ASIAN AND AFRICAN STUDIES, *10*, 2001, 1, 11–43

THE PHONOLOGICAL DISTANCES BETWEEN TURKIC LANGUAGES BASED ON SOME PHONOLOGICAL FEATURES OF CONSONANTS

Yuri TAMBOVTSEV
Department of English and Linguistics, Novosibirsk Pedagogical University,
P. O. Box 104, Novosibirsk-123, 630123 Russia

The Turkic languages have a long history of its development, during which they had contacts with various languages, among them Mongolian, Tungus-Manchurian, Ugric, Samoyedic, Slavonic, Iranian and other languages. All these contacts influenced the Turkic languages to some extent. They also had intensive contacts between themselves because the peoples, who spoke the Turkic languages were nomads. Usually a nomadic tribe has extensive and intensive contacts with other tribes or peoples. The aim of this paper is to discover the phono-typological distances between the selected Turkic languages. Some of them had more contacts, some of them had less contacts but all of them had various contacts, which certainly had a direct or indirect influence on them. Ancient Mongolians ousted some of the ancient Turkic tribes from the places of their original settlements, which led to a great mixture among the Turkic peoples (Gumiljov, 1994: 22 – 23).

The application of statistical methods to the problems of determination of linguistical relationship was discussed by many linguists from the point of view of measuring the quantitative weight of very slight but concrete linguistic evidence, and deciding, free from subjectivity as far as this is humanly possible, whether the weight is sufficient to carry proof. One cannot help agreeing with H. K. J. Cowan that it is not always fully realized that the results achieved by the traditional comparative method in linguistics are really based not on objective proof, but on subjective impressions and the evidence unconsciously implied the bulk of some quantitative data (Cowan, 1962: 58 – 60).

Every language has its own spirit or internal energy, which is embodied into different amounts of material entities, i.e. speech sounds. Phonetics is the only branch of linguistics which studies the material entities. This is why the classification on the basis of the occurrence of speech sounds in the language sound chains is based on a solid fundamental. A science is said to reach its fully developed stage when having used exact methods of investigation it obtains a true classification of the objects it studies. Using statistical techniques and accurate

quantificational procedures, a linguist can create a typological scheme of languages inside a certain genealogical language family taking into account the similarity of sound chains (Tambovtsev, 1994-a, 1994-b, 1997). Joseph H. Greenberg considers the term "type" to have a connotation opposite to arbitrariness to the highest possible degree. One cannot help agreeing with his opinion that a type is a class sharing essential, not merely accidental, properties (Greenberg, 1957: 68).

Speaking about the evolution of the use of statistical methods in the analysis of literary style (stylometry), which is also a branch of mathematical linguistics, David Holmes, a mathematician, remarks that the works in the field of mathematical linguistics, which are usually packed with tables, charts and statistical analyses, are often not understandable to the majority of the linguists (Holmes,1998: 111). In this work, we tried to introduce some of the simplest statistical methods, so that the reasoning should be clear even to the scholars who are against modern methods in linguistics and who, as S. Wells puts it, "have neither the inclination, nor even the intellectual ability to understand, let alone to assess" (Wells, 1996: 28). It is advisable to use statistical methods in linguistics on the phonetic or phonological level because the features, which we select for the analysis, fulfil Bailey's requirements, since they are salient, structural, frequent, easily quantifiable, and relatively immune from conscious control (Bailey, 1979). By measuring and counting the distribution of the selected features in the speech chain of a language, we hope to uncover the peculiar characteristics of this language, which make it similar or dissimilar to another language (Tambovtsev, 1994-a; 1994-b). This work uses the methods of phonetic typology, which refer to a field of quantitative linguistics in which the language classification is made on the basis of certain phonetic features. The concept of phonetic typology so far has been discussed without the methods of phonological statistics: it is possible to categorize languages into different typological types on the basis of the certain sound mosaic picture given by the frequency of certain phonological groups (Tambovtsev, 1984, 1991). Every language has its own sound chain which is unique. This speech sound chain creates a unique picture of phonemic frequencies. Then this phonemic frequency pattern can be compared to the analogical pattern of some other languages (Tambovtsev, 1977, 1983). Even before the conscious introduction of the phonostatistical methods in linguistics, as N. Z. Gadjieva points out, phonemic frequencies were used in Turkic languages to distinguish dialects of the Uzbek language long ago: for example A. K. Borovkov classified the Uzbek dialects of the Samarkand – Buhara group as the one which has a greater frequency of occurrence of the phoneme /a/ in the first syllable. V.V. Reshetov defined some dialects in Uzbek on the basis of the prevailing frequency of the /e/ affix. N.A. Baskakov used some frequent consonantal transitions as the feature to differentiate dialects in Nogaj (Gadjieva, 1980: 102). It is possible to transfer the phonemic frequencies into the phonological distances on the basis of the differences between the frequencies in question. Viktor Krupa correctly points out that in 1928 Jan Czekanowski was one of the first to use the correlation techniques giving the typological

distances between some Indo-European languages on the basis of several diagnostic features. Jan Czekanowski compared pairs of languages and looked for the verification of four possible outcomes: a) both languages shared the selected features; b) the features occurred only in the first language; c) the features occurred only in the second language; d) the features were both absent from both languages (Krupa,1970: 60). This technique was used in determining the phonological distances between the dialects of the Karelian language, a member of the Finno-Ugric language family (Tambovtsev, 1984-a). Jan Czekanowski also introduced in linguistics the method of clustering languages basing on several features and many other useful quantitative techniques (Czekanowski, 1957). It should be noted that Jan Czekanowski did not invent his basic method of investigation. It was used in mathematical statistics and is known there as the tetrachoric test of independence and association, and is actually a modified form of the "chi-squared" test. This tetrachoric test ("tetra"- in Greek means "four") is so called because of the four outcomes described above and can be found in any desk-book on statistics (e.g. Butler, 1985: 120). Nevertheless, Jan Czekanowski should be praised for introducing into linguistics many methods, which were much earlier used by the anthropologists and biologists (Czekanowski, 1957). As early as 1909, Jan Czekanowski introduced the notion of the construction of the linked unlocked graph which resembles the method of the nearest neighbour used in linguistics much later (Czekanowski, 1909; 1957). The pioneering works of Jan Czekanowski gave rise to numerous works in this direction by C.D. Chretien, A.L. Kroeber, J. Ellis, A. Ellegard, J. Levin, C.R. Sankaran, A. D. Taskar, P.C. Ganeshsundaram, Yu. Tambovtsev, etc. They used the same basic tetrachoric principle, though their formulas were different. V. Krupa gave a good analysis of these formulas, so there is no need for us to do it (cf. Krupa, 1970). Investigating the speech chains in different languages from the phonostatistical point of view, I came to the conclusion that comparing just the inventories or the inventories and phonemic frequencies is not an adequate tool since it can show two similar languages as distant because some phoneme "A" in language "A" may have a correspondence in phoneme "B", but not "A" in language "B". Viktor Krupa also emphasizes the need for commensurability as well as the need to proceed beyond comparing merely the inventories of the phonemic systems (Krupa, 1970: 60). Later, I understood Krupa's remark that the demand for commensurability must lead to the construction of a set of distinctive features suitable for the description of all languages included in the investigation and resorted to the set of features constructed on basic salient consonantal groups, defined by the work of the active speech organ, the manner of articulation and the work of the vocal cords. I have to note further that the frequencies of phonological groups or in other words the occurrence of certain types of phonemes in the speech chain of this or that particular language can be proved. As a matter of fact, it was proved that the occurrence of linguistic units in a language has a high degree of orderliness (Butler, 1985; Kenny, 1986; Ledger, 1989; Tambovtsev, 1986; Zipf, 1935). On the basis of these directional principles, the aim of this paper is to find peculiar frequency characteristics of some

phonological groups in the speech chains of the Turkic languages and Mongol and, then, to construct the phono-typological distances between them. This will help us later to define the common traits in the sound typology of the Uralic and Altaic languages from a new typological viewpoint. It goes without saying that we cannot use here Czekanowski's method based on the four features (a, b, c, d) described above. It was necessary to develop some other method which could build the linguistical distances on only one feature "A", that is when the features are shared by both languages. I have applied a method, called Euclidean distance, though a modified "chi-squared" method, I used elsewhere, could also be applied here. The method of calculating the distances on the basis of Euclidean space shall be described in detail below.

Phonostatistical investigations of some Finno-Ugric languages published by the author (Tambovtsev, 1977, 1979, 1981, 1985-b, 1986, 1995) make it possible to compare them to Hakas, Kazah, Yakut, Uzbek, Turkmen, Turkish and other Turkic languages whose frequency characteristics were also studied by the author (Tambovtsev, 1991). It shall allow us to speak of the universal peculiarities of the Ural-Altaic languages. At the modern stage one cannot say for sure whether Ural-Altaic languages should be considered as one language unity. The same is true even of the Altaic languages. One can see two possibilities: a) the Altaic languages diverged; b) the Altaic languages converged. The first possibility is vividly demonstrated by the modern Scandinavian languages. They used to be very close and have a common protolanguage, though later in the course of its development they reformed losing a number of common typological features and acquiring a number of new different ones, so that now their former closeness could be restored to see if they were really related. An opposite example is the Sprachbund of the Balkan languages, when the languages acquired common typological features as a result of their long contact. It was rather surprising for me to find out that animals have more or less identical causes for similarities: a) the similarity was caused by the common features of the common ancestor; b) the similarity was caused by the acquired features which were developed by the common trend due to the common ancestor, that is similarity because of parallel evolution; c) the similarity was caused by acquired features which turned to be common, that is similarity because of convergence (Mair, 1971: 232). R. Austerlitz believed that the contacts of languages should be considered in dynamics, when they are repeated in layers, so that the waves (Wellentheorie) of related languages or dialects are deposited, one on top of the other, creating even more diversity, which makes the task of the comparative linguist even more difficult (Austerlitz, 1982: 53).

At the first step of constructing the typological distances, we have chosen to take into account only the consonantal groups leaving the vowel group frequencies for the second step, that is when we are unable distinguish between two languages, only on the basis of data on the frequency of occurrence of these consonantal groups. Consonants bear the main semantic meaning in the word, while vowels serve to prepare the articulators for forming the following consonant because very often it is not possible to have a consonantal cluster of this or that

kind (Tambovtsev, 1979, 1981). So, one can say that often consonants are more important for communicational understanding. Ludmila G. Zubkova studied in detail the functioning of vowels and consonants in many languages and discovered the same tendency, for example in Russian the consonants express the lexical meaning, while vowels express grammatical meaning. She argues that in Russian prose the nouns have 28.4% of vowels and 3% of consonants in morphs, because the noun is the less grammatical part of speech. The verb, which is the most grammatical part of speach, the exponents of morphs are 40.9% of vowels and 25.9% of consonants (Zubkova, 1988: 80). The other reason, why I took consonants was that consonants usually form into groups from the point of articulation (better, more correctly). Phonetically, consonants are sounds made by a closure or narrowing in the vocal tract so that the air flow is either completely blocked, or so restricted that audible friction is produced. This is why consonant articulations are relatively easy to feel, and as a result are most conveniently described in terms of place (active organ of speech) and manner of articulation. In addition, a routine phonetic description of consonants involves information about the mode of vibration of the vocal cords, etc. (Crystal, 1980: 82). Our assumption is that a certain language having a certain articulation basis, prefers this or that sound, which is later embodied through the different frequency pattern.

We understand the articulation basis as a set of movements of the speech organs and their conditions required for the articulation of the sound of a certain language. The articulation basis of a certain language (a dialect) is the sum of articulatory habits which children acquire during their growth from the utterances of the adults around them, so that their speech organs can more readily pronounce this type of consonant, rather than another. So the frequency data of consonantal use may serve as a tool for comparative typological studies of Uralic and Altaic languages.

We have defined our consonantal groups from the point of view of place of articulation (or the work of the active organs of speech), manner of articulation, and the work of vocal cords, because in this case their most salient features are involved. During the development of phonetics many other features (e.g. Jakobsonian) spring out and then go, but the articulatory features were the most salient for the Old Indian linguists long before our era, and so they remain. I shall not criticize the feature system invented by Roman Jakobson, and later developed by Noam Chomsky and Morris Halle, because it was done smashingly by such outstanding phoneticians as L.V Bondarko, L.R. Zinder (Bondarko et al., 1966), and Peter Ladefoged (Ladefoged, 1971). I am also convinced that the feature system theory does not give any better clues than were long ago defined in articulatory phonetics. The work of the active speech organs or place (point) of articulation is one of the three above mentioned main parameters for consonants used in the phonetic classification of speech sounds, referring to which active organ works: lips (labial), the tip and front of the tongue (front), the middle of the tongue (medio-lingual or palatal) and back of the tongue with uvula (velar or back). Pharyngeal or glottal consonants are seldom found in languages, so it is not necessary to create a separate group for them: they should enter

the group of back consonants. The fact that there are four features from this point of view allows us to have two charts: 1) labial-front; and 2) palatal-back consonants.The Turkic languages are depicted according to these parameters in Fig. 1 and Fig. 2.

The other three important features are provided by the classification of consonants from the point of view of the manner of articulation, that is the classification, referring to the kind of articulatory process used in a consonant's production. Usually, within consonants several articulatory types are recognized, based on the type of closure made by the speech organs. If complete closure occurs, then the consonant is called occlusive. Affricates are also included in this group. A fricative consonant is the result of a narrowing without complete closure. Sonorants are the consonants produced with a relatively free air flow and voicing but at the same time having all the consonantal characteristics, i.e. the complete or incomplete closure. Thus, we define 1) sonorants; 2) occlusives and 3) fricatives. It is necessary to add here the fourth characteristic, which is voicing in order to have two more charts: Fig. 3 for sonorant – occlusive; Fig. 4 for fricative – voiced. The theory of classification recommends use of several features derived from different points of view to obtain true results (Itkonen, 1980: 334 – 336; Rozova, 1986: 159). According to these demands the features should not be intersecting, that is the frequencies of the sonorants should not be included in the total frequencies of the voiced consonants (Tambovtsev, 1990; 1990-a; 1988; 1991). Therefore, in this study we took the frequencies of the voiced occlusives and voiced fricatives. In addition to phonostatistical data obtained for some Finno-Ugric languages, I have also considered elsewhere the phonemic frequency data on some Tungus-Manchurian languages, which together with the Turkic languages comprise the Altaic super language family or Sprachbund (Tambovtsev, 1984, 1987, 1988, 1989).

The corpus of Turkic and other Ural-Altaic languages was fed into computers by the author and many native speakers of Ural-Altaic languages who were students at Novosibirsk University during 1972-1987. The author thanks them all for their generous help. The author also received great aid from the scholars of the department of Statistics and Probability of the Mathematical Faculty of Novosibirsk University, especially Dr. Arkadiy Shemiakin. The samples were large enough to yield statistically stable and significant results. Usually the samples were greater than 100,000 phonemes, though some of them were 500,000 and more phonemes. The sample volumes are discussed in detail elsewhere (Tambovtsev, 1977, 1985, 1986; Tambovtsev et al., 1982). However, mathematical statistics require that the sample on average should not be less than 30,000 phonemes (Tambovtsev, 1994-a; 1994-b; Tambovtsev et al., 1993; 55-56). In my view, it is more desirable to base our considerations on the frequencies of the entire phonemic classes (e.g. vowels and consonantal groups (e.g. labial, front, palatal, back, sonorant, occlusive, fricative and voiced), rather than on the frequencies of separate individual phonemes because they are more statistically stable, i.e. linguistically reliable (Tambovtsev, 1991: 157-164). The investigation of Hakas, Kazah, Uzbek, Turkmen and Jakut showed that the com-

parison between the Turkic languages though having very similar phonetic inventories, should be made between phonemic classes or groups, since some phonemes in one Turkic language may have no identical counterpart in another Turkic language. It is also true for the Finno-Ugric, Tungus-Manchurian and other languages (Tambovtsev, 1983; 1984; 1991). The five Turkic languages studied before (Tambovtsev, 1991) were considered from the point of view of only four consonantal groups, that is from the point of view of the work of the active organ of speech (or the place of articulation), and even these four features (frequency of occurrence of the labial, front, palatal and back consonants) allowed us to find certain distances between them. It is interesting to see if the positions of these languages will change if we introduce some more features. In this work some more groups of four features are added: defined by the manner of articulation (sonorant, occlusive and fricative consonants) and by the work of the vocal cords (voiced consonants). The late Vladimir M. Nadel'aev (1910 – 1985) recommended to call "voiced" consonants as "weak", and sonorant consonants as "superweak", because his phonetic studies showed that in this case the most important feature is the degree of tenseness of the basic active organ of speech, forming the constriction in the vocal tract, but not the work of the vocal cords (Nadel'aev, 1985: 11 – 12). He often underlined that articulation base influences the frequency of occurrence of certain vowels and consonants. Articulation base was regarded by him as the most conservative structure of the language which usually remains the same. It allowed him to put forward the ideas of using the data of experimental phonetics as the clue for defining the ethnogenetic hypothesis concerning the Mongolian, Turkic and other peoples (Nadel'aev, 1986: 34, 53 – 63).

It is believed that these 8 basic articulatory features are quite sufficient for every language to be separate in the Euclidean space, though, in principle, it is possible to make the measuring grid finer, if necessary, by adding some more features, for example splitting the labial consonants into labio-labial and labio-dental, or splitting labials into sonorant labials, occlusive labials and fricative labials. However, such a detailed grid may have many empty cells, numerically embodied in zeros. It may give a false picture, that is two similar languages which do not have identical phonemic divisions will be regarded as distant, that is placed far apart, while in fact they are similar, that is close in the Euclidean space. One can consider this situation with the help of the formula for computing linguistic distances, provided below. In one of my previous works (Tambovtsev, 1991) I have also applied the classification of consonants, leaving the vowel data for the second stage of classification. However, it turned out that consonantal features can measure the distances between the five Turkic languages taken for the study, all right. We did not need to resort to the additional features, that is the vowel features, or the additional consonantal features.

Many linguists believe that some of the Turkic languages are closer to Mongolian. O. Behtlingk found Mongolian words in the first Jakut tests (Behtlingk, 1964: 37). W. Radloff was so impressed by the Mongol-Jakut similarities on all levels (phonetic, morphological, lexical and even syntactical), that he thought

that Jakut, a language of unknown origin, first turned into a Mongolian language and then into a Turkic one (Ubrjatova, 1969: 29 – 30). Though later Radloff's idea was rejected, one should explain the abundance of Mongolian elements in Jakut. E.I. Ubrjatova considered Jakut that has a Mongolian substratum, since Turks do not borrow verbs (Ubrjatova,1969: 29 – 30). N. Shirobokova is sceptical about the possibility of giving the exact time of Jakut-Mongolian contacts since there are some views that they were in the 12th,14th, 15th, or even 16th centuries (Shirobokova, 1980: 140-141). Comparing Jakut and Burjat lexis, she comes to the conclusion that Jakut also greatly influenced Burjat (Shirobokova, 1980: 147). Our phonostatistical investigation may show whether Jakut is on the outskirts of the Turkic group from the phonological point of view. This study may also help to give some support to this or that side in the long discussion of the proponents and opponents of the Altaic hypothesis. In my opinion V.I. Rassadin is certainly correct when he calls upon the scholars to provide new evidence for or against the Altaic language unity, rather than repeat old well-known reasoning for or against it (Rassadin, 1988: 103 – 108).

Let us consider the theoretical possibility of the influence of a certain language on the Turkic languages. One can consider at least 6 possible outcomes. First of all it is the so-called chain influence, when a certain language (let us call this language "A") of a Mongolian, Turkic or any other origin influenced a Turkic language ("B"), but was also influenced by it; and when language ("B") influenced and at the same time was influenced by another Turkic language ("C") and so on (Fig. 1 – 1). Naturally, when one language is in contact with the other language they both influence each other. So, when one declares that one language influenced the other, it is also implied that it was also subject to some sort of influence, though this influence could be much less. The forward and backward directions of the influence are shown on the scheme (Fig. 1) by the arrow directed both ways. Therefore, further the words "and was subject to some influence" should be implied automatically, when the word "influenced" is used. A bit more complicated case is (Fig. 1 – 2) when a certain language "A" influenced a Turkic language ("B"), which in its turn influenced two more Turkic languages. The third possibility (Fig. 1 – 3) is when a certain language "A" influenced a Turkic language ("B") which influenced another Turkic language "C", but at the same time was influenced by the Turkic language "D", which in its turn was influenced by the Turkic language "C". The fourth case (Fig. 1 – 4) shows the influence of Mongolian ("A") on the Turkic languages "B" and "D" and they both in their turn influence another language "C". The fifth case (Fig. 1 – 5) shows the influence of a certain language "A" on all the Turkic languages ("B, C, D"). It looks like a common linguistical situation in ancient times, for instance concerning Mongolian, since Mongolian influenced a great number of languages (Turkic, Tungus-Manchurian, Slavonic, etc.). It is more probable that Mongolian could influence the Turkic languages more because of their geographical closeness. The sixth situation is also rather common for Turkic languages, since it is the case (Fig. 1 – 6) when languages influence each other.

We shall depict all the 13 Turkic languages on 4 different schemes according to their data of the frequency of occurrence of consonants (Fig. 1, 2, 3, 4) and give a preliminary analysis. After that we shall give a detailed statistical analysis.

The measurement along the scale "**labial-front**" showed that Uzbek, Turkmen and Azeri are closer to each other than to other selected Turkic languages. Altai, Ujgur, Karakalpak and Jakut are far away from each other. Jakut seems to be further away from all the Turkic languages, except for Azeri and Kazah and closer to Burjat. Hungarian is closer to Uzbek, Turkmen, and Baraba Tatar than to the other Turkic languages, though it is much more close to them, than Altai, Karakalpak, Jakut or Ujgur are. Though we know that Russian was greatly influenced by the Turkic languages, it is distanced from all of them, except Karakalpak (Fig. 1). The measurement along the scale "**palatal-back**" seems to keep to the same tendency. That is, Hungarian is closer to them than Russian is. Altai moved into the middle of this set of Turkic languages and became closer to Kazah, Baraba Tatar, and Uzbek. Ujgur remained on the outskirts of the Turkic language community. Azeri and Kumandin moved to the outskirts (Fig. 2). The scale "**sonorant – occlusive**" distributes Turkic languages so that Karakalpak, and Kumandin go to the outskirts while Baraba Tatar, Uzbek, Turkish, Kirgiz and Jakut are in the middle. The closest neighbour of Turkish is Uzbek, though it belongs to the Karluk branch of Turkic languages in Baskakov's classification and should be close to Ujgur (Baskakov, 1981: 68). According to this scale (Fig. 3) Hungarian is as close to Ujgur as Kazah. It is much closer to Ujgur than Russian. The scale "**fricative – voiced**" brings Turkish, Kumandin, Kazah and Ujgur into the centre of the set, while Azeri, Baraba Tatar and Altai turn out to go to the outskirts. According to the distribution of the fricative and voiced consonants in the speech sound chain, Hungarian is not close to any of the Turkic languages (Fig. 4). These four phonological scales give a vivid picture of the positions of the selected 13 Turkic languages in relation to each other. However, one should bear in mind that the method of the nearest neighbour used here has some drawbacks. For instance, introducing into this set of languages some more Turkic languages besides Azerbaidjani, Turkish, Tatar, Karakalpak, Ujgur, Kirgiz, Kazah, Turkmen, Uzbek, Hakas, Jakut, Kumandin and Altai, may change this picture of closeness a little bit, but the pattern remains the same, since the majority of the Turkic languages were taken for this phono-typological study.

It is not our aim to verify the phonemic inventories in these 13 Turkic languages. We also took the established classification of the consonantal groups, a more detailed account of them may be found elsewhere (Tambovtsev, 1991, 1994-a; 1994-b). The limited space of the journal does not permit us to give the separate phonemic frequencies of all the 13 Turkic languages. Some of them can be found in our previous work (Tambovtsev, 1991), while the rest are provided in this paper (Tab. 3 – 7). Nevertheless, the total percentage of frequency of occurrence of the four consonantal features defined by the active organ of articulation, three features – by the manner of articulation and one – by the work of the vocal cords are provided for all computed 13 Turkic languages in

question (Tab. 1, 2). The totals of the percentage of groups defined by the manner of articulation and the work of vocal cords are discussed here for the first time and were not provided in the previous work of the author (Tambovtsev, 1991).

Let us do a detailed phonostatistical analysis by every of the 8 features. First of all, it is interesting to see how Turkic languages use the labials, which of the Turkic languages use them similarly and which of these Turkic languages use the greater proportion of the labials, since it is believed that the Turkic languages do not use labials to the same extent that other languages use them, for example Finno-Ugric or Slavonic, or Tungus-Manchurian (Tambovtsev, 1983). It turned out that Turkic languages use labials in the limits of 5.98% – 12,80%. So, the range (R) is equal to 6.82 and the mean range (MR) is equal to 0.52. The lowest use of labials is in Altai and the highest – in Karakalpak. We obtained the following ordered series of the Turkic languages: Karakalpak – Turkish – Ujgur – Azeri – Uzbek – Kumandin – Tatar – Turkmen – Kirgiz – Hakas – Kazah – Jakut – Altai. This series (Tab. 1) shows that in Kirgiz the functioning of labials is similar to Hakas and in Kazah – to Jakut. It may be so that a certain language influenced Hakas and Kazah in some way, but more than Turkish or Karakalpak. At the same time one can notice that Karakalpak and Turkish have a similar typology of distribution of labials, and so do Turkish and Ujgur, Ujgur and Uzbek, Uzbek and Kumandin, etc., that is every neighbouring pair of languages. One can calculate the distances between these 13 Turkic languages to each other on the one hand, and these Turkic languages to any other on the basis of this parameter. We shall show these distances graphically with respect to two parameters: the ratio of labial and front consonants in the sound chain. Turkic labials function within the limits of the interval 5.98% – 12.80%; R = 6.82; MR = 0.52. Should this be considered large or small? Surely, it should be considered large, since 20 Uralic (i.e. Finno-Ugric and Samoyedic) languages have a narrower interval of 7.71% – 13.72%; i.e. R = 6.01 and the mean range (MR) is equal to 0.30, which is much less than that of the Turkic languages. The united Mongolian and Tungus-Manchurian group (7 languages) is distributed in the limit of 7.52% – 12.46%; R = 4.94; MR = 0.71. The three Mongolian languages have the following range and mean range: R = 1.02 and MR = 0.34. Its unity with the Tungus-Manchurian languages makes the group more dispersed. The seven Paleo-Asiatic languages give the interval of 6.43% – 11.34%; R = 4.91; MR = 0.70. Eight Slavonic languages have the labial limit: 12.46% – 16.66%; R = 4.2; MR = 0.52, which may be some sort of measure, since the Slavonic languages are known to have close typological similarity. Therefore, if the mean range is less than 0.52, one can speak of a greater typological similarity, which can be heard even by an untrained ear. One should take into account that the sound chains in the Slavonic languages sound rather similar. The range is a very blunt statistical tool, the mean range is better, but one must use more sophisticated statistical tools, which are more robust and much finer and more exact. This is why the coefficient of variance and the modified "chi-square" criterion are introduced further. The coefficient of variance shows the fluctuation of variables,

so that the less this coefficient, the less the fluctuation. It has a very good property, that is while using it, one should not think about commensurability, when comparing two or more different data. It is possible to compare any values of any sort, which may be distributed according to different statistical laws. The coefficient of variance of the labial consonants of these 13 Turkic languages is 20.54%, which is higher than that of Finno-Ugric languages (14.42%).This means that they fluctuate more than in the Finno-Ugric languages. One should measure the limits of this fluctuation by the "chi-squared criterion". Though the fluctuation of labials in Turkic languages is greater than in the Finno-Ugric languages, it is within the theoretical limits of the homogeneous distribution. Later, we shall see that the distribution of labials in these Turkic languages is rather smooth, that is homogeneous. Measured by the "chi-square" criterion, it gives the value of 4.47, while the critical value for the "chi-squared" criterion is 21.026. It means that there are no values in the Turkic labial series, which should be regarded as "alien". The set of Turkic languages is not as compact as the set of Finno-Ugric languages by the Q-parameter (0.21 and 0.14). The Q-parameter allows us to compare compactness of the sets with different number of members. One must divide the received "chi-square" value by the theoretical value, which could be found in any table of the chi-square distribution (e.g. in Butler, 1985: 176). One should correctly count the degrees of freedom, which is here 12. The number of the degrees of freedom is very important because the critical value of the "chi-square" criterion depends on it. Luckily, it is easy to calculate it, since it is equal to the number of the members minus one (13 – 1 = 12). If one puts Mongolian into the set of these Turkic languages, then its homogeneity increases a bit, that is the value of the coefficient of variance decreases (cf. 20.54% and 20.34%). This means that the labials in Mongolian function similarly to the labials in the Turkic sound chain. In order to prove, that not all languages improve the homogeneity of the set of Turkic languages, let us take a language, which can be supposed not to be influenced by the Turkish language. Mansi (Vogul), one of the Finno-Ugric language family, could be selected as such a language. If one puts it into the Turkic set, then the value of the coefficient of variance becomes greater (cf. 20.54% and 23.50%), and so does the Q-coefficient (cf. 0.21 and 0.29). It means, that the introduction of Mansi into this set of Turkic languages makes it more dispersed, unlike the case with Mongolian.

One can state that front consonants of these Turkic languages use a rather narrow zone of 32.35% – 40.24%; R = 7.89; MR = 0.61. It should be considered to be a small interval since other language families or groups show greater values. In this case, the 20 Uralic languages have a limit within 24.79% – 36.78%; R = 11.99; MR = 0.60.The 7 Mongolian and Tungus-Manchurian languages have a limit within 17.31% – 36.57%; R = 19.26; MR = 2.75.The 7 Paleo-Asiatic languages: 20.01% – 36.74%; R = 16.64; MR = 2.38. The 8 Slavonic languages: 27.42% – 35.46%; R = 8.04; MR = 1.01. However, as indicated earlier, one should use a more sophisticated statistical methods than ranges or means. The coefficient of variance of the front consonants is less than the equivalent coefficient of the labials (6.26% and 20.54%). This means that ac-

cording to the parameter of the occurrence of the front consonants, Turkic languages are more evenly distributed. The value of the degree of homogeneity is also less (Q = 0.08), which indicates the smoothness of distribution. The coefficient of variance of the distribution of the front consonants in Finno-Ugric languages is higher (11.61%).The Turkic languages happen to have the following ordered series, that is from the highest to the lowest: Altai – Hakas – Azeri – Kirgiz – Kazah – Tatar – Turkmen – Karakalpak – Uzbek – Turkish – Jakut – Kumandin – Ujgur. Therefore, by this parameter Altai is very close to Hakas, Azeri to Kirgiz, etc. It is interesting to note that Jakut is close to Kazah and Altai with respect to labials, while with respect to front consonants Jakut is close to Turkish and Kumandin. One can see that Mongolian is not alien to these Turkic languages, because the homogeneity becomes higher, if we introduce Mongolian there. The coefficient of variability (6.26% versus 6.13%) behaves favourably, while the value of Q remains the same, but rather low (0.08 and 0.08), which shows a good level of compactness. It does not happen with the introduction into this set of just any language. If Mansi is introduced, then both coefficients become greater (V = 7.13% and Q = 0.10). So, the compactness of the set becomes less.

Palatal consonants in Turkic are generally in the limits of 5.22% – 0.74%, but the Kumandin language gives a peak of 14.32. Without Kumandin the range is 4.48; MR = 0.37, while with Kumandin – R = 13.58; MR = 0.97. It looks like Kumandin does not belong here by this parameter. In Uralic this limit is 10.46% – 1.33%; R = 9.13; MR = 0.46. In the united group of Mongolian and Tungus-Manchurian languages: 10.51% – 1.48%; in the Paleo-Asiatic group: 11.81% – 1.01%; in the Slavonic group: 9.31% – 1.45%. The palatals in these Turkic language have the greatest coefficient of variance (91.91%). However, the Kumandin value ruins the homogeneity and the value of Q is greater than unity (Q = 1.78). In order to see if the value of palatals in Kumandin is alien, let us calculate the homogeneity of this series without Kumandin. The value of coefficient of variance goes drastically lower and becomes 43.62%, which is more than 2 times less, and Q becomes much less, only 0.29. The ordered series of Turkic languages with respect to palatals, from the highest to the lowest is as follows: Kumandin – Jakut – Turkmen – Azeri – Kirgiz – Uzbek – Turkish – Karakalpak – Kazah – Altai – Tatar – Hakas – Ujgur. Here, Kumandin is rather far away from its neighbour in the ordered series Jakut, while Jakut is close to Turkmen, less close to Azeri, even less close to Kirgiz, etc. and far away from Ujgur. Actually, it is of great importance to consider the frequency of occurrence of palatal consonants because of the classification of Turkic languages by Sergey E. Malov, who divided them into old and new groups on the basis of occurrence of the palatal phoneme /j/. In most Turkic languages that were taken for this phonostatistical investigation the only palatal consonant is /j/: for example in Altai, Hakas, Kirgiz, Uzbek, Karakalpak, Kazah, Turkmen, Uygur, etc. In those Turkic languages, where it is not the only palatal consonant, as in Azeri (72.33%) and Turkish (99.27%) the palatal consonant /j/ comprises the majority of the total occurrence of the palatal consonants. In Jakut (47.11%) it is nearly

half of the occurrence and in Baraba Tatar it is less than that (34.54%). Malov's criterion of the division of the Turkic languages into old and new groups is the occurrence of /j/ in the new languages, where in the old ones some other sounds occur. From the phonostatistical point of view, Malov's "new Turkic languages" should have a greater frequency of occurrence of /j/. According to Malov, the new Turkic languages are: Azeri, Altai, Bashkir, Gagauz, Kazah, Karaim, Karakalpak, Kirgiz, Kumandin, Kumyk, Kyochak, Nogay, Oyrot, Pecheneg, Polovets, Salar, Tatar, Turkish, Turkmen, Uzbek, Ujgur, Chatagay, and Chulym. To the group of "old Turkic languages" S. E. Malov put the following languages: Chuvash, Bulgarian, Hazarian, Runic-Oguz, Proto-Uygur, Tofalar, Proto-Chatagay, Hakas, Shor, and Yellow-Uygur (Malov, 1952: 139). Therefore according to S. E. Malov, Hakas should be in the group with low occurrence of palatals, while Azeri, Altai, Kazah, Karakalpak, Kirgiz, Tatar, Turkish, Turkmen, Uzbek, and Ujgur. Our data on the frequency of occurrence of palatals also show that Hakas has a very low frequency of occurrence of /j/.

From the point of view of the distribution of back consonants in Turkic languages it is notable that Ujgur takes the first place with 13.47% while Azeri is in the last place with 6.84%, R = 6.63; MR = 0.51. The introduction of Mongolian leaves the range the same and makes MR smaller (0.47). It tells us that the set becomes more compact with Mongolian. In Uralic languages back consonants have a limit from 17.48% – 7.44%; R = 10.04; MR = 0.50. In the united group of Mongolian and Tungus-Manchurian languages the interval is 12.81% – 8.50%; R = 4.31; MR = 0.62; in the Paleo-Asiatic group: 23.21% – 13.66%; R = 9.55; MR = 1.36; in the Slavonic group: 6.87% – 5.63%; R = 1.24; MR = 0.15. So, the limit zone for functioning of the back consonants in the united group of Mongolian and the selected Turkic languages may not seem very narrow, if one takes into account the Slavonic data. The coefficient of variance of Turkic languages from the point of view of the distribution of back consonants is very small (15.11%). The value of Q also shows great homogeneity (Q = 0.15). It is much smaller than unity, which means that the set of Turkic is very homogeneous. It is even smaller than the analogical parameter for the Finno-Ugric languages (V = 18.67%; Q = 0.19). One can claim that from the point of view of the distribution of the back consonants, the set of Turkic languages is more compact than the set of the Finno-Ugric languages. In the ordered series, we have: Azeri – Kirgiz – Karakalpak – Turkmen – Turkish – Baraba Tatar – Uzbek – Altai – Kumandin – Kazah – Hakas – Jakut – Ujgur. Turkmen has a similar distribution of back consonants to Turkish and Turkish to the language of the Baraba Tatar, who live in Siberia and seem to have had contacts with the ancient Turks. It is not surprising that Turkmen is close to Turkish.

Sonorant consonants in the Turkic languages function in the interval of 30.68% – 18.65%; R = 12.03; MR = 0.92. The introduction of Mongolian makes the united group more compact (MR = 0.86). Now let us consider the limit zones for sonorant consonants in other languages in the Uralic group: 34.76% – 17.45%; R = 17.31; MR = 0.87; in the united Mongolian and Tungus-Manchurian group: 28.32% – 20.05%; R = 8.27; MR = 1.18; in the Paleo-Asiat-

ic group: 32.63% – 19.11%; R = 13.52; MR = 1.93; and in our standard group – Slavonic: 25.06% – 20.25%; R = 4.81; MR = 0.60. The united group of Mongolian and Turkic languages have the following ordered series with respect to sonorants (from the highest to the lowest value in the speech chain): Karakalpak – Turkmen – Azeri, Tatar – Uzbek – Turkish – Jakut – Kirgiz – Hakas – Altai – Mongolian – Kazah – Ujgur – Kumandin. Thus, Mongolian, Kazah and Altai speech chains have a similar manner of using sonorants. Jakut is close to Turkish and Kirgiz. So, it seems that judging by a new parameter, Jakut is ever closer to a new language. We shall discuss this phenomenon in detail later. The coefficient of variance and Q show that from the point of this feature, the set of Turkic languages is rather compact (V =12.03%; Q = 0.21). Turkic languages are more compact in this feature than Finno-Ugric (V = 18.08%; Q = 0.48) languages. Introducing Mongolian into the set of Turkic languages, makes this set more compact (V = 11.69%; Q = 0.20). Again one can claim that Mongolian is not alien to the Turkic languages.

With respect to occlusive consonants, Turkic languages function in the limit zone of 37.26% – 18.94%; R = 18.32; MR = 1.41. The introduction of Mongolian into this group, makes this group more compact (MR = 1.31). In the Uralic languages it is 32.50% – 16.35%; R = 16.15; MR = 0.81. In the united group of Mongolian and Tungus-Manchurian languages: 25.46% – 21.72%; R = 3.74; MR = 0.53; in the Paleo-Asiatic group: 28.21% – 21.21%; R = 7.00; MR = 1.00; in the Slavonic group: 21.91% – 14.96%; R = 6.95; MR = 0.87. So, one can see that the mean limit zone of functioning of occlusives in the united Mongolian and Turkic group is greater than that in the Mongolian and Tungus-Manchurian, Paleo-Asiatic or Slavonic groups, though without Mongolian the Turkic group is even more dispersed. The united group has the following ordered series with respect to occlusive consonants: Kumandin – Altai – Kirgiz – Ujgur – Tatar – Uzbek – Kazah – Turkish – Mongolian – Jakut – Karakalpak – Hakas – Turkmen – Azeri. Again, Mongolian is close to some other Turkic languages: this time Turkish and Jakut, while Jakut is also close to Karakalpak. Judged by this feature, the Turkic languages are not very compact (V =20.33%; Q = 0.57). Finno-Ugric languages are more compact here (V = 13.56%; Q = 0.23). The introduction of Mongolian makes the group more compact (V = 19.70%; Q = 0.54). By this parameter, Mongolian also belongs to the group of Turkic languages.

Functioning of fricative consonants define the zone of 13.66% – 5.73%; R = 7.93; MR = 0.61. It means that among all the Turkic languages, one cannot find a language, whose speech chain has more fricative consonants than 13.66% and less than 5.73% with respect to the total number of all phonemes in the sound chain of a certain language. The introduction of Mongolian makes this group more compact (MR = 0.57). The number of fricatives in the phonemic chains of the Uralic languages is 16.91% – 4.58%; R = 12.33; MR = 0.62; in the Mongolian and Tungus-Manchurian languages: 18.76% – 2.94%; R= 15.82; MR = 2.26; in the Paleo-Asiatic languages: 17.64% – 3.44%; R = 14.20; MR = 2.03; in the Slavonic languages: 17.43% – 9.44%; R = 7.99; MR = 1.00. So, the frica-

tive consonants in the united group of Mongolian and Turkic languages function in a very narrow limit zone. The ordered series with respect to fricative consonants is the following: Hakas – Kumandin – Azeri – Jakut – Kazah – Mongolian – Ujgur – Turkish – Uzbek – Karakalpak – Turkmen – Tatar – Kirgiz – Altai. As we can see from this series, Mongolian is close to Kazah and Ujgur, while Jakut is close to Azeri and Kazah. By this parameter the Turkic languages (V = 22.16%; Q = 0.27) are less compact than the Finno-Ugric languages (V = 14.57%; Q = 0.17). The introduction of Mongolian into the group of Turkic languages makes this group more compact (V = 21.32% and Q = 0.26). One can state by this parameter that Mongolian clearly belongs to this group.

The 8th feature is the function of voiced consonants, which give the limit zone of 15.00% – 7.22%; R = 7.78; MR = 0.60. In the Uralic languages this limit zone is 13.00% – 0.00%; R = 13.00; MR = 0.65; in the Mongolian and Tungus-Manchurian languages: 19.99% – 9.91%; R = 10.08; MR = 0.76; in the Paleo-Asiatic languages: 16.37% – 0.00%; R = 16.37; MR = 2.34; in the Slavonic languages: 26.48% – 19.23%; R = 7.25; MR = 0.91. The ordered series of Turkic languages with respect to voiced consonants is as follows: Turkmen – Azeri –Kirgiz – Uzbek – Altai – Kazah – Turkish – Hakas – Ujgur – Kumandin – Jakut – Karakalpak – Tatar. Here, Jakut is close to Kumandin and Karakalpak. The compactness by this feature is not very great (V = 22.18%; Q = 0.32). However, the Finno-Ugric languages are even less compact (V =46.21%; Q = 1.11). Measuring the closeness of languages in the scale of order is not so informative as measuring it in the numerical scale. Depicting all the eight features numerically in the graphical form (Fig. 1– 4) will give us a clearer picture of the closeness of these 13 Turkic languages. Moreover, the ordered series and the numerical scales of Turkic languages can give some idea of the general place for them among the other world languages.

The graphical representation of these 13 Turkic languages, which are in fact the main Turkic languages, may help us to establish a new classification of Turkic languages based on the sound picture they create in speech. Actually, the distances between the Turkic languages may create a true classification from the point of view of dynamic phonology which investigates not only the true phonemic inventory of a language, but also the sound picture of a language created by the frequency of each phoneme resulting in the frequencies of phonemic groups. It is important to stress that our four graphs are built on the values of features that do not depend on each other. In the theory of classification (Rozova, 1986: 30 – 31), it is quite necessary to have such features that are the most informative, not correlated to the others and not derived from the other features (Tambovtsev, 1994-a; 1994-b). We considered the 8 consonantal features to be the most efficient since they involve all the three possible points of view basing on the articulatory characteristics.

If one takes into consideration functioning in the Turkic sound chain the features of "labiality" (i.e. the state of being labial) and "frontality" (i.e. the state of being front), one can find out that Jakut, Altai, Ujgur and Karakalpak are far away from the rest of the eight Turkic languages. Jakut and Altai use the labial

consonants in a very low proportion. Nevertheless, Jakut is different from Altai in the way the front consonants are used: Jakut use them very little (like Ujgur), while Altai uses the greatest proportion of them, in fact, more than any of the other eleven Turkic languages. Hakas and Kirgiz are very close to each other. Azeri approaches them, though it is also close to Tatar, though to a lesser extent. Kazah, Turkmen, Tatar, Uzbek and Turkish form another compact group, though Turkish has a tendency towards Ujgur and Karakalpak, while Kazah tends to be similar to Jakut, Turkmen or Kirgiz. It must be pointed out that the precise graph is not locked. This means that it is not possible to travel from one Turkic language to another without travelling some certain distances twice. It also means that the set of these 12 languages is not so compact as it could be if the scheme were convex (Fig. 1).

Let us take into consideration the functioning of such features as palatal and back consonants frequencies. One can see that again Jakut and Ujgur are at the periphery of these Turkic languages while Kirgiz, Karakalpak, Turkish, Uzbek, Kazah, Altai and Tatar form a compact set of languages. Ujgur, Hakas and Jakut overexploit the back consonants while Azeri overexploits palatal consonants. The graph is neither locked nor convex. One can understand from Figs. 1 and 2 that by these four parameters (labial, front, palatal, back) which show the active organ (or place of articulation) that Turkish is close to Uzbek and Karakalpak, Hakas and the other languages by two parameters are close to certain languages and by the two other features to other languages different from the previous ones; for example by the first two parameters Hakas is close to Kirgiz and Altai, while two second parameters bring Hakas into the vicinity of Ujgur and Tatar. Later, we shall try to explain this phenomenon.

The functioning of sonorant and occlusive consonants gives the following chain of Turkic languages (Fig. 3): Karakalpak, Turkmen, Azeri, Hakas, Jakut, Uzbek, Tatar, Turkish, Kazah, Ujgur, Kirgiz and Altai. Karakalpak and Turkmen are the most sonorant while Kazah and Ujgur are the least sonorant. Altai and Kirgiz are the most occlusive, Azeri and Turkmen are the least occlusive. Turkish is close to Tatar and Uzbek, Jakut is close to Uzbek and Hakas. All in all the aggregation is not compact, languages do not centre around some compact and convex set of languages. The schemes (Figs. 1 and 2) showing the consonantal active organ are more compact. The following scheme built on the parameters of fricativeness and voicedness is also more compact (Fig. 4). Hakas, Tatar, Azeri and Altai are on the periphery. Turkish is close to Ujgur and Kazah; Jakut – to Ujgur. The set of Turkic languages by these two parameters is not convex and is not very compact.

We can see from these four schemes that some of the languages, like Jakut or to some extent Altai, tend to sound differently from the other ten Turkic languages. Actually it is often stated in Turkic linguistics that Jakut takes a special place in the Turkic language family since in the process of its development Jakut has been strongly influenced by the Tungusic and Mongolic languages (Musaev, 1984: 12; Ubrjatova, 1982: 4-6). The sound picture of the Jakut language is rather different from the other 11 Turkic languages, as if Jakut has

drifted away from them. Actually, our data reflect the phono-statistical difference of Jakut, which developed on the outskirts of the Turkic world. Mongolian ousted Jakut and Tungusic tribes, which separated the Jakuts from the other, related Turkic peoples (Ubrjatova, 1960: 3-7). The fact that the Turkic tribes mixed a lot during their historical development may give the answer to the problem of why one cannot see a clear-cut picture of similarity everywhere. The picture is very vague because by one parameter one language is close to the second, and by the next parameter, it is close to the third, and so on. This is why it is very hard to construct only one true classification of Turkic languages like those produced in Finno-Ugric, Samoyedic or Tungus-Manchurian linguistics (Tambovtsev, 1994-a; 1994-b).

At present, Turkic linguistics has at least 15 classifications of Turkic languages based on different features and thus yielding different results: those of I. N. Berezin, W. Radloff, F. E. Korsh, A. N. Samojlovich, I. Adelung, A. Palmblad, A. A. Balbi, G. Klaproth, I. Hammer, G. Ramstedt, M. Raesaenen, V. Bogoroditzky, S. E. Malov, N. A. Baskakov and L. Ligeti. The present typological analysis from the point of view of phono-statistics may help to establish another modern classification. This work is an introduction to a new field of linguistic investigation called dynamic philology (Zipf, 1935) which may help Turkic linguistics define basic features and reduce the 15 classifications to one basic classification, which would reflect phonetical and other similarities between Turkic languages. Our new approach gives a true picture of the distances between the Turkic languages each of which took its choice and chance in the course of phonological development, as every world language does (Herdan, 1966). One cannot help agreeing with a well-known Turkologist A. N. Kononov who used to emphasize that all proposed classifications (even those proposed recently) cannot represent clearly the phonetic and grammatical structure of all Turkic languages by uniting them in several groups (Kononov, 1982: 322). Many other scholars who worked in Turkology (Bogoroditskiy, 1953; Korsh, 1910; Samoylovich, 1922; Shcherbak, 1970; Baskakov, 1981; Klimov, 1971; etc.) stated and it is a common place in Turkology that during thousands of years the Turkic languages mixed a lot and this resulted in very similar phonological systems and lexis (Musaev, 1984: 33). N.Z. Gadjieva considers such convergence processes between Turkic languages as very common (Gadjieva, 1980: 120). N. A. Baskakov remarks that the classification of Turkic languages into the groups of Oguz, Kypchak, Karluk, Ujguro-Oguz, Bulgarian and others is rather conventional (Baskakov, 1981: 60–69). G. A. Klimov and A.M. Shcherbak think that the long period of mixing of the Turkic languages led to some new similarities between them (Klimov, 1971: 11; Shcherbak, 1970: 20). E. N. Nadjip believed that the Turkic protolanguage had mixed Ujguro-Oguzo-Kypchak features (Nadjip, 1970: 87). Many specialists in Turkic linguistics think that it is the great similarity which hinders the construction of a true classification of the Turkic languages. They also think that it is hard to apply the comparative method to the Turkic languages because of their similarity (e.g. Musaev, 1984: 32). Another well-known Turkologist A. M. Shcherbak takes into

account the fact that similarity was caused by a double process, when the genealogical similarity was added to the later typological similarity, as the development of languages is not a pure process of splitting of the protolanguages in the line of Schleicher's genealogical tree. It is a more complex process where divergence and convergence of languages and dialects go together resulting in a mixture of different genealogical and typological features, making the borders muzzy and blurred (Shcherbak, 1970: 17–22). A well-known specialist in ethnography L. N. Gumiljov studied the history of the ethnographical development of the ancient Turks, Ujgurs, Oguz, Kypchaks etc. for 1,500 years and came to the conclusion that scholars should take into account the successions of zigzags and impulses in the development of some peoples (Gumiljov, 1993). Unfortunately, linguists and ethnographers often work separately which hinders their understanding. In my view, the ethnographical data may explain some linguistical changes. According to L. N. Gumiljov the way of life of certain people may be influenced by the adaptation of these people to the landscape. On the basis of this idea we may understand why the Turkic languages have similar phonological systems: the peoples who spoke Turkic languages were nomadic tribes who moved very fast on the vast steppe regions. This meant that their phonological system had to provide stable and good communication. A phonological process of this kind may require either the same or similar sound types, which should be clearly distinguishable from each other. In order to communicate and to be understood a language should have easily recognizable units (Tambovtsev, 1994-a; 1994-b). So, in the course of its development the phonological system of a language chose "good" phonemes which were easy to recognize (Tambovtsev, 1976). In any language the phonemic frequencies are in a state of dynamic homeostasis, that is the state of equilibrium produced by a balance of phonological functions and of lexical composition within a certain period of time (Tambovtsev, 1994-a; 1994-b).

Let us consider in detail what was said about this or that Turkic language regarding its place in relation to other Turkic languages. Discussing the Azeri (or Azerbaidjani) language N.Z. Gadjieva remarks that one can put it in the group of Oguz languages only conventionally (Gadjieva, 1966: 66). Actually, Azeri has some features which are common to such Oguz languages as Turkish, Turkmen and Gagauz but at the same time Azeri has some features common to the Kypchak group of languages (Gadjieva, 1979: 175 – 180). In our schemes Azeri is not close to either Turkish or Turkmen. It always stands apart from the rest of the 11 Turkic languages (cf. Fig. 1 – 3). Altai is considered to be a fair Kypchak language related to Kirgiz, though the Northern Altai dialects belong to the Oguzo-Ujgur group (Baskakov, 1966: 506; 1978: 38). In our phonological schemes Altai is close to Kirgiz by palatal, back, sonorant, occlusive, fricative and voiced consonantal characteristics. L.V. Dmitrieva believes Tatar (Siberian) is close to Tatar (Volga), Bashkir, Kazah and Altai (Dmitrieva, 1966: 155). D.G.Tumasheva considers Tatar (Siberian) to be closer to Kypchak, though having Oguz and Karluk features (Tamasheva, 1968), that is Bashkir, Kazah, Karakalpak and Nogai. Our schemes show that the language of the Siberian Tatars

is close to Karakalpak only with respect to fricative and voiced consonants, but with respect to palatal and back consonantal characteristics Tatar is close to Altai and Kazah. Turkish is said to enter the Oguz group of Turkic languages, though it has Kypchak features on all levels: phonetic, morphological and syntactic (Gadjieva, 1979: 180). Modern Ujgur is said to enter the Karluk group of Turkic languages (Kaydarov, 1966: 366), where Uzbek is also placed. V.G. Karpov and A.H. Baskakov believe Hakas belongs to the Oguzo-Ujgur group together with Tuvin and Jakut (Karpov, 1966: 428); Baskakov, 1981: 68 – 69). E.I. Ubrjatova considers Jakut to be in the Oguz group while N.A. Baskakov places it in the Oguzo-Ujgur group (Ubrjatova, 1960: 2 – 12; 1966: 403; Baskakov, 1981: 60 – 69). If E.I. Ubrjatova is correct, then Jakut should be close to Turkish or Azeri in our schemes, but it is not (cf. Figs. 1 – 3). The fact the Jakut is always on the outskirts of the set of these Turkic languages does not allow us to accept the assumption of E. I. Ubrjatova that Jakut preserved the features of the ancient Turkic languages. The point of view of E. N. Nadjip who believes that the ancient proto-Turkic language was a good mixture of all the modern Turkic languages, that is all their features, so it should be somewhere in the centre of our schemes. According to our data, it is Turkish, Tatar, Kazah or Kirgiz. However, we can speak about the ancient state of the Turkic languages only conventionally since we cannot measure to what extent they changed later under the influence of each other and other languages. This is why we cannot help agreeing with A. N. Kononov who does not consider any of the above mentioned 15 classifications of Turkic languages to be satisfactory (Kononov, 1982: 322). This means that our phono-typological analysis may help to improve one of the 15 classifications or to establish a new one, which could be accepted as true.

Let us try to reduce the abundance of distances from different points of view to the total of such distances. It is possible to do this with the help of the formula calculating the Euclidean distance. It was adapted to our purposes in the following form:

D (distance) = to the square root of – (Lb1 – Lb2) squared + (Fn1 – Fn2) squared + (Pt1 – Pt2) squared + (Bk1 – Bk2) squared + (Sn1 – Sn2) squared + (Oc1 – Oc2) squared + (Fc1 – Fc2) squared + (V1 – Vd2) squared

Where

Lb1 – is the frequency of occurrence of labial consonants in the first language;

Lb2 – is the frequency of occurrence of labial consonants in the second language;

Fn1 – is the frequency of occurrence of front consonants in the first language;

Fn2 – is the frequency of occurrence of front consonants in the second language;

Pt1 – is the frequency of occurrence of palatal consonants in the first language;

Pt2 – is the frequency of occurrence of palatal consonants in the second language;

Bk1 – is the frequency of occurrence of back (velar) consonants in the first language;
Bk2 – is the frequency of occurrence of back (velar) consonants in the second language;
Sn1 – is the frequency of occurrence of sonorant consonants in the first language;
Sn2 – is the frequency of occurrence of sonorant consonants in the second language;
Oc1 – is the frequency of occurrence of occlusive consonants in the first language;
Oc2 – is the frequency of occurrence of occlusive consonants in the second language;
Fc1 – is the frequency of occurrence of fricative consonants in the first language;
Fc2 – is the frequency of occurrence of fricative consonants in the second language;
Vd1 – is the frequency of occurrence of voiced consonants in the first language;
Vd2 – is the frequency of occurrence of voiced consonants in the second language.

Having computed the data of the 8 features (Tab. 1 and 2), I found the united distances between these Turkic language. So, one can see that Azeri sounds very much like or similar to Turkmen (6.62), Uzbek (7.39), Turkish (7.89) or Kazah (8.78). At the same time, Kumandin (24.04), Altai (15.80), Ujgur (11.52) or Baraba Tatar (10.83) do not sound very similar to Azeri, while Kirgiz (8.89), Hakas (8.89), Karakalpak (9.47) and Jakut (10.75) are in the intermediate position (cf. Tab. 9). In the same way one can see the distances between these 12 Turkic languages and Altai (Tab. 10), Kazah (Tab. 11), Karakalpak (Tab. 12), Kirgiz (Tab. 13), Baraba Tatar (Tab. 14), Turkish (Tab. 15), Turkmen (Tab. 16), Ujgur (Tab. 17), Uzbek (Tab. 18), Hakas (Tab. 19), Jakut (Tab. 20) and Kumandin (Tab. 21). It is also possible to range all the distances between these 13 Turkic languages (Tab. 22), in order to see the scheme of closeness. Actually, the closest are Uzbek and Turkish (2.99), Uzbek and Kirgiz (4.15), Uzbek and Kazah (4.15), Turkish and Kazah (4.24), Turkish and Ujgur (4.41), and Ujgur and Kazah (4.790). The greatest distances are Kumandin – Karakalpak (24.05) and Kumandin – Azeri (24.07), see Tab. 23. It is also possible to consider the position of each of the 13 Turkic languages in the centre or outskirts of the Turkic group of languages judged from the point of view of the totals of the distances. It is assumed that the less the sum of the distances, the more central a position the language takes. Turkish happened to take the most central position with the total of the distances of 89.39. Kazah (93.15) and Uzbek (93.31) are also not far away from the centre while Altai (158.78) and especially Kumandin (249.37) are really on the outskirts (Tab. 23). It appears that Kumandin was very much influenced by the substratum language or languages, which were, in fact, of Finno-Ugric or Samoyedic origin. Irina J. Sel'utina came to this conclusion

on the basis of the peculiar Kumandin vowel and consonant data, obtained by the methods of experimental phonetics. The peculiar type of the Kumandin language was stressed by a number of linguists (Sel'utina, 1983; 1998). However, it may be so that Kumandin was influenced by an unknown language which also influenced Japanese, if one follows the hypothesis of the Altaic origin of Japanese. It was extremely interesting for me to find out that the Japanese language is the closest to the Altaic languages, and among them Kumandin (23.30) was the closest. Japanese is not so close to other languages. Our next investigation shall be devoted to measuring the phonostatistical distances between the rest of the computed Ural-Altaic languages. In my opinion it is necessary to analyse the similarity of all languages of the world on the basis of the phonetico-typological scales, introduced by the author. The investigation of this sort may give the first clues for further interesting linguistical discoveries.

In order to find some standard for assessing the distances, it is necessary to take two closely related languages. In this case they are Mongolian and Burjat. Using this standard, one can assess the distances correctly. The closest sound picture to Mongolian is created by Uzbek (5.88) and the least similar – by Karakalpak (13.68). Kirgiz (6.62), Kazah (7.07) and Turkmen (8.39) are also close to Mongolian. Azeri (8.42), Turkish (8.47), Hakas (9.12) and Ujgur (10.27) are not so similar in the way they sound. The less similar are Jakut (11.11), Altai (11.23) and Baraba Tatar (12.20).

Material

Consonants were grouped according to the following phonetic charts:

1) Mongolian – Sanzheev,1960; Kasjanenko, 1968; Todaeva, 1981; Nadel'aev, 1985; 1987;

1) For Altai – Baskakov, 1966; 2) Azeri (Azerbaidjani) – Gadjieva, 1966; 3) Hakas – Chankov, 1957:9; 4) Jakut – Ubrjatova, 1982:12–37; Barashkov, 1953:26–49; 5) Karakalpak – Baskakov, 1966:303; 6) Kazah – Kenesbaev et al., 1966:320–321; 7) Kirgiz – Yunusaliev, 1966:484; Ahmatov, 1968:191; Batmanov, 1953:78; 8) Kumandin – Sel'utina, 1983; 1998; 9) Tatar (Baraba) – Tumasheva, 1977; Dmitrieva, 1966:157–158; 1981:4–7; 10) Turkish – Sevortian, 1955:32–44; 11) Turkmen – Azimov et al., 1966:93; 12) Ujgur – Talipov, 1984:6–18; Borovkov, 1935:6–19; Najip, 1954:12; 13) Uzbek – Reshetov, 1966:343.

The limit of the journal space does not permit us to give details on the material (texts) and the phonemic frequency data of the Turkic languages, selected for the investigation, but the information can be easily found elsewhere (Tambovtsev, 1991). The author expresses his great gratitude for the advice, consultations and materials on the Turkic languages to the late Vladimir Mihailovich Nadel'aev, a well-known researcher of Mongolian and Turkic languages.

REFERENCES

Ahmatov T.K. Zvukovoy stroy sovremennogo kirgizskogo yazyka. – Frunze: Mektep, 1968. –191 pp.

Azimov P., Amansariev D., Sariev K. Turkmenskiy yazyk. – In: Yazyki narodov SSSR. Tyurkskie yazyki (YNSTU), t. 2. – Moskva: Nauka, 1966, pp. 91–111.

Bailey, R. W. Authorship attribution in forensic setting. – In: D.E. Ager, F.E. Knowles, and J. Smith (eds.) Advances in Computer-aided Literary and Linguistic Research. – Birmingham: AMLC, 1979.

Barashkov P.P. Zvukovoy stroy yakutskogo yazyka. – Yakutsk: Saha sir.; 1953. – 98 pp.

Baskakov N.A. Altaiskiy yazyk. – YANSTU, t. 2. – Moskva: Nauka, 1966, pp. 506–522.

Baskakov N.A. Arealy i marginal'nye zony v razvitii tyurkskih yazykov. – In: Narody i yazyki Sibiri. Areal'nye issledovaniya. – Moskva: Nauka, 1978, pp. 37–46.

Baskakov N.A. K istoriko-tipologicheskoy fonologii tyurkskih yazykov. – Voprosy yazykoznaniya, N 1, 1981, pp. 60–69.

Batmanov I.A. Sovremennyi kirgizskiy yazyk. Fonetika. – Frunze: Mektep, 1953. – 189 pp.

Bektaev K.B. Statistika fonemosochetaniy v karakalpakskom yazyke. – In: Statistika kazahskogo teksta. Trudy gruppy "Statistiko-lingvisticheskoe issledovanie i avtomatizatsiya", Volume 3. – Alma-Ata: Nauka, 1973, pp. 635–642.

Behtlingk O. Über die Sprache der Jakuten, 1964.

Borovkov A. Uchebnik uygurskogo yazyka. – Leningrad: LVI, 1935. – 247 pp.

Bogoroditskiy V.A. Vvedenie v tatarskoe yazykoznanie v svyazi s drugimi tyurkskimi yazykami. – Kazan': Gosizd, 1953.

Bondarko L.V. , Zinder L.R. O nekotoryh differentsial'nyh priznakah russkih soglasnyh fonem. – In: Voprosy yazykoznaniya, N 2, 1966.

Butler Ch. Statistics in Linguistics. – Oxford: Basil Blackwell, 1985. – 214 pp.

Chankov D.I. Soglasnye hakasskogo yazyka. – Abakan: HNIIYAL, 1957. – 107 pp.

Cowan H.K.J. Statistical determination of linguistic relationship. – In: Studia Linguistica, 16, 1962, pp. 57–96.

Crystal D. A First Dictionary of Linguistics and Phonetics. – London: Andre Deutsch, 1980. – 390 pp.

Czekanowski J. Zur Differentialdiagnose der Neandertalgruppe. – Korblatt Deutsch. Ges. Anthrop., XL, 1909, pp. 44–47.

Czekanowski J. Wstęp do Historii Słowian. – Poznań: Instytut Zachodni, 1957. – XIX + 513 pp.

Djevdet-zade H., Kononov A. Grammatika sovremennogo turetskogo yazyka. Fonetika. – Leningrad: Gosizd, 1934. – 268 pp.

Dmitriev N.K. Turetskiy yazyk. – Moskva: Vost. lit., 1960. – 95 pp.

Dmitrieva L.V. Yazyk barabinskih tatar. – In: Yazyki narodov SSSR. Tyurkskie yazyki (YANSTU), t. 2, pp. 157–158.

Dmitrieva L.V. Yazyk barabinskih tatar. Materialy i issledovaniya. – Leningrad: Nauka, 1981. – 225 pp.

Gadjieva N.Z. Azerbaydjanskiy yazyk. – YANSTU, t. 2, pp. 66–90.

Gadjieva N.Z. Tyurkoyazychnye arealy Kavkaza. – Moskva: Nauka, 1979, 263 pp.

Gadjieva N.Z. K voprosu o klassifikatsii tyurkskih yazykov i dialektov. – In: Teoreticheskie osnovy klassifikatsii yazykov mira (TOKYM). – Moskva: Nauka, 1980, pp. 100–126.

GREENBERG J. H. The nature and uses of linguistic typologies. – In: International Journal of American Linguistics. Vol. 23, N 2, April, 1957, pp. 68–77.

GUMILJOV L. N. Drevnie tyurki. – Moskva: Nauka, 1967. – 504 pp.

GUMILJOV L. N. Drevnyaya Rus' i Velikaya Step'. – Moskva: Mysl', 1989. – 764 pp.

GUMIJOV L. N. Tysyacheletie vokrug Kaspyja. – Moskva: Mishel i K, 1993. – 336 pp.

GUMILJOV L. N. Orda. Proishozhdenie sobytiy. – "Arabeski" istorii (ed. A.I. Kurkchi). – Moskva: Tanais, 1994, pp. 18–24.

HERDAN G. The Advanced Theory of Language as Choice and Chance. – Berlin: Springer-Verlag, 1966. – 459 pp.

ITKONEN E. Qualitative vs Quantitative Analysis in Linguistics. – In: PERRY Th. Evidence and Argumentation in Linguistics. – Berlin, 1980, pp. 336–366.

KAYDAROV A.T. Uygurskiy (novouygurskiy) yazyk. – In: Yazyki narodov SSSR. Tyurkskie yazyki (YANSTU), t. 2, pp. 363–386.

KARPOV V.G. Hakasskiy yazyk. – YANSTU, t. 2, pp. 428–445.

KASJANENKO Z. K. Sovremennyy mongol'skiy yazyk. – Leningrad: Leningrad University Press, 1968. – 112 pp.

KENESBAEV S. K., KARASHEVA N. V. Kazahskiy yazyk. – YANSTU, t. 2, pp. 320–339.

KENNY A. Stylometric Study of the New Testament. – Oxford: Clarendon Press, 1986. – 124 pp.

KLIMOV G.A. Voprosy metodiki sravnitel'no-geneticheskih issledovaniy. – Leningrad: Nauka, 1971. – 86 pp.

KONONOV A.N. Grammatika sovremennogo turetskogo literaturnogo yazyka. – Moskva – Leningrad: AN SSSR, 1956. – 569 pp.

KONONOV A.N. Istoriya izucheniya tyurkskih yazykov v Rossii. – Leningrad: Nauka, 1982. – 360 pp.

KRUPA V. A Typological Comparison of Polynesian Phonetic Systems. – In: Beiträge zur Linguistik und Informationsverarbeitung, N 20, 1970, pp. 58–92.

KORSH F.E. Klassifikatsiya turetskih plemen po yazykam. – Etnograficheskoe obozrenie, 1910, t. 1-2.

LADEFOGED P. Preliminaries to Linguistic Phonetics. – Chicago: The University of Chicago Press, 1971. – 122 pp.

LEDGER G.R. Re-counting Plato. A Computer Analysis of Plato's Style. – Oxford: Clarendon Press, 1989. – 254 pp.

MAJR E. Printsipy zoologicheskoy sistematiki. – Moskva: Mup, 1971. – 454 pp.

MALOV S. E. Drevnie i novye tyurkskie yazyki. – Izvestiya AN SSSR, OLYA, t. 2, 1952, 2–139 pp.

MUSAEV K.M. Leksikologiya tyurkskih yazykov. – Moskva: Nauka, 1984. – 228 pp.

NADEL'AEV V. M. Sostav fonem v zvukovoy sisteme sovremennogo mongol'skogo yazyka. – In: Fonetika sibirskih yazykov. – Novosibirsk: Institut IFF, 1985, pp. 3–24.

NADEL'AEV V.M. U istokov tuvinskogo yazyka. – In: Issledovaniya po tuvinskoy fonologii. – Kyzyl: IJL, 1986, pp. 53–63.

NADEL'AEV V. M. Sovremennyy mongol'skiy yazyk. – Yazyki Sibiri i Mongolii. – Novosibirsk: Institut IFF, 1987, pp. 3–74.

NADJIP E.N. Uygurskiy yazyk. – Moskva: MGU, 1954. – 191 pp.

NADJIP E.N. O srednevekovyh literaturnyh traditsiyah i smeshannyh pis'mennyh yazykah. – Sovetskaya tyurkologiya, 1970, N 1, pp. 87–89.

RESHETOV V.V. Uzbekskiy yazyk. – Yazyki narodov SSSR. Tyurkskie yazyki, t. 2. – Moskva: Nauka, 1966, pp. 340–362.

RADLOFF W. Die jakutische Sprache in ihrem Verhaeltnisse zur den Turksprachen. – St. Peterburg, 1908.

RASSADIN V.I. Rol' kontaktov v obrazovanii tyurko-mongol'skoy yazykovoy obshchnosti. /The role of contacts in forming Turkic-Mongolian language unity/. – In: Voprosy yazykoznaniya, 4, 1988, pp. 103–108.

ROZOVA S.S. Klassifikatsionnaya problema v sovremennoy nauke. – Novosibirsk: Nauka, 1986. – 223 pp.

SAMOYLOVICH A.N. Nekotorye dopolneniya k klassifikatsii turetskih yazykov. – Petrograd: PGU, 1922. – 15 pp.

SEVORTYAN E.V. Fonetika turetskogo literaturnogo yazyka. – M.: AN SSSR, 1955. – 152 pp.

SEL'UTINA I. J. Kumandinskiy konsonantizm. Eksperimental'no-foneticheskoe issledovanie. /The System of Consonants of the Kumandin Language. Experimental Phonetical Investigation/. – Novosibirsk: IIFF, 1983. 238 pp.

SEL'UTINA I. J. Kumandinskiy vokalizm. Eksperimental'no-foneticheskoe issledovanie. /The System of Vowels of the Kumandin Language/. – Novosibirsk: Sibirskij Hronograf, 1998. – 184 pp.

SHIROBOKOVA N. N. O yakutsko-mongol'skih kontaktah /On Yakut-Mongolian Contacts/. – In: Narody i yazyki Sibiri. – Novosibirsk: Nauka, 1980, pp. 140–147.

TALIPOV T. Uygur tili fonetikisi. – Almuta: Mektep, 1984. – 128 pp.

TAMBOVTSEV Yu.A. Raspoznavanie fonem chelovekom. – In: Empiricheskoe predskazanie i raspoznavanie obrazov. Vychislitel'nye sistemy. Volume 67. – Novosibirsk: IM SOAN, 1976, pp. 161–164.

TAMBOVTSEV Yu.A. Nekotorye harakteristiki raspredelenia fonem mansiyskogo yazyka. – In: Sovetskoe Finno-Ugrovedenie, XIII, N 3, 1977, pp. 195–198.

TAMBOVTSEV Yu.A. Raspredelenie glasnyh fonem v mansiyskoy poezii. – In: Sovetskoe Finno-Ugrovedenie (SFU), XV, N3, 1979, pp. 164–167.

TAMBOVTSEV Yu.A. Zakonomernosti chastotnogo funktsionirovaniya dolgih i kratkih glasnyh v udarnyh i neudarnyh slogah mansiyskogo slova. – In: Sovetskoe Finno-Ugrovedenie, XVII, N 2, 1981, pp. 105–109.

TAMBOVTSEV Yu. Empiricheskoe raspredelenie chastotnosti fonem v kazimskom dialekte hantiyskogo yazyka. – In: Lingvostatistika i vychislitel'naya lingvistika. Trudy po lingvostatistike. Uchonye zapiski Tartuskogo gosuniversiteta, N 628, 1982, pp. 121–134.

TAMBOVTSEV Y.A. Selected consonantal characteristics of some Finno-Ugric languages from a phonostatistical point of view. – In: Finnisch-Ugrischen Forschungen XLV, H. 1-3, 1983, pp. 154–163.

TAMBOVTSEV Yu.A. Empirical distribution of the phonemes in Oroch (typological analysis). – In: Archiv orientální, volume 52, N 3, 1984, pp. 285–294.

TAMBOVTSEV Yu. A. Phoneme frequency and closeness quotient: establishing relationship degrees by phonostatistics. – In: Ural-Altaic Yearbook, N 56, 1984 – a, pp. 103–119.

TAMBOVTSEV Yu.A. The consonantal coefficient in selected languages. – In: The Canadian Journal of Linguistics, 1985, 30 (2), pp. 179–188.

TAMBOVTSEV Yu.A. The consonantal quotient as a means to indicate structural differences. – In: Journal of Pragmatics, 10, 1986, pp. 741–748.

TAMBOVTSEV Yu.A. Review on: Mansiyskiy yazyk. ROMBANDEEVA E.I. & VAHRUSHEVA M.P. Leningrad: Prosveshchenie, 1984. – In: Ural-Altaic Yearbook (Bloomington), N 59, 1987, pp. 168.

TAMBOVTSEV Yu.A. Typological linguistical distances. – In: Literary and Linguistic Computing, Vol. 3, N 3, 1988, p. 205.

TAMBOVTSEV Yu.A. The compactness of Uralic language group on the basis of certain phonological features. – In: Mitteilungen der Societas Uralo-Altaica. Heft 8. Viertes Internationales Symposion "Uralische Phonologie" 4-8 September 1989 in Hamburg, Hamburg: Universität Hamburg, 1989, pp. 186–187.

TAMBOVTSEV Yu.A. Kompaktnost' finno-ugorskoy yazykovoy semyi po dannym konsonantnogo koeffitsienta. – In: Linguistica Uralica, V. 26, N 1, 1990, pp. 13–20.

TAMBOVTSEV Yu.A. Selected phonostatistical features of the Hakas language. – In: Ural-Altaic Yearbook, N 63, 1991 (Bloomington), pp. 155–165.

TAMBOVTSEV Yu.A. Otsinka kompaktnosti grup mov na fonetychnomu rivni. – In: Inozemna filologiya, N 103, 1992, pp. 79–84.

TAMBOVTSEV Yu.A. (with Homitska I.Yu.) Statystychny parametry stiliv angliyskoyi movy. – In: Inozemna filologiya, N 106, 1993, pp. 54–59.

TAMBOVTSEV Yu.A. Dinamika funktsionirovaniya fonem v zvukovyh tsepochkah yazykov razlichnogo stroya. – Novosibirsk: NGU, 1994-a. – 136 pp.

TAMBOVTSEV Yu.A. Tipologiya uporyadochennosti zvukovyh tsepey v yazyke. – Novosibirsk: NGU, 1994-b. – 199 pp.

TAMBOVTSEV Yu.A. The distances of the Ugric languages based on the typological consonantal features. – In: Summaria acroasium in sectionibus et symposiis factarum. Pars II. – Jyväskylä: Moderatores, 1995, p. 116.

TUMASHEVA D.G. Yazyk sibirskih tatar. T. 2. – Kazan': KGU, 1968. – 272 pp.

TUMASHEVA D.G. Dialekty sibirskih tatar: opyt sravnitel'nogo issledovaniya. – Kazan': KGU, 1977.

SHCHERBAK A.M. Sravnitel'naya fonetika tyurkskih yazykov. – Leningrad: Nauka, 1970.

UBRYATOVA E.I. Yakutskiy yazyk i ego otnoshenie k drugim yazykam, a takzhe k yazykam mongol'skim i tunguso-man'chzhurskim. – Moskva: Nauka, 1960. – 16 pp.

UBRYATOVA E.I. (ed). Grammatika sovremennogo yakutskogo literaturnogo yazyka. – Moskva: Nauka, 1982. – 496 pp.

WELLS, S. In Memory of Master William Peter. – In: Times Literary Supplement, 1996, 4843: 28.

YUNUSALIEV M.B. Kirgizskiy yazyk. – In: Yazyki narodov SSSR. Tyurkskie yazyki (YANSTU), t. 2, pp. 482–505.

ZUBKOVA L. G. Sootnoshenie zvukovyh edinits so znachashchimi v tipologicheskom aspekte (Onomasiologicheskiy i semasiologicheskiy podhody v fonologii). – In: Voprosy yazykoznaniya, N 3, 1988, pp. 69–83.

Table 1

PERCENTAGE OF FREQUENCY OF OCCURRENCE OF FOUR CONSONANTAL GROUPS, DEFINED BY THE ACTIVE ORGAN (I.E. PLACE OF ARTICULATION) IN 13 TURKIC LANGUAGES, % OF ALL PHONEMES IN THE SPEECH CHAIN

Cons. group Language	Labial	Front	Palatal	Back	Total of consonants
1. Azeri	9,66	36.61	3.97	6.84	57.08
2. Turkish	10.41	33.50	2.81	10.91	57.63
3. Tatar	9.04	34.60	2.13	11.35	57.12
4. Kazah	7.41	34.95	2.57	12.09	57.02
5. Karakalpak	12.80	34.40	2.65	10.25	60.11
6. Kirgiz	8.43	36.27	3.07	9.86	57.63
7. Ujgur	9.83	32.35	0.74	13.47	56.39
8. Turkmen	8.50	34.50	4.10	10.90	58.00
9. Uzbek	9.42	34.34	2.84	11.48	58.08
10. Altai	5.98	40.24	2.54	11.71	60.47
11. Hakas	7.82	36.99	1.14	12.59	58.54
12. Jakut	6.10	32.77	5.22	13.27	57.36
13. Kumandin	9.22	32.47	14.32	11.92	67.93

Table 2

PERCENTAGE OF FREQUENCY OF OCCURRENCE OF FOUR CONSONANTAL GROUPS, DEFINED BY THE MANNER OF ARTICULATION (SONORANT, OCCLUSIVE, FRICATIVE) AND BY THE WORK OF THE VOCAL CORDS (VOICED) IN 13 TURKIC LANGUAGES, % OF ALL PHONEMES IN THE SPEECH CHAIN

Cons. group Language	Sonorant	Occlusive	Fricative	Voiced
1. Azeri	26.66	18.94	11.48	14.56
2. Turkish	24.89	23.22	9.52	11.60
3. Tatar	25.73	23.54	7.85	7.22
4. Kazah	22.97	23.25	10.80	11.91
5. Karakalpak	30.68	20.97	8.46	8.89
6. Kirgiz	24.31	25.76	7.56	14 .28
7. Ujgur	22.93	23.62	9.84	9.90
8. Turkmen	29.40	20.50	8.10	15.00
9. Uzbek	25.24	22.16	10.68	14.20
10. Altai	23.63	31.11	5.73	11.94
11. Hakas	23.95	20.93	13.66	10.26
12. Jakut	24.29	22.14	10.92	9.17
13. Kumandin	18.65	37.26	12.02	9.23

Table 3

FREQUENCY OF OCCURRENCE OF AZERI VOWELS AND CONSONANTS, % OF ALL PHONEMES IN THE SPEECH CHAIN. TOTAL – 91,706 PHONEMES.

N	Phoneme	%	N	Phoneme	%	N	Phoneme	%
1.	ý	10,77	12.	j	2.87	23.	v	1.45
2.	a	9.92	13.	s	2.75	24.	k ‘	1.10
3.	i	8.54	14.	u	2.61	25.	O	1.04
4.	n	7.46	15.	k	2.23	26.	x	0.93
5.	r	6.55	16.	y	2.20	27.	c	0.88
6.	d	5.81	17.	o	2.14	28.	c ‘	0.68
7.	l	5.70	18.	e	2.13	29.	q	0.63
8.	m	4.08	19.	s	2.01	30.	f	0.51
9.	i	3.57	20.	z	1.73	31.	p	0.31
10.	b	3.31	21.	g	1.60	32.	z	0.02
11.	t	3.02	22.	h	1.45		Total:	100%

Table 4

FREQUENCY OF OCCURRENCE OF TURKISH VOWELS AND CONSONANTS, % OF ALL PHONEMES IN THE SPEECH CHAIN. TOTAL – 107,192 PHONEMES.

N	Phoneme	%	N	Phoneme	%	N	Phoneme	%
1.	a	12.48	12.	d_	3.40	23.	p	1.10
2.	e	8.79	13.	u	3.29	24.	v	0.99
3.	k	7.29	14.	j	2.79	25.	o	0.87
4.	i	7.14	15.	g	2.46	26.	dz	0.71
5.	r	6.28	16.	b	2.43	27.	f	0.62
6.	n	5.47	17.	o	2.32	28.	a :	0.62
7.	m	5.27	18.	u	1.99	29.	o :	0.18
8.	l	5.08	19.	s	1.70	30.	i :	0.13
9.	i	4.50	20.	z	1.60	31.	u :	0.08
10.	t	4.38	21.	c	1.42	32.	k ‘	0.02
11.	s	3.45	22.	h	1.16	33.	l ‘	0.00
							Total:	100.00%

Table 5

FREQUENCY OF OCCURRENCE OF TATAR (SIBERIAN BARABA) VOWELS AND CONSONANTS, % OF ALL PHONEMES IN THE SPEECH CHAIN. TOTAL 67,569 PHONEMES.

N	Phoneme	%	N	Phoneme	%	N	Phoneme	%	N	Phoneme	%
1.	a	12.67	12.	k	2.76	23.	u	1.12	34.	oi	0.31
2.	n	8.01	13.	j	2.13	24.	ts	0.98	35.	e	0.25
3.	i	7.63	14.	m	2.11	25.	ai	0.94	36.	oi	0.25
4.	a	6.80	15.	o	2.02	26.	c	0.83	37.	j	0.14
5.	i	6.62	16.	q	1.91	27.	o	0.82	38.	y	0.14
6.	r	6.04	17.	d	1.73	28.	y	0.82	39.	h	0.10
7.	t	6.03	18.	s	1.60	29.	z	0.57	40.	yj	0.09
8.	l	5.58	19.	n	1.32	30.	o	0.55	41.	yj	0.06
9.	p	4.68	20.	b	1.32	31.	ai	0.50	42.	x	0.04
10.	k	3.93	21.	g	1.28	32.	v	0.41	43.	ei	0.03
11.	s	3.21	22.	e	1.26	33.	o	0.39	44.	f	0.01
									45.	z	0.01
									Total:		100.00%

Table 6

FREQUENCY OF OCCURRENCE OF UJGUR VOWELS AND CONSONANTS, % OF ALL PHONEMES IN THE SPEECH CHAIN. TOTAL – 48,331 PHONEMES.

N	Phoneme	%	N	Phoneme	%	N	Phoneme	%
1.	i	12.25	12.	k	3.42	23.	g	0.98
2.	a	10.90	13.	d	2.91	24.	uj	0.79
3.	u	6.83	14.	s	2.43	25.	v	0.75
4.	ý	6.48	15.	s	2.06	26.	j	0.74
5.	l	6.31	16.	p	2.04	27.	aj	0.67
6.	k	5.63	17.	c	1.81	28.	x	0.52
7.	m	5.33	18.	z	1.75	29.	x,	0.49
8.	t	4.94	19.	b	1.69	30.	oj	0.49
9.	n	4.79	20.	q	1.35	31.	z,	0.28
10.	r	4.67	21.	e	1.32	32.	ts	0.21
11.	o	3.88	22.	n,	1.09	33.	z	0.18
						34.	f	0.02
						Total:		100.00%

Table 7

FREQUENCY OF OCCURRENCE OF KIRGIZ VOWELS AND CONSONANTS, % OF ALL PHONEMES IN THE SPEECH CHAIN. TOTAL – 29,935 PHONEMES.

N	Phoneme	%	N	Phoneme	%	N	Phoneme	%	N	Phoneme	%
1.	a	12.67	12.	u	3.24	23.	c	1.55	34.	e:	0.08
2.	n	8.35	13.	b	3.21	24.	z	1.34	35.	x	0.02
3.	k	5.94	14.	j	3.07	25.	e	1.27	36.	v	0.02
4.	i	5.67	15.	m	3.06	26.	k,	0.59	37.	e	0.01
5.	e	5.62	16.	yy	3.03	27.	y:	0.35	Total: 100.00%		
6.	t	4.96	17.	g	2.76	28.	q	0.28			
7.	r	4.95	18.	O	2.44	29.	o:	0.28			
8.	l	4.61	19.	s	2.22	30.	n,	0,27			
9.	d	4.58	20.	p	2.14	31.	a:	0.26			
10.	i	3.87	21.	z	1.94	32.	O:	0.18			
11.	o	3.27	22.	S	1.77	33.	y:	0.13			

Table 8

FREQUENCY OF OCCURRENCE OF ALTAI CONSONANTS, % OF ALL PHONEMES IN THE SPEECH CHAIN. TOTAL – 190,525 PHONEMES. VOWELS – 79,105; CONSONANTS – 111,420.

N	Phoneme	%	N	Phoneme	%	N	Phoneme	%	N	Phoneme	%
1.	r	7.56	6.	d	3.52	11.	c	2.06	16.	s	1.25
2.	k	7.49	7.	b	2.74	12.	g	2.02	17.	z	1.21
3.	t	7.12	8.	j	2.54	13.	n,	1.99	18.	z	0.75
4.	n	5.07	9.	s	2.50	14.	m	1.69	19.	k,	0.20
5.	l	4.78	10.	dz	2.43	15.	p	1.53	20.	v	0,02
									21.	x	0.01

It is enough to provide the frequency of occurrence of consonants.
The data on the frequency of the phonemes in Yakut, Hakas, Kazah, Uzbek and Turkmen can be found in Tambovtsev, 1991.

Table 9

DISTANCES BETWEEN AZERI AND THE OTHER 12 TURKIC LANGUAGES, DERIVED BY THE AUTHOR ON THE BASIS OF THE EUCLIDEAN SPACE METHOD

1. Azeri – Turkmen (6.62); 2. Uzbek (7.39); 3. Turkish (7.89); 4. Kazah (8.78); 5. Kirgiz (8.89); 6. Hakas (8.89); 7. Karakalpak (9.47), 8. Jakut (10.75); 9. Baraba Tatar (10.83); 10. Ujgur (11.52); 11. Altai (15.80); 12. Kumandin (24.07).

Table 10

DISTANCES BETWEEN ALTAI AND THE OTHER 12 TURKIC LANGUAGES, DERIVED BY THE AUTHOR ON THE BASIS OF THE EUCLIDEAN SPACE METHOD

1. Altai – Kirgiz (7.96); 2. Kazah (10.86); 3. Baraba Tatar (11.40); 4. Turkish (12.00); 5. Uzbek (12.28); 6. Ujgur (12.69); 7. Jakut (13.45); 8. Hakas (13.64); 9. Turkmen (14.26); 10. Azeri (15.80); 11. Karakalpak (15.88); 12. Kumandin (18.56).

Table 11

DISTANCES BETWEEN KAZAH AND THE OTHER 12 TURKIC LANGUAGES, DERIVED BY THE AUTHOR ON THE BASIS OF THE EUCLIDEAN SPACE METHOD

1. Kazah – Uzbek (4.15); 2. Turkish (4.24); 3. Ujgur (4.79); 4. Hakas (4.89); 5. Jakut (5.04); 6. Kirgiz (5.67); 7. Baraba Tatar (6.47); 8. Turkmen (8.42); 9. Azeri (8.78); 10. Karakalpak (10.58); 11. Altai (10.86); 12. Kumandin (19.26).

Table 12

DISTANCES BETWEEN KARAKALPAK AND THE OTHER 12 TURKIC LANGUAGES, DERIVED BY THE AUTHOR ON THE BASIS OF THE EUCLIDEAN SPACE METHOD

1. Karakalpak – Baraba Tatar (7.07); 2. Turkish (7.35); 3. Turkmen (7.77); 4. Uzbek (8.81); 5. Azeri (9.47); 6. Ujgur (9.85); 7. Jakut (10.56); 8. Kazah (10.58); 9. Hakas (10.64); 10. Kirgiz (10.78); 11. Altai (15.88); 12. Kumandin (24.05).

Table 13

DISTANCES BETWEEN KIRGIZ AND THE OTHER 12 TURKIC LANGUAGES, DERIVED BY THE AUTHOR ON THE BASIS OF THE EUCLIDEAN SPACE METHOD

1. Kirgiz – Uzbek (4.15); 2. Turkish (5.53); 3. Kazah (5.67); 4. Turkmen (7.72); 5. Baraba Tatar (7.95); 6. Altai (7.96); 7. Ujgur (8.16); 8. Azeri (8.89); 9. Jakut (9.19); 10. Hakas (9.42); 11.Karakalpak (10.78); 12. Kumandin (18.86).

Table 14

DISTANCES BETWEEN BARABA TATAR AND THE OTHER 12 TURKIC LANGUAGES, DERIVED BY THE AUTHOR ON THE BASIS OF THE EUCLIDEAN SPACE METHOD

1. Baraba Tatar – Turkish (5.15); 2. Ujgur (5.58); 3. Kazah (6.47); 4. Jakut (6.52); 5. Karakalpak (7.07); 6. Uzbek (7.22); 7. Hakas (7.91), 8. Kirgiz (7.95); 9. Turkmen (9.36); 10. Azeri (10.83); 11. Altai (11.40); 12. Kumandin (20.33).

Table 15

DISTANCES BETWEEN TURKISH AND THE OTHER 12 TURKIC LANGUAGES, DERIVED BY THE AUTHOR ON THE BASIS OF THE EUCLIDEAN SPACE METHOD

1. Turkish – Uzbek (2.99); 2. Kazah (4.24); 3. Ujgur (4.41); 4. Baraba Tatar (5.15); 5. Kirgiz (5.53); 6. Jakut (6.31); 7. Turkmen (6.90); 8. Hakas (7.03); 9. Karakalpak (7.35); 10. Azeri (7.89); 11. Altai (12.00); 12. Kumandin (19.59).

Table 16

DISTANCES BETWEEN TURKMEN AND THE OTHER 12 TURKIC LANGUAGES, DERIVED BY THE AUTHOR ON THE BASIS OF THE EUCLIDEAN SPACE METHOD

1. Turkmen – Uzbek (5.56); 2. Azeri (6.62); 3. Turkish (6.90); 4. Kirgiz (7.72); 5. Karakalpak (7.77); 6. Kazah (8.42); 7. Jakut (9.29); 8. Baraba Tatar (9.36); 9. Hakas (10.07); 10. Ujgur (10.24); 11. Altai (14.26); 12. Kumandin (23.56).

Table 17

DISTANCES BETWEEN UJGUR AND THE OTHER 12 TURKIC LANGUAGES, DERIVED BY THE AUTHOR ON THE BASIS OF THE EUCLIDEAN SPACE METHOD

1. Ujgur – Turkish (4.41); 2. Kazah (4.79); 3. Baraba Tatar (5.58); 4. Uzbek (6.05); 5. Jakut (6.32); 6. Hakas (7.03); 7. Kirgiz (8.16); 8. Karakalpak (9.85); 9. Turkmen (10.24); 10. Azeri (11.52); 11. Altai (12.69); 12. Kumandin (19.92).

Table 18

DISTANCES BETWEEN UZBEK AND THE OTHER 12 TURKIC LANGUAGES, DERIVED BY THE AUTHOR ON THE BASIS OF THE EUCLIDEAN SPACE METHOD

1. Uzbek – Turkish (2.99); 2. Kazah (4.15); 3. Kirgiz (4.15); 4. Turkmen (5.56); 5. Ujgur (6.05); 6. Baraba Tatar (7.22); 7. Jakut (7.24); 8. Azeri (7.39); 9. Hakas (7.42); 10. Karakalpak (8.81); 11. Altai (12.28); 12. Kumandin (20.05).

Table 19

DISTANCES BETWEEN HAKAS AND THE OTHER 12 TURKIC LANGUAGES, DERIVED BY THE AUTHOR ON THE BASIS OF THE EUCLIDEAN SPACE METHOD

1. Hakas – Kazah (4.89); 2. Jakut (6.93); 3. Turkish (7.03); 4. Ujgur (7.03); 5. Uzbek (7.42); 6. Baraba Tatar (7.91); 7. Azeri (8.89); 8. Kirgiz (9.42); 9. Turkmen (10.07); 10. Karakalpak (10.64); 11. Altai (13.64); 12. Kumandin (22.25).

Table 20

DISTANCES BETWEEN JAKUT AND THE OTHER 12 TURKIC LANGUAGES, DERIVED BY THE AUTHOR ON THE BASIS OF THE EUCLIDEAN SPACE METHOD

1. Jakut – Kazah (5.04); 2. Turkish (6.31); 3. Ujgur (6.32); 4. Baraba Tatar (6.52); 5. Hakas (6.93); 6. Uzbek (7.24); 7. Kirgiz (9.19); 8. Turkmen (9.29); 9. Karakalpak (10.56); 10. Azeri (10.75); 11. Altai (13.45); 12. Kumandin (18.87).

Table 21

DISSTANCES BETWEEN KUMANDIN AND THE OTHER 12 TURKIC LANGUAGES, DERIVED BY THE AUTHOR ON THE BASIS OF THE EUCLIDEAN SPACE METHOD

1. Kumandin – Altai (18.56); 2. Kirgiz (18.86); 3. Jakut (18.87); 4. Kazah (19.26); 5. Turkish (19.59); 6. Ujgur (19.92); 7. Uzbek (20.05); 8. Baraba Tatar (20.33); 9. Hakas (22.25); 10. Turkmen (23.56); 11. Karakalpak (24.05); 12. Azeri (24.07).

Table 22

ALL THE DISTANCES BETWEEN 12 TURKIC LANGUAGES, DERIVED BY THE AUTHOR ON THE BASIS OF THE EUCLIDEAN SPACE METHOD

1. Uzbek – Turkish (2.99); 2. Uzbek – Kirgiz (4.15); 3. Uzbek – Kazah (4.15); 4. Turkish – Kazah (4.24); 5. Ujgur – Turkish (4.41); 6. Ujgur – Kazah (4.79); 7. Hakas – Kazah (4.89); 8. Jakut – Kazah (5.04); 9. Turkish – Baraba Tatar (5.15); 10. Turkish – Kirgiz (5.53); 11. Uzbek – Turkmen (5.56); 12. Ujgur – Baraba Tatar (5.58); 13. Kirgiz – Kazah (5.67); 14. Uzbek – Ujgur (6.05); 15. Jakut – Turkish (6.31); 16. Jakut – Ujgur (6.32); 17. Baraba Tatar – Kazah (6.47); 18. Jakut – Baraba Tatar (6.52); 19. Turkmen – Azeri (6.62); 20. Turkmen – Turkish (6.90); 21. Jakut – Hakas (6.93); 22. Hakas – Ujgur (7.03); 23. Hakas – Turkish (7.03); 24. Baraba Tatar – Karakalpak (7.07); 25. Uzbek – Baraba Tatar (7.22); 26. Jakut – Uzbek (7.24); 27. Turkish – Karakalpak (7.35); 28. Uzbek – Azeri (7.39); 29. Hakas – Uzbek (7.42); 30. Turkmen – Kirgiz (7.72); 31. Turkmen – Karakalpak (7.77); 32. Turkish – Azeri (7.89); 33. Hakas – Baraba Tatar (7.91); 34. Baraba Tatar – Kirgiz (7.95); 35. Kirgiz – Altai (7.96); 36. Ujgur – Kirgiz (8.16); 37. Turkmen – Kazah (8.42); 38. Azeri – Kazah (8.78); 39. (8.81); 40. Kirgiz – Azeri (8.89); 41. Hakas – Azeri (8.89); 42. Jakut – Kirgiz (9.19); 43. Jakut – Turkmen (9.29); 44. Turkmen – Baraba Tatar (9.36); 45. Hakas – Kirgiz (9.42); 46. Azeri – Karakalpak (9.47); 47. Ujgur – Karakalpak (9.85); 48. Hakas – Turkmen (10.07); 49. Ujgur – Turkmen (10.24); 50. Jakut – Karakalpak (10.56); 51. Kazah – Karakalpak (10.58); 52. Hakas – Karakalpak (10.64); 53. Jakut – Azeri (10.75); 54. Kirgiz – Karakalpak (10.78); 55. Baraba – Tatar – Azeri (10.83); 56. Altai – Kazah (10.86); 57. Baraba Tatar – Altai (11.40); 58. Ujgur – Azeri (11.52); 59. Turkish – Altai (12.00); 60. Uzbek – Altai (12.28); 61. Ujgur – Altai (12.69); 62. Jakut – Altai (13.45); 63. Hakas – Altai (13.64); 64. Turkmen – Altai (14.26); 65. Azeri – Al-

tai (15.80); 66. Altai – Karakalpak (15.88); 67. Kumandin – Altai (18.56); 68. Kumandin – Kirgiz (18.86); 69. Kumandin – Jakut (18.87); 70. Kumandin – Kazah (19.26); 71. Kumandin – Turkish (19.59); 72. Kumandin – Ujgur (19.92); 73. Kumandin – Uzbek (20.05); 74. Kumandin – Baraba Tatar (20.33); 75. Kumandin – Hakas (22.25); 76. Kumandin – Turkmen (23.56); 77. Kumandin – Karakalpak (24.05); 78. Kumandin – Azeri (24.07).

Table 23

TOTALS OF DISTANCES AS INDICATORS OF THE CENTRAL POSITION AMONG THE TURKIC LANGUAGES

1. Turkish – 89.39; 2. Kazah – 93.15; 3. Uzbek – 93.31; 4. Kirgiz – 104.28; 5. Baraba Tatar – 105.79; 6. Ujgur – 106.56; 7. Jakut – 110.47; 8. Hakas – 116.12; 9. Turkmen – 119.77; 10. Azeri – 130.90; 11. Karakalpak – 132.81; 12. Altai – 158.78; 13. Kumandin – 249.37.