be considered similar. In linguistics it should not be different (**Tambovtsev, 2003a; 2003b**).

Conclusions

1. The mean frequency of occurrence of labial consonants in the families of the Altaic language unity: Mongolic – 7.28%; Turkic – 8.71%; Tungus-Manchurian – 10.12%. The world languages taken for this study demonstrated that they are distributed in the range of 1.70% to 17%. The mean is 10.51%. We can state that the languages which employ lesser frequency underexploit the labial consonants while those which employ the greater frequency overexploit them in their speech chains. Our data clearly demonstrate that all Altaic languages in general underexploit the use of the labial consonants. This may be explained as the Asiatic depression of labials. Surely, we could not embrace all the languages of the world but our sample 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 a good clue for understanding how human language works.

2. The least dispersed language taxon is the Mongolic family (V=7.55%). This means that the languages of this taxon are very typologically close. The American Indian languages are quite dispersed, which indicates that their speech sound chains are rather different in structure. This may be explained by the fact that we took many families of the American-Indian languages (V=44.09%). Tungus-Manchurian (17.59%) and Turkic (V=18.94%) are not very compact.

3. The use of the t-test can demonstrate similar and peculiar tendencies in the distribution of the labial consonants in different language 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 Turkic languages is very similar to that of the American Indian languages. The Turkic family is different in the use of labials both to the Tungus-Manchurian (TTM=1.021) and Mongolic (TTM=1.540) taxa, and therefore cannot be considered as part of a family with them.

Tables

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

Tab. 2
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.	Mongolic	7.52		3.	Kalmyk	6.65
2.	Buriat	7.67				
	Statistical data					
	Mean	7.28			SI	0.30
	S	0.55			V %	7.55

Tab. 3
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				
	Statistical data					
	Mean	10.12			SI	3.17
	S	1.78			V %	17.59

Tab. 4
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.	Mansi (Northern)	13.56		11.	Mordovian-Moksha	11.26
2.	Mansi (Konda)	12.29		12.	Mordovian-Erzia	13.72
3.	Hanty (Kazym)	12.60		13.	Vodian	11.95
4.	Hanty (Eastern)	10.45		14.	Vepsian	11.11
5.	Hungarian	10.04		15.	Karelian-Tihvin	9.66
6.	Komi-Zyrian	10.21		16.	Karelian-Livvik	11.16
7.	Komi-Permian	11.15		17.	Karelian-L'udik	8.66
8.	Udmurt	13.66		18.	Finnish	8.73
9.	Mari-Lawn	9.47		19.	Estonian	10.21
10.	Mari-Mountain	9.99		20.	Saami	14.44
	Statistical data					
	Mean	11.19			SI	2.82
	S	1.68			V %	15.04

Tab. 5
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.	Nenets	12.14		3.	Nganasan	7.71
2.	Sel'kup	11.99		4.	Kamasin	13.99
	Statistics data					
	Mean	11.46			SI	7.08
	S	2.66			V %	23.21

Tab. 6
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.	Chookchee	8.76		4.	Eskimo-Naukan	7.76
2.	Koriak	10.00		5.	Eskimo-Imaklin	6.72
3.	Itel'men	6.43				
	Statistical data					
	Mean	7.93			SI	2.18
	S	1.48			V %	18.61

Tab. 7
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.	Hebrew	13.34		5.	Assirian	13.39
2.	Arabic	13.42		6.	Somalian	7.62

3.	Neo-Aramaic	11.92		7.	Sokotrian	11.18
4.	Hausa	9.93				
	Statistical data					
	Mean	11.54			SI	4.75
	S	2.18			V %	18.89

Tab. 8
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.	Hindi	9.97		4.	Marathi	9.51
2.	Bendali	10.06		5.	Gipsy	10.61
3.	Gudjarati	11.35				
	Statistical data					
	Mean	10.30			SI	0.50
	S	0.71			V %	6.85

Tab. 9
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.	Gilian	15.18
2.	Dari (Afganistan)	12.85		6.	Osetian	12.26
3.	Tadjak	13.11		7.	Kurdish	16.25
4.	Talysh	12.81		8.	Pashto	12.82
	Statistical data					
	Mean	13.38			SI	2.33
	S	1.53			V %	11.40

Tab. 10
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.	Russian	12.63		7.	Slovenian	12.54
2.	Ukranian	13.01		8.	Polish	16.66
3.	Belorussian	14.45		9.	Slovak	12.79
4.	Serbian	11.96		10.	Czech	13.57
5.	Bulgarian	12.91		11.	Sorbian	14.83
6.	Macedonian	11.67				
	Statistical data:					
	Mean	13.35			SI	1.90
	S	1.38			V %	10.34

Tab. 11
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
	Statistical data:					
	Mean	12.25			SI	3.84
	S	1.96			V %	16.00

Tab. 12
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.	Italian	10.38		4.	French	13.96
2.	Spanish	9.79		5.	Rumanian	10.22
3.	Portuguese	11.10		6.	Moldavian	11.06
	Statistical data					
	Mean	11.08			SI	2.24
	S	1.50			V %	13.49

Tab. 13
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.	Eglish	13.05		5.	Danish	11.95
2.	Dutch	12.03		6.	Norwegian	10.60
3.	German	9.88		7.	Swedish	11.00
4.	Gothic	10.56				
	Statistical data					
	Mean	11.30			SI	1.19
	S	1.09			V %	9.65

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

#	Language	%		#	Language	%
1.	Greek	10.81		3.	Albanian	12.07
2.	Armenian	10.32				

Tab. 15
The Frequency of Occurrence of the Labial Consonants in the Sound Chain of the Taxon of the Isolated Paleo-Siberian Languages (% to all phonemes).

#	Language	%		#	Language	%
1.	Ket (Yug)	8.36		3.	Nivhian	11.34
2.	Yukaghir	11.10				

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

#	Language	%		#	Language	%
1.	Japanese	6.94		3.	Ainu	9.28
2.	Korean	10.00				

Tab. 17
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.	Georgian	13.35		4.	Avarian	9.75
2.	Adygian	12.22		5.	Abhazian	9.17
3.	Kabardian	10.70		6.	Chechenian	7.51
	Statistical data					
	Mean	10.45			SI	4.67
	S	2.16			V %	20.67

Tab. 18
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.	Chinese	9.13		4.	Tibetan	12.67
2.	Thai	12.63		5.	Dungan	8.22
3.	Burmanese	8.79				
	Statistical data					
	Mean	10.29			SI	4.75
	S	2.18			V %	21.19

Tab. 19
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. 20
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.	Tagalog	10.50		4.	Dajak	8.77
2.	Indonesian	11.96		5.	Maori	7.11
3.	Hawaian	7.87		6.	Marquis	7.80
	Statistical data					
	Mean	9.11			SI	4.70
	S	2.17			V %	23.79

Tab. 21
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.	Djingili	11.35		4.	Ngandi	9.92
2.	Mangarayi	14.51		5.	Nyangumada	10.40
3.	Ngaanyatjarara	8.42		6.	Nunggubuyu	12.47
	Statistical data					
	Mean	11.18			SI	4.54
	S	2.13			V %	19.07

Tab. 22
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				
	Statistical data					
	Mean	7.96			SI	12.35
	S	3.51			V %	44.09

Tab. 23
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	8.78		12.	Afro-Asiatic	11.54
6.	Tungus-Manchurian	10.12		13.	Indo-European	12.22
7.	Sino-Tibetan	10.29				

Tab. 24
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.	Australian (aborig.)	19.07
2.	Indo-European	14.66		9.	Caucasian	20.67
3.	Finno-Ugric	15.04		10.	Sino-Tibetan	21.19
4.	Tungus-Manchurian	17.59		11.	Samoyedic	23.21
5.	Paleo-Asiatic	18.61		12.	Austronesian	23.79
6.	Afro-asiatic	18.89		13.	American Indian	44.09
7.	Turkic	18.94				

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. 30
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 T.	2.496
2.	Kypchak group of T.	2.091				

Tab. 31
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

Baskakov, 1969 – Baskakov N.A. Vvedenie v izuchenie t'urkskih jazykov. – Moskva: Nauka, 1969.

Bertagaev, 1968 – Bertagaev T.A. Mongol'skie Jazyki [Mongolic languages]. - In: Jazyki Narodov SSSR. Volume 5. Mongol'skie, Tunguso-Man'chzhurskie I Paleoaziatskie Jazyki. – Leningrad: Nauka, 1968, p. 7 – 33.

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

Butler, 1985 – Butler Ch. 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 D. An Encyclopedic Dictionary of Language and Languages. - Oxford: Blackwell, 1992.

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

Gadjieva, 1979 – Gadjieva N. Z. T'urkojazychnye arealy Kavkaza [The Turkic language areals of the Caucasus]. – Moskva: Nauka, 1979.

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

Jahontov, 1980 – Jahontov S. E. Otsenka stepeni blizosti rodstvennyh jazykov. [Estimation of the degrees of the closeness of the related languages]. - In: Teoreticheskie osnovy klassifikatsii jazykov mira. Problemy rodstva. [Theoretical foundations of the classification of world languages. The problems of relatedness]. - Moskva: Nauka, 1982, p. 148 – 157.

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.

Kornfilt, 1990 – Kornfilt J. Turkish and the Turkic languages. – In: The World's Major Languages (edited by B. Comrie). – Oxford: Oxford University Press, 1990, 619 – 644.

Kotvich V. – 1962. Study of on Altai languages. IIL, M.

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

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.

Sherbak, 1994 – Sherbak A. M. Vvedenie v sravnitel'noe izuchenie t'urkskih jazykov. [Introduction to the comparative study of the Turkic languages]. – Sankt-Peterburg: Nauka, 1994.

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

Serebrennikov, 1982 - Serebrennikov B. A. Problema dostatochnosti osnovanij v gipotezah, kasajushchihs'a geneticheskogo rodstva jazykov. [The problem of the sufficient foundation in the hypotheses concerning the genetic relatedness]. - In: Teoreticheskie osnovy klassifikatsii jazykov mira. Problemy rodstva. [Theoretical foundations of the classification of world languages. The problems of relatedness]. - Moskva: Nauka, 1982, p. 6 - 62.

Sunik, 1968 – Sunik O.P. Tunguso-man'chzhurskie jazyki [Tungus-Manchurian languages]. – In: Jazyki Narodov SSSR. Volume 5. Mongol'skie, Tunguso-Man'chzhurskie I Paleoaziatskie Jazyki. – Leningrad: Nauka, 1968, p. 58 – 67.

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, 2003a – 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, 2003b – Tambovtsev, Yuri. Izmerenie fonostatisticheskih rasstojanij mezhdu ural'skimi jazykami. [Measuring phonostatistical distances between Uralic languages]. – In: Fenno-Ugristica # 25, (Tartu), 2003, p.120 – 168.

Tambovtsev, 2003c - Tambovtsev, Yuri. Phononologal similarity between Tofa and some other languages of Asia based on the frequency of occurrence of certain typological consonantal features. – in: Fenno-Ugrica # 25, (Tartu), 2003, p. 169 – 179.

Tambovtsev, 2004 - Tambovtsev, Yuri. Uralic language taxon: natural or atificial? (Typological compactness of Uralic languages and other language taxa: branches, subgroups, groups, families and super-families). - In: Fenno-Ugristica # 26, (Tartu), 2004, p. 200 - 233.

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

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.

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 – The Psychobiology of Language. Houghton Mifflin, New York