Yuri Tambovtsev
Novosibirsk Pedagogical University
Novosibirsk, Russia

Language Taxons and the Naturalness of their Classification

Abstract

Language subgroups, groups, families and unities were investigated from the point of view of their dispersion as first proposed in Tambovtsev, 1986. We consider how compact this or that language taxon (i.e. subgroup, group, family or unity) is on the basis of distribution of certain consonantal groups in the speech sound chain. Therefore, one can speak about a compact or diffuse language family. If a language taxon is compact, then its internal connections are shorter than its outer connections. The same notion of compact object is accepted in pattern recognition. The more compact the family, the more correctly its languages are chosen. If we put in the family a language which does not belong to the family, then the dispersion of the family rises, thus it becomes less compact. If the language has a similar sound chain, then the dispersion of the group remains the same or becomes less. It means that the family became more compact. In this case we speak about the typological properties of the families. We measure the dispersion of a family by the sum of dispersions of eight phonostatistical features: frequency of occurrence of labial, front, palatal, velar, sonorant, occlusive, fricative and voiced consonants. It is important that the features do not intersect. The values of the coefficient of variance (V) and T coefficient show the degree of dispersion. The principle is the greater the dispersion, the less compact the family.

We have chosen the coefficient of variance and T-coefficient since they both keep to the law of commensurability. Comparing different languages of different language families and different morphological structures was possible since all of them have the same eight phonetic features mentioned above. We have considered Indo-European, Turkic, Mongolian, Tungus-Manchurian, Samoyedic, Finno-Ugric, Paleo-Asiatic, Austronesian, Australian and American Indian language families. Measuring the typological density of language taxons (subgroups, groups, families, unities) may help to understand how natural (or correct) they are.

Introduction

Every language has its own speech sound chain, by which it is different from the other languages in the world. It is possible to compare sound picture of a language to the sound picture of some other language because they yield different frequencies of occurrence of their speech sounds. If the speech sound chain of some language is similar to the speech sound chain of the other language, one can see that the frequencies of occurrence of speech sounds in these languages are similar. Languages with similar sound pictures form compact subgroups, groups, families and other language taxons. The aim of this article is to consider the degrees of compactness, or density, of the set of world languages classified into branches, subgroups, groups, families and other language taxons, from the point of view of the frequency of occurrence of their speech sounds. It is necessary to understand that if these language taxons are compact enough they may be considered to be natural. The analysis of the classifications in natural sciences shows that compactness is the sign of a natural classification with all its special properties (Rosova, 1986: 71 - 98).

Natural and articifial classifications

Specialists in classifications believe that a good classification should be natural. Usually pattern recognisers discuss the quality of classification from the point of view of its naturalness. By the term "natural" classification they define some ideal classification, which should be the result of the development of the classification methods. Natural classifications are believed to possess the following properties. First of all, natural classification should not need a large number of classificatory features. Secondly, the features should be essential. Thirdly, a natural classification should foresee the unknown properties of the object (Zagoruiko, 1999: 36 - 47; Zagoruiko, et al., 2004: 28).

Having analysed different types of classifications in geology, biology and other natural sciences, Stalia S. Rozova came to the conclusion that in fact the idea of a "natural classification" coincides with the idea of a "good" classification. At the same time she warns us that it is very hard to define a "natural" class. The trouble usually lies in the search for the fundamentals of classifications of this sort. The first difficulty arises in finding the essential features on which the other features must depend. However, in many cases the choice of such features is often not only difficult, but completely impossible. Classification activities often begin in situations where scholars do not know the object of their research completely. Therefore they do not know which of the selected features are really primary and which are secondary (Rozova,1986: 74).

Some language taxons seem to be natural. One can give the example of such a set of languages. Take, for instance, the East Slavonic language taxon, including Belorussian, Russian and Ukrainian. It is possible to prove it because they are typologically close and the direct communication of speakers is possible. The situation is not so obvious with some other language taxons. Let us take the example of the languages and dialects which constitute the Ugric group of languages: Hungarian, Mansi, and Hanty. In fact, Hungarians do not understand either Mansi, or Hanty. The speakers of the languages of the Ob-Ugrian branch of the Ugric subgroup of the Finno-Ugric group of the Uralic family, i.e., Mansi and Hanty, usually cannot understand each other either. Communication even in different dialects of Mansi (and Hanty!) often is not possible at all. The Konda and Sosjva dialects of the Mansi language are so different that communication between the speakers of these dialects is not possible. One should expect that the speakers of different dialects of a language understand each other. However, this is not the case with the Ugric languages. It is also true for many dialects of the Hanty language, not to speak of Hanty and Mansi, as they are said to be separate languages.

Taxonomy is always a sort of classification. So, we can say that classifications create taxons. Natural classifications create natural taxons and articicial classifications create artificial taxons. The example of an artificial taxon may be a set of languages whose listings begin with the letter "m" in the alphabetic catalogue the library. Let us just give some of the languages which begin with the letter "m", taken at random: "Mabida, Macedonian, Madu, Magahi, Malay, Mangarayi, Mansi, Marathi, Mari, Maykulan, Mbaatyana, Megeb, Moldavian, Mongolian, Mordovian, etc."

Everybody knows Mendeleev's classification of chemical elements, "The System of Chemical Elements". It is a fair example of a natural classification, which can foretell the properties of a chemecal element because of its position in this classification. However, it is possible to construct an artificial classification of the chemecal elements, if one puts the elements just in the alphabetic order. This artificial classification is not able to foretell the properties of some chemical element due to the position of the element in the alphabetic list, yet it is very useful for learning purposes.

Specialists in the theory of classifications usually think it quite essential to define first of all two types of classifications: natural and artificial (Rozova, 1986: 45). Summing up all the points of view on the construction of natural and artificial classifications, we can say that natural classifications are basic and fundamental, while artificial classifications are optional and subjective. However, one cannot help agreeing with S. S. Rozova, N. I. Kondakov, M. S. Strogovich, and other specialists in the field of theoretical classifications, who analysed many classifications in many sciences and humanities and came to conclusion that it is often hard to judge if the classification is natural or artificial, especially at the initial stages of some sciences or humanities (Kondakov, 1971: 151; Rozova, 1986: 46 - 49). They point out that usually scholars try to build a natural classification because they consider natural classifications most important and "good". More often than not a natural classification is a sort of ideal. Genetic classifications are said to be natural. S. S. Rozova shows that usually genetic classifications, which were built at the early stages of development of some sciences and thought to be natural at the early stage, turn out, in fact, to be artificial at the later stages of the development and should be reconsidered and changed (Rozova, 1986: 84 - 98). This, perhaps, is the case with Uralistics now. She warns against the considerations of some hypothesis as facts (Rozova, 1986: 87 - 92).

Some time ago, the Finno-Ugric and Samoyedic languages were considered to be separate language families. However, now is is fashionable to unite them into one genetic family (Austerlitz, 1990: 569). Although some linguists believe the united set of Finno-Ugric and Samoyedic languages called "Uralic family" is a natural taxon of languages, some other linguists (e.g. Ago Kuennap, Angela Marcantonio, Kalevi Wiik, etc.) do not believe them to be a family, that is, a genetically related language taxon, which can be called natural. It it necessary to remark, of course, that this depends on how a language family is defined. Usually, one understands a family as a genetically related language entity, that is, a close set of genetically related languages. It is supposed to form a natural taxon. Many linguists believe Turkic languages to form a natural taxon, since they are very similar and the direct communication is usually possible. Some specialists in Finno-Ugric and Samoyedic studies are quite sceptical that all Uralic languages, especially Finno-Ugric and Samoyedic, are genetically related. That is they do not believe Uralic taxon of languages to be a natural language taxon. The demonstration of a genetic relationship depends on finding words of similar phonological shape having equivalent meaning. That means that if languages are related, their speeech sound chains are similar.

Language Taxonomy

Usually the languages of the world are classified into taxons on the basis of some words which have similar or identical sound forms, at the same time having similar or identical meanings. We are trying to study some of the defined language taxons by a new method called typologo-metrical. Here, we shall explain the method in detail by the example of the taxon of Uralic languages. The taxon of Uralic languages is known to include Finno-Ugric and Samoyedic languages. We should analyse the typological similarity of the sound chains of the Finno-Ugric and Samoyedic taxons to determine if they are similar enough to belong to one and the same language family. If they are not similar, than we should come to the conclusion that their combination into one language family is artificial.

Let us consider one point, which may be the same for natural and artificial classifications, the usefulness of these classifications. Sometimes this point, especially at the early stages of the development of some science or in the humanities, leads scholars astray. A useful artificial classification may be taken for a natural classification. Thus, strange as it may sound, both natural and artificial classifications may be quite useful. A list of Finno-Ugric, Samoyedic, Turkic or the languages of the world in alphabetic order is a fair example of a useful classification, which is at the same time artificial. The order of the languages in these classifications, and thus, the neighbouring languages on the list, have nothing to do with the origin or typology of these languages. Moreover, this order may be different in English and in Russian because the order of the letters is different. Nevertheless, this artificial classification of languages is quite useful, especially for different sorts of catalogues or lists. In fact, in describing Turkic languages, we took the principle of the alphabetic order since there are at least 15 classifications of Turkic languages, which may be called natural, since they take into account some important and essential typologo-genetic features. At the same time, the artificial language classifications select some arbitrary features, which are not important or essential for this or that set of languages (Tambovtsev, 2001-b). In this case, an artificial classification is more correct, because a natural classification may be misleading. Nina Z. Gadzhieva does not believe it is possible to yield one classification of the Turkic languages, which should be true from all aspects. On the contrary, she emphasises that different features may give different classifications. She strongly believes that the use of computers and the methods of mathematical linguistics may help to correct the existing classifications of Turkic languages (Gadzhieva, 1980: 125).

We shall study different language taxons on this basis of the new method in linguistsics, i.e., the degrees of compactness of the main phonetic features. Let us discuss the notion of compactness and how to measure the degree of compactness of different language taxons. In this form the notion of typological compactness from the phonological point of view was introduced in linguistics in 1986. It was based on the frequency of occurrence of some certain important and essential articulatory features. Several criteria of mathematical statistics were used to measure compactness (Tambovtsev, 1986).

Establishing the Strict Hierachy of Language Taxons

However, before discussing the degrees of compactness of different language taxons, one must establish the exact order of the language taxons. The ordered series of taxons must begin with the smallest taxon and end with the largest. By the smallest taxon we mean the language taxon which includes the least number of languages. It is quite logical to begin with the notion of a branch as the smallest language taxon. Thus, we can propose to define the following ordered series of language taxons, from the smallest to the largest:

1. branch
2. subgroup
3. group
4. family
5. unity
6. phylum
7. union
8. community

Language taxonomy is known to be tightly linked with language typology and language classification. Typology is considered to be the method of research which is based on the separation of a set of some objects into certain types. The type is meant to be a taxonomic unit. As a result, one can acheive a sort of taxonomy, which in linguistics can be understood as a classification. Nickolai G. Zagorujko points out that the structure of a taxon is better if more similar objects are united into one taxon. The diversion of the individual characteristics of the objects from the mean is minimal. The requirements for "similarity" or "closeness" are based on the notion of compactness, which is put forth by different scholars who deals with taxonomy (Zagorujko, 1972: 90).

One has to define a set of languages as a branch, i.e., the smallest language taxon. One of the options is to define Ob-Ugric languages (Mansi and Hanty) as a branch of the Ugric subgroup of the Finno-Ugric group of the Uralic family. In its own turn the Uralic family may enter the Ural-Altaic language unity. It is quite logical, but may or may not be a natural classification of the languages in question. Unfortunately, in linguistics the notion of a branch, subgroup, group, etc. is not paid attention to, so they are mixed. A branch is often wrongly called a subgroup or a group. Even a language family is sometimes called a group, though sometimes it is called a language unity. Thus, one can see that the definitions of language taxons are not stable. In fact, there is no one-to-one correspondence between the terms and the natural subdivisions or divisions, which are generally accepted and fixed.

Therefore, it is better to use for language sets some general term like "a taxon". We propose by a language taxon to mean some sort of a set of languages. Actually, by our typologo - metrical method we try to construct some sort of typologo-metric classification for Finno-Ugric and Samoyedic languages, known as Uralic languages. However, it is still a great enigma if they are a closely related family from a typological point of view. They may be a conglomeration of languages, mechanically put together, just for some sort of convinience to classify them. Thus, in this case, one should call them an artificial classification. If they are sufficiently close from the phono-typological point of view, they should be called a natural classification. A natural classification is apt to be a genetic one with greater probability. After calculating Uralic compactness on the one hand, and Finno-Ugric and Samoyedic compactness on the other hand, one can draw certain conclusions. So, we can receive the values of compactness for these taxons: a) Ugric; b) Ugric-Permian; c) Ugric-Volgaic; d) Finno-Ugric; e) Uralic; and many others.

After that it is advisable to compare these values of compactness with those of the Turkic, Tungus-Manchurian and other taxons of the world languages. Thus, we are trying to build up a new systematics of the Finno-Ugric, Samoyedic and other languages, defined in accordance with their presumed or natural relationships and based on some certain set of the selected features.

Crisis of Some Language Classifications

It appears that nowadays a crisis of a scientific paradigm exists in the field of Uralistics. One can notice the main features of this crisis, which were or are the same as in the other sciences or humanities. These features are well described by T. S. Kuhn in his book "The Structure of Scientific Revolutions" as the crisis of the old scientific paradigm and the creation of the new scientific theories (Kun, 1977: 96 - 109). Kuhn is quite correct to stress that the old scientific paradigm never goes away peacefully. Usually, scholars strongly and negatively react to new theories and to those scholars who introduce new theories. Kuhn points out that what the scholars never do, it is to rush to the support of the new theory (Kun, 1977: 110 - 119). It is necessary to point out that there were quite negative reactions to the ideas of Nokolai S. Trubetzkoy to his doubts on the genetic closeness of the Indo-European languages.

We can see the similar negative reaction of the majority of the specialists in Uralistics to the new theories of Ago Kuennap, Angela Marcantonio, Wiik Kalevi and others, who try to reject the old scientific paradigm in Uralistics.

We made up our mind to introduce some new data on the typology of sound chains in the Uralic languages. Our data may help either to make the old Uralic paradigm stronger or may give new evidence for rejecting it. It is easy to explain psychologically why the old scientific paradigms are more stable and why many scholars would rather cling to a false (but old) paradigm, than switch over to the true (but new and unknown) one. It is quite cosy to remain in the embrace of the old and known paradigm. One can always close his or her eyes to its inconsistences and drawbacks. Many Uralic linguists got used to the old classification, which they first studied as students. They do not want to think about it twice since they usually work on some other linguistic problems, which do not concern the classification of languages. Usually, many linguists do not want to disturb "sleeping dogs". They do not believe that this or that linguistic classification must be checked again and again. Fortunately, in Uralistics there are some other linguists who think that with growing linguistic knowledge the old linguistic classifications should be verified. That is, every new linguistic fact should be used to verify the old linguistic classifications. If more and more new linguistic facts contradict the old classification, then it would have to be reconsidered on the basis of the new leverl of linguistic knowledge. The linguists with modern linguistic thinking argue that the old linguistics classifications must be verified and checked again and again and reconsidered if necessary again and again. However, in Uralistics, as well as in linguistics in general, old classifications are not reconsidered after an abundance of new linguistic facts has been received. One must bear in mind a simple idea: what was good and logical several centuries ago, i.e., at the old level development of linguistics, may be neither good nor logical at the advanced development of linguistics, of course, if we want to call this entity "science". Any linguist must understand the difference between a linguistic fact, which may remain true, though discovered several centuries ago, and a linguistic theory, which can be altered or rejected when the abundance of new linguistic facts are discovered.

Some outstanding linguists, such as Boris A. Serebrennikov, urged linguists to return to the established language taxons (classifications) again in order to verify them on the basis of certain laws of logics. He stressed that each established genetic language family, i.e., a language taxon or a classification, is not a fact but a hypothesis (Serebrennokov, 1982: 6).

Compactness of Language Taxons

We built our definitions and ideas about compactness on the fundamentals of pattern recognision in order to be able to solve some of the problems in Uralistics. Actually, the problems in other fields of linguistics are often similar and cannot be solved in any other linguistical way, i.e. remaining inside the frames of reasoning and data of classical linguistics.

It is important to bear in mind that in this form the notion of compactness is usually used in the Sciences, not in the Humanities, although we have omitted mathematical formalism. We understand "compact" as "neatly fitted, firmly put together, closely united or packed, not gangling or spare; concentrated in a limited area or small space, compressed, condensed, having density". One should note that if we remove the unnecessary mathematical formalism of pattern recognition, then this notion is very similar to the notion of compactness in philosophy, science, technology and everyday life (EK, 1975:457; Hornby, 1984: 115; Kondakov, 1975: 254; Ozhegov, 1970: 280; Petrova, 1964: 314; Webster, 1965: 168). In linguistics it was not used in the way we use it. It looks as if we introduced the concept into typology for the first time in our own works in the seventies of the previous century. One should not mix the term "compact" in pattern recognition and in acoustics; which was later used in experimental phonetics. It is true that there the term "compact" was used in the acoustical sense, as one of the features of a vowel or consonant sound. The term was set up by Jakobson and Halle in their distinctive feature theory of phonology. Compact sounds are defined acoustically as those which have a relatively high concentration of acoustic energy in a narrow, central part of the sound spectrum. It is a common notion in acuostic phonetics and is generally accepted (Crystal, 1980: 72).

We understand "compactness" as it is understood in pattern recognition with some references to languages, if a language is understood as an object. That is, we understand it as the degree of how compact the languages are placed in the language taxons: branches, subgroups, groups, families, unities and other language superfamilies from the point of view of the frequency of occurrence of certain groups of consonants in certain languages. In other words, it is the value of the total distance between the languages inside a language taxon, i.e. branch, subgroup, group, family or any other language superfamily. In fact, we believe that every language subgroup, group, family, superfamilly, unity or Sprachbund should be considered as a taxon. Taxonomic approach was taken from biosystematics and pattern reconition. This approach requires us to select certain linguistic features. We have selected the frequency of occurrence of certain phonological classes (vowels and consonants) and groups (labial, front, palatal, velar, sonorant, occlusive, fricative and voiced consonants). The compactness of a language subgroup, group or family is calculated on the basis of the frequency of occurrence of the mentioned features.

We found only two cases when the notion of compactness is used in linguistics in our way. We believe that in these cases, the notion of compactness, which we described earlier, does not differ from the meaning of this term used in their linguistic works. First of all, we mean the works of V. V. Martynov, T.V. Gamkrelidze and Vjach. Vs. Ivanov. Unfortunately they don't define the term "compactness". We can presume that this term is used in its common logical sense that is in the sense reflected in the dictionaries. Thus, V. V. Martynov speaks about "a compact language massive", meaning that the native speakers of a certain language live together (Martynov, 1983: 6). Tomaz V. Gamkrelidze and Vjacheslav Vs. Ivanov speak about "the compactness of territorial language spread" (Gamkrelidze et al., 1984: 44).

Sometimes this term is used in the sense close to ours in archeology. So, Tret'jakov speaks of "compact heaps of ancient relics" (Tret'jakov, 1970: 81), "compactness of the territories of the peoples" (Tret'jakov, 1970: 3) or "compact settling" (Tret'jakov, 1982: 118).

It is necessary to emphasise that in our works we used a more precise definition of compactness (Tambovtsev, 1986; 1991). It was possible to receive some new linguistic results because we based our investigation on the clearly cut and exact definition of compactness. It is accepted and used in pattern recognition and statistical methods of classification (Arkadjev et al., 1964: 29-34; 1971; Bongard, 1967: 30-31; Dide, 1985: 65-71; Eliseeva et al., 1977: 9-14; Mirkin, 1976:114-116; 1980: 233-236; Vasil'jev, 1969: 16-18; Zagorujko, 1972: 21).

Some scholars speak about "condensation" which is in fact the same as compactness (Mirkin, 1976: 114-146; 1980: 232-238). Yuri D. Apresjan uses the term "condensation" in linguistics. Basing on condensations his algorithm constructs classes of verbs in Russian (Apresjan, 1966: 141-144). We consider that in this case both terms: "compactness" and "condensation" mean the same.

Let us consider in detail what is compactness or condensation. B. G. Mirkin considers condensation to be such a set of objects in which any inner link is shorter than any outer link (Mirkin, 1980: 235). Further we shall se that the definition of compactness is more or less the same in the opinion of different scholars.

The first scholars who defined compactness were A. G. Arkadjev and E. M. Braverman, who dealt with methods of pattern recognition. They considered compactness as the set which satisfies the following conditions:

A. the number of marginal points is much less than the total number of points

B. any two inner points may be linked by a rather smooth line coming only through the points of the same set; and as a consequence

C. almost any inner point has only the points of the same set in a rather vast neibourhood (Arkadjev et al., 1964: 28)

Arkadjev and Braverman developed the idea of compactness in their subsequent works. Actually, they remark, if the group (set) is compact, then the objects inside it should be linked tightly, and on the contrary, the objects of different groups should be far away (Arkadjev et al., 1971: 27).

Another pattern recogniser V. I. Vasil'ev believes that if the set of points is compact, then:

A. a smooth transition from one image to another is always possible inside one and the same pattern, so that all intermediate images are recongnised as the images of one and the same pattern; on the contrary it is not possible to transit from the elements of one image to the elements of the other image without obtaining some intermediate elements which have uncertain origin.

B. if the deformation of the image in any direction is not great, then they remain inside the same image (Vasil'ev, 1969: 16-17). However, V. I. Vasiljev further cites all the requirements of Arkadjev and Braverman, which we discussed above. Therefore, they are not provided here.

M. M. Bongard remarks that if the set of points is compact, then they all are situated in a space closely (Bongard, 1967: 30-31).

N. G. Zagorujko believes that one often operates with the notion of compactness in pattern recognition, by which one means that the realisation of one and the same image is reflected in geometrically close points of the sample space (Zagorujko, 1972: 21). Eliseeva I. I. and V. O. Rukavishnikov speak about compactness and condensation as about the places where the points consentrate (Eliseeva et al., 1977: 40). Thus, it is easy to define a taxon as compact. One should find the inner and outer links of the languages inside it and check if the mean of the inner links is less than the mean of the outer links.

The majority of pattern recognisers agree with the definition of compactness given by V. M. Glushkov and his co-authors in Encyclopaedia of Cybernetics, who regard the hypothesis of compactness as a supposition that the subset of the images to be recognised is simple in a certain sense. The notion of simplicity may be realised differently. For instance, the classes of images are called compact if they may be separated from each other with the help of hyper-planes or when each class of images can be represented in the form of the unity of several convex sets.

The criterion of compactness reflects the idea that the images of one class should be more similar, than those of different classes (EK, 1975: 229). If a set of objects (in our case - languages) is compact, then it forms a taxon, i.e. a subgroup, group, family or superfamily.

This is why the notion of taxon is closely connected with the notion of compactness. In fact, the value of compactness may be regarded as the total of the inner distances inside a taxon.

Every language may be regarded as a separate object. We must analise this object and define some features on the basis of which we shall form taxons. The feaures must be basic. Simon G. Kordonskij states that there are two forms of theoretical descriptions. Classificational description introduces the notion of a taxon while the experimental description introduces the notion of a type. Taxons fix the objects, which exist. Types fix functioning of objects, i.e. the way the objects exist (Kordonskij, 2001: 19). Languages may be regarded as objects and as sets of functions. This is why languages may be both taxons and types.

Yuri A. Shreider understands languages as inner systems, which could be classified. Therefore he understands the schemes of language classifications as the outer systems. In fact, by the outer system he considers some class of objects, which have some common features. Moreover, he thinks that these object may be united because and only because of the fact that they form a natural system. Developing his ideas, S. G. Kordonskij adds that common features may be a part of both inner and outer systems. The outer system may fall into classes in the natural way (Kordonskij, 1983: 141). It is quite important for linguists to understand if existing language taxons are real or imaginary, and can be changed by the next generation of linguists. We believe that our investigation may help linguists to regard existing language classifications as natural systems. In our case the great value of compactness indicates that the classification is natural, thus the taxons are natural too. On the contrary, if the value of compactness is small, then the classification is not natural. Actually, it may show that one should not unite some languages in one taxon (group).

Uniting some languages into subgroups, groups, families and superfamilies, we separate them from the medium of the rest world languages. As G. P. Mel'nikov correctly remarks, going over from the medium to the object, a subject discovers an object as a violation of the qualitative property, i.e. as the violation of the homogenous property of the quality (Mel'nikov, 1978: 22). In this case compactness may be an indicator of holism. In a sense, the value of compactness shows the limits of allowed differences between the languages in a taxon, which can be measured by the theoretical values of the "chi-square" criterion. If the introduction of a language violates the compactness of a taxon too much, then the obtained value of chi-square criterion is greater than its theoretical value. One can see that this certain language does not belong to this certain taxon (Tambovtsev, 1994-a: 23-69). G. P. Mel'nikov calls it the borders of the allowed difusion of the functional states. Transferring his remark on a language taxon as a holistic object, one can speak about the limits of the chages under which the given object remains in the limits of the stable functioning (Mel'nikov, 1978: 55).

In other words, the stability of a taxon as a holistic object aims at the influence on theose languages whose occurrences differ too much from the mean of the taxon. If the language does not want to accomodate its typology, then it is ousted from the taxon. Sometimes only one language violates the compactness of the taxon. However, there may be cases when many languages greatly fluctuate from the mean. If these differences are greater than the theoretical value of the chi-square criterion, then one should not consider this taxon holistic. In this case it is not a group, but a mechanical mixture, i.e. random conglomeration of languages. It is not what can be called a natural class (taxon).

We believe that the value of the compactness of a language taxon may be the indicator of the measure of how much systemic this or that group is, if this group is a holistic object. In the opinion of G. P. Mel'nikov, any object may be regarded as a system, if it is defined how systemic it is (Mel'nikov, 1978: 68). Gennadij P. Mel'nikov underlined the importance of treating the languages of the world as some sort of a system, which should be studied by the methods of mathematical statistics (Mel'nikov, 2003: 278 - 281; 347, etc.).

As it was mentioned earlier, it is very important to select the most essential features in order to construct a taxon. The features chosen are the most basic for any language. They are the frequency of occurrence of the articulatory consonantal groups defined by:

A. the work of the active organ of speech: labial, front, palatal and velar sometimes this is called the place of articulation;

B. manner of articulation: sonorant, occlusive and fricative

C. work of the vocal cords: voiced

This is the most complete consonantal classification from the point of view of articulation. Thus we can call these features the most basic and essential for any human language. Therefore, we obtained eight feature for consonants:

1) labial
2) front
3) palatal
4) velar
5) sonorant
6) occlusive
7) fricative
8) voiced

The above have equal typological weights.

The frequency of occurrence of the consonantal groups mentioned above is calculated in per cent to all the phonemes in the speech sound chain. This is why the frequency of vowels is tacitly present, though in this work it is not vividly (exactly, explicitly, obviously) used. If it is not possible to find out the differences in the compactness of language taxons, based only on consonants, then the actual frequencies of vowels should be introduced.

The ethalon for comparing typological similarities of speech sound chains of different languages is the value of compactness of different language taxons. It is quite clear now. However, it is not clear how to calculate the value of compactness. We must develop a certain method for linguistics since nobody has done it before. We have developed sevaral methods. Now let us consider them. The easiest method to calculate compactness is the calculation of the standard deviation. However, the easiest way is not always the best. We can't use it for reasons of commensurability. Nevertheless, one can use standard deviation if one studies the variability of a certan group of consonants in one and the same language. For instance, one can calculate the variability of labial consonants in the texts of different writers in English. If one wants to compare the data, which have various values of their means, then one should use the coefficient of variance, which keeps to the laws of commensurability. Actually, coefficient of variance allows us to compare data of different sorts and origins.

We must emphasise that it is quite necessary to keep to the principle of commensurability. Usually this is not discussed in linguistics. Nevertheless, it is one of the most basic principles in any scientific investigation (Drujanov, 1982: 28 - 52; Hurgin, 1979: 101 - 115; Petrov, 1974: 6; Zagorujko, 1981: 32 - 34).

In comparing objects (in this case - languages) one should keep to one and the same principle and the comparison must use universal parameters (Kondakov, 1971: 151). We believe compactness to be a universal parameter since we can measure it on the taxons of languages of different genetic origin and morphologic structure.

To measure compactness one should first measure its invese, i.e. difusion or deviation. Actually, a lot of methods were invented to measure deviation or variance. One of them is the coefficient of variance and the other is the value of the T-coefficient. One should bear in mind that the values of both coefficients are inverse to the value of compactness. In other words, the greater their value, the less the value of compactness, and on the contrary, the less their value, the greater the value of compactness. Before explaining how to calculate the coefficient of variance and T-coefficient, let us consider the calculation of the standard deviation, which is the basis for calculation of the coefficient of variance.

Very often lisguists compare means of occurrence of some linguistic units and don't go further than that. Actually, the mean value is more reliable than the mode or median (Hurgin, 1979: 105). One should take the mean because it represents the distribution better than the mode or median (Pavlovskij, 1967: 55). Nevertheless, to see the variance a linguist should use the standard deviation because it is the measure of despersion and changability. Christopher Butler believes that the standard deviation is difficult to interpret in common-sense terms, but recommends it, for it has properties which make it very suitable for further statistical work (Butler, 1985: 37). Standard deviation measures how the values of some variable are grouped around their mean. The value of the standard deviation shows the degree of this variability. It is measured by the sum of the values of the differences raised to the second power. The greater the value of the standard deviation of some linguistic unit, the more dispersed the values around their mean. The greater dispersion, the greater the changability of some linguistic phenomenon. And on the contrary, if dispersion becomes less, then the changability is less (Tambovtsev, 1994: 33 - 37).

Let's discuss the properties of standard deviation in detail. Compactness is really closely connected to variability. Christopher Butler is quite correct to state that first of all we must answer the question: how closely do the data cluster round the mean. We must also answer the other question before considering measures of variability in some detail: why we bother to measure this property at al? We agree with him that variability may be of interest in its own right (Butler, 1985: 35). In this case, it is important because we attempt to find out a compactness profile of a language taxon. On the one hand, we can claim a difference in the typology of speech sound chains of some languages much more confidently if the values are widely spread. On the other hand, we can see typological similarity between languages if the values cluster closely round the mean of a language taxon. One must bear in mind, that neither the range, nor the mean of some linguistic values is evident enough while comparing two language taxons. It is quite necessary to calculate the value of the standard deviation to realise the true distribution inside a language taxon. Let us show it graphically on a simple one-dementional case, though our conclusions are true for multi-dimentional cases. Theoretically one can imagine several cases (more details in Tambovtsev, 2003: 77 - 124).

It is easy to see that if two taxons have an equal number of members, then their means are equal to each other if the range is the same. However, the values of the standard deviation are quite different, and, thus, the values of compactness are quite different too:

A) [__[__[__[__[__[__[__[. B) [_[_[_[__________[_[_[_[

It is the most complex case. Nevertheless, even in this case, one can measure the truee difference between the values of compactness in these taxons.

Practically, in linguistics, one encounters the situation when the number of the membrs in the taxons and their ranges are different. We should consider these cases simpler. However, even then, it is better to calculate the value of the standard deviation in order to see how much their compactnesses differ.

It is quite clear that if one taxon has the same number of memebers as another one, then its compactness is less if its range is greater:

A) [_[_[_[_[_[_[_[. B) [___[___[___[___[___[___[

On the contrary, if the range in both taxons is the same, then the taxon with the greater number of members has the greater value of compactness:

A) [_[_[_[_[_[_[_[. B) [____[____[____[

We considered all the possible situations in which the value of standard deviation works better than the mean or the range.

One can easily find the formula of standard deviation in any book on mathemetical statistics:

$S= \sqrt{} \sum (x - M)^2 / n - 1$
where S - standard deviation
x - the value of some parameter of a language
M - the mean value
n - the number of languages

We recommend the reader to take (n-1) because then we obtain an unbiased estimate: it is therefore safest to use it (Butler, 1985: 37; Tambovtsev, 2003: 11 - 16).

Let's consider the properties of the coefficient of variation. As it was mentioned above the coefficient of variation in such situations, where it is necessary to compare the variability of distributions described in different units. In our case, it is not possible to compare the compactness of labial and front consonants with the help of standard deviation since their values differ too much. For instance, the value of the standard deviation of the labial consonants of the Ugric languages (Tab.3) is much less (nearly twice) than the standard diviation of the front consonants (cf. 1.49 and 1.90). However, the coefficient of variation of the front consonants, on the contrary, is twice as less (cf. 12.65% and 6.09%). Actually, the value of the standard deviation of labial consonants of the Germanic group of Indo-European family is 1.14, while the value of standard deviation of the front consonants is 5.38. We shall see further that in fact that the values of standard diviations of labial and front consonants are incommensurable and that is they cannot be compared directly. There is no basis for comparison since the mean of the labial consonants in Ugric languages is 11.79, while the mean of the Ugric front consonants is 3 times greater (31.18). The same imcommensurability one can observe in other language laxons. For instance, labial mean in Germanic languages comprises 11.42 while the front mean is 37.78. We must use some sort of measure, which takes into account the great absolute differences. Usually, to compare such great differences the values of the coefficient of variation is applied since it is a reliable relative measure (Richmond, 1964: 89 - 90; Tambovtsev, 2004: 11 - 16):

$$V = S / M \bullet 100\%$$

where V- the coefficient of variation
S - the standard deviation
M - the mean value

In fact, Chris Butler, Raimond G. Piotrovski, Yuri Tambovtsev and others have shown that the coefficient of variance allows us to compare any data with any data in linguistics by the values of the coefficient of variance without bothering about incommensurability of data (Butler, 1985: 37 - 43; Tambovtsev, 2004: 11 - 16).

We can reconsider the compactness of the labial consonants. The coefficient of variance of labial (V=10.01%) and front (14.25%) consonants of the Germanic group shows that the compactness is more or less the same (cf. 10.01% and 14.25%), though their standard diviations are quite different (cf. 1.14 and 5.38). One can see that coefficient of variation yields a more reliable result. One must not forget that the value of compactness is converse to that of the coefficient of variation. Thus, the taxon of Germanic languages is more compact from the point of view of the labial (V=10.01%), than front (14.25%) consonants. It is also possible to compare the compactness of different language taxons from the point of view of labial consonants. For instance, the Iranian group of Indo-European (V=15.06%) is less compact by the parameter of the labial consonants while Indic group is more compact (V=6.85%). Coefficient of variation allows us to compare the labial campactness of the groups of the Indo-European family mentioned above with those of families: Mongolian (V=7.55%), Tungus-Manchurian (16.19%) or American Indian languages (29.89%).

The value of the coefficient of variation as well as T-coefficient may measure the linguistic stability of a sample text. We can understand it in the way the notion of stability is used in cybernetics, i.e. stability is the ability of a system to return to some stationary state from any other different state (Glushkov, 1975: 468). Or in other words, stability is a property allowing some distortions to occur within certain limits, which are small enough (Glushkov, 1975: 478).

Naturally, there arises a question how great these allowed limits should be. It is easier to find out such limits for for the T - coefficient, as we shall see further. We do it with the help of the theoretical values for the "chi-square" criterion, given in statistical tables (e.g. Bol'shev et al., 1983).

However, it is more difficult to understand how great the allowed limits should be for the coefficient of variation since there are no theoretical limits for it. In linguistics as well as in the Humanities and in all the Sciences connected with man and his activity, these allowed limits are derived emperically, i.e. from practice. Unfortunately, different scholars allow different limits. Some of them consider it tolarable to have its value up to 50% (Martynenko, 1988: 62). V. N. Sis'kov is sure that the taxon is homogenious and stable if the value of coefficient of variation is less than 33% (Sis'kov, 1971: 10). G. Ja. Martynenko considers the set of texts of the writers of the end of the 19th century and the beginning of the 20th century stable and integral by some of the syntax features because their coefficient of variation is much less than 33% (Martynenko, 1988: 150 - 154). R. G. Piotrovkij calculates the stability of the use of the English definite article (V=7.12%) and the German word "power" in the texts of FRG (V= 90.00%) and GDR (160.00%). He thinks that the distribution of the article may be called stable (Piotrovskij et al., 1977: 243).

G. Ja. Martynenko correctly points out that the greater the value of the coefficient of variation, the greater the probability that the set is not homogenious. He calls it the criterion of fluctuation (Martynenko, 1988: 62). V. I. Sis'kov believes that the high value of this coefficient may indicate to the mixture of some two distributions with quite different means (Sis'kov, 1975: 101). A. I.Venchikov and his colleagues think that the value of this coefficient should not be greater than 50% (Venchikov et al., 1974: 21). G. N. Zajtsev considers critical the value greater than 105%. Actually, he proposes the following scale: 1) small variation: 0% - 4%; 2) normal variation: 5% - 44%; 3) considerable variation: 45% - 64%; 4) great variation: 65% - 84%; 5) very big variation: 85% - 104%; 6) abnormal variation: 105% and more. In the limits of normal variation he defines the so-called lower norm: 5% - 24% (Zajtsev, 1990: 39). So, his norm (44%) is greater than the critical value (33%) for other scholars.

In order to set up a crucial limit for our phonemic variations, let us consider how the coefficient of variation behaves in speech acoustics.

Let us consider fluctuation of the acoustic duration of speech sounds. The value of coefficient of variation of the voiced part of English consonants is 26.56% - 27.49%; tempo of speech - 7.69% - 16.04%; duration of Russian vowels 14.97% - 27.83% (Bondarko, 1981;

Bondarko et al, 1983). The value of the coefficient of variation of Kumandin vowel duration is in the limits of 5.49% - 18.04% (Tambovtsev, 2000-a: 285); in Ket the limit is 20.51% - 34.97% (Tambovtsev, 2000-b: 459).

Compactness of Different Language Taxons

Further we shall use the value of compactness to judge if some taxonomy or clusterisation is made correctly. It is shown in detail and vividly in the following Tables (cf. Table 1 - 3). One can use the fact that if two groups are mingled together, then the compactness of the mixture is usually greater than the compactness of each group. In this case, one can judge how homogenious a taxon is. One can measure it with the help of the value of TMB - coefficient. If the value of the TMB - coefficient becomes smaller, then the language, which is introduced in a taxon, belongs to this taxon. If it becomes greater, then the introduced language does not belong to this group since its sound chain is typologically too different. In fact, this coefficient verifies the similarity of the sound chain of a language to the similarity of the other languages in a language taxon. The more likely is a language to the other languages of the taxon, the greater it raises its compactness.

Usually genetically related languages have similar sound chains. The most similar are the consonants. This is why we took consonants as the basis for our study. The classical comparative method is built on the comparison of sounds. In this sense our method gives similar results.

Let us consider the sound similarities of the genetically related languages (Barannikova, 1973: 277 - 280):

Table 1
The Slavonic group of the Indo-European language family.

	mother	brother	sister	children	grand-child
Russian	mat'	brat	s'estra	d'et'i	vnuk
Ukrainian	mati	brat	sestra	d'iti	vnuk
Belorussian	matsi	brat	s'astra	dzetsi	unuk
Bolgarian	majka	brat	sestra	detsa	vnuk
Serbian	mati, majka	brat	sestra	dete	unuk

Czech	mater, matka	bratr	sestra	dite, deti	vnuk
Polish	mac', matka	brat	siostra	dzici	wnuk

Approximately the same similar sound picture is found in other genetically related languages.

Table 2
The Germanic group of the Indo-European language family.

	father	daughter	water	summer	good
German	Vater	Tochter	Wasser	Sommer	gut
English	father	daughter	water	summer	good
Dutch	vader	dochter	water	zomer	good
Swedish	fader	dotter	vatten	sommar	got

Table 3
The Roman group of the Indo-European language family and Latin.

	horse	cow	cat	land	big	bring
Latin	caballus	vacca	cattus	terra	grandis	portare
French	cheval	vache	chat	terre	grand	porter
Italian	cavallo	vacca	gatto	terra	grande	portare
Spanish	caballo	vacca	gato	tierra	grande	portar

Table 4
The Turkic language family.

	mother	father	daughter	horse	land	no
Tatar	ana	ata	kyz	at	zhir	juk
Azeri	ana	ata	gyz	at	(er)	joh
Kazah	ana	ata	kyz	at	zher	zhok
Turkmen	ana	ata	kyz	at	er-	jok
Uzbek	ona	ota	kiz	ot	er	juk
Bashkir	ana	ata	kyz	at	er	juk
Hakas	ine	aba	hyzy	at	chir	choh
Kumandin	ene	aba	kys	at	cher	chok
Altaj-kizhi	ene	ada	kyzym	at	ter	t'ok
Jakut	ije	aga	kys	at	sir	suoh
Tuvin	avaj	achaj	kyzym	at	cher	chok

One can see that the frequency of occurrence of all groups of consonants and even separate phonemes is equal or almost equal in these related languages. Genetic relatedness shows typological closeness from the point of view of the sound chains. Therefore, if two languages have similar sound chains they may be genetically related, but at the same time they are typologically similar. We can judge typological closeness by the value of compactness. Sometimes, however, languages become typologically close due to contact which they have had for a long time. This is why it is important to trace the contact of languages.

World languages are distributed in a rather narrow limit from the point of view of the frequency of occurrence of the eight defined groups of consonants. We shall see it quite vividly further while putting in order the values of each of the eight groups. In fact, these limits can be a sort of language universals.

We begin our analysis with the most compact language families. The most compact turned out to be the Mongolian language family. The Samoyedic and Turkic language families are also very compact. Then we shall analyse the other language families (Tungus-Manchurian, Finno-Ugric, Paleo-Asiatic, Indo-European, Austronesian, etc.). We shall also analyse the values of compactness of groups in the language families where groups are well-defined (e.g. Finno-Ugric and Indo-European). It is also worth while to consider superfamilies, unities and Sprachbunds such as Uralic, Altaic, Ural-Altaic and Balkan Sprachbund.

Compactness of Finno-Ugric Language Taxon

Let us consider the Finno-Ugric language taxon, beginning from the Ugric subgroup, which is said to include Mansi, Hanty and Hungarian. As a matter of fact, it is necessary to point out that we compare Hungarian not only to Finno-Ugric, but also to Turkic languages, because during its long history of development Hungarian had a lot of contact with Turkic languages. Daniel Abondolo, Bela Kalman and other linguists underline that Hungarian is not typical among the Uralic family (Abondodolo, 1990: 577).

At the first sight it is quite striking that Hungarian makes the compactness of the Turkic taxon higher (MV= 18.42, MTMB= 0.21). However, if one takes into consideration the ethnic contacts of the Hungarian during their history it is not so striking. It may be because of the ancient contacts of Hungarians with Turkic peoples in Siberia and then for the period of the Hungarians living for more than one thousand years on the Volga river. The speech communication between Hungarians and the Ob-Ugrians, i.e. Mansi (Vogul) and Hanty (Ostjak), is not possible. We must point out to the fact that the frequency of occurrence of eight consonantal groups turns out to be quite different. It is important to bear in mind that A. Marcantonio came to the conclusion that Hungarian is not a Uralic language. She writes about the history of attribution of Hungarian either to the Finno-Ugric family (J. Budenz) or to the Turkic family (A. Vambery). She points out that 52% of the Hungarian lexicon should not be considered Uralic. She finds only 19% of the Budenz core lexicon to be surely Finno-Ugric (Marcantonio, 2002: 37 - 48).

Let us indicate that the dispersion of the whole Finno-Ugric family (MV= 24.14%, MTMB= 0.47) is greater than that of its parts. Actually, it may be a sort of indication of the whole taxon having gaps, thus being rather dispersed than homogenious. Consequently, the Ugric and Baltic-Finnic languages obviously show different tendencies in the use of consonantal groups. The mean of the coefficient of variance (MV) in the subgroup of Ugric languages is 27.66%, MT = 0.47. The dispersion of the Baltic-Finnic subgroup is less (MV= 23.24%, MTMB= 0.35). The dispersion of the Volgaic subgroup (MV= 17.90%, MTMB= 0.13) is less than that of the Baltic-Finnic subgroup.

Let us make some experiments in mixing up different subgroups of the Finno-Ugric languages. Let us mix up the Ugric and Permic taxons. The dispersion of this united Ugro-Permic taxon (MV= 26.46%, MTMB= 0.46) is less than that of the Ugric taxon (MV= 27.66%, MTMB= 0.47). This fact indicates the similarity in the disribution of consonants in the Ugric and in Permic languages. One can state that Ugric and Permic languages are typologically closer to each other than to the other Finno-Ugric languages. Robert Austerlitz seems to have no solid foundation to put Permic and Volgaic languages into one taxon with the Baltic-Finnic languages (Austerlitz, 1990: 570). Our data show quite vividly that Ugric languages are closer to Volgaic languages (cf. Table 2, 3).

In order to prove this statement, we have to mix up the Ugric taxon with the Volgaic and Baltic-Finnic taxons.

The united taxon of Ugric and Volgaic languages has a greater dispersion (MV= 26.35%, MTMB= 0.45), than the united taxon of Volgaic and Baltic-Finnic languages (MV= 23.22%, MTMB= 0.35). It means that Volgaic and Baltic-Finnic languages have more typologically similar tendences. It is interesting to see if these tendences are preserved if we put some isolated Asiatic languages into the Finno-Ugric family. It may reveal if these isolated languages naturally belong there. Let us depict the ordered series (showing the increasing dispersion) after the introduction there the following languages: Ket (MV= 24.76%, MTMB= 0.49), Yukaghir (MV= 24.90%, MTMB= 0.50), Korean (MV= 24.91%, MTMB= 0.49), Japanese (MV= 25.06%, MTMB= 0.49), Nivhi (MV= 25.81%, MTMB= 0.54). Even Chinese shows a more similar typological tendency, than these genetically isolated languages (MV=

23.75%, MTMB= 0.46). The least similarity is demostrated with the Finno-Ugric languages Sweet Grass Cree, an American Indian language of Canada (MV= 26.62%, MTMB= 0.56).

Compactness of Samoyedic Language Taxon

Analysing the Samoyedic family, one should bear in mind that some Samoyedic languages do not have voiced consonants (e.g. Nenets and Sel'kup) while the Nganasan language suddenly has a great number of them in its sound chain (10.37%). It is a great problem since it makes them impossible to compare by this particular parameter. We have to resort only to the other 7 features, if we want to keep to the principle of commensurability. Nevertheless, it is still possible to compare the degrees of dispersion directly because we take the mean values of the coefficient of variance (V%) and TMB coefficient. The introduction of the isolated Asiatic language into the Samoyedic family (MV= 18.29%, MTMB= 0.16) may shed light on the problem of their typological similarity. It gives us the following ordered series: Japanese (MV= 19.09%, MTMB= 0.18), Ket (MV= 20.03%, MTMB= 0.18), Korean (MV= 21.57%, MTMB= 0.19), Yukaghir (MV= 27.40%, MTMB= 0.38), Nivhi (MV= 30.38%, MTMB= 0.49). It is quite vivid that all these isolated languages make the compactness of the Samoyedic family less. So, they do not belong there naturally.

Compactness of Uralic Language Taxon

Now let us consider the Uralic language taxon, which has been recently strongly criticised. Our investigation revealed that Uralic language taxon (MV= 28.31%, MTMB= 0.57) is more disperced than the Samoyedic (MV= 18.29%, MTMB= 0.16) or Finno-Ugric (MV= 24.14%, MTMB= 0.47) family. It shows that Uralic taxon is not natural, but rather artificial, i.e. created by linguists. Our data may give the Uralic linguists another impetus to reconsider the Uralic language taxon.

Let us make some experiments with the introduction (MV= 28.31%, MTMB= 0.57) of some isolated Asiatic languages into it to check their typological similarity. The introduction of these languages yielded the following ordered series: Ket (MV= 28.32%%, MTMB=

0.68), Korean (MV= 28.61%, MTMB= 0.60), Japanese (MV= 28.62%, MTMB= 0.63), Yukaghir (MV= 28.62%, MTMB= 0.65), Nivhi (MV= 29.15%, MTMB= 0.68), Sweet Grass Cree (MV= 29.77%, MTMB= 0.69). These values show that none of the given isolated language naturally belongs to Uralic taxon.

Comparing the Compactness of Uralic Language Taxon with other Language Taxons of World Languages

Let us consider Turkic language family (MV= 18,77%, MTMB= 0.21), wich is much more compact than Finno-Ugric (MV= 24.14%, MTMB= 0.47) family. Our data on campactness verify our intuitive linguistic impression that Turkic languages are more similar from the point of view of their sound pictures. Our data also confirm the closeness of the Turkic languages. Many linguists in the field of Turkic studies underline the closeness of Turkic languages (Isaev, 1978: 109; Musaev, 1984: 16). Probably, the Turkic languages are so close because in addition to their genetic relatedness they kept their intensive contacts for a long time, constantly mixing up with each other (Musaev, 1984: 33).

Besides, less time passed after the Turkic tribes lived compactly together on the limited rerritory in the Central Asia, than when the Finn-Ugric tribes lived together on a vaster territory (Baskakov, 1981: 60-69; Musaev, 1984: 31-33). The same is true for the Indo-European, Paleo-Asiatic and other language families. It goes without saying that languages in contact acquire some common features, and on the contrary, after losing contacts they have diverse tendencies, thus losing common features and becoming typologically diverse.

Let us deal with the Tungus-Manchurian family. Analysing the compactness of the Tungus-Manchurian taxon, cinsisting of 6 languages (Evenki, Nigidal, Nanai, Ul'ch and Oroch), one can conclude that its compactness is rather high. Indeed, its dispersion (MV= 18.60%, MTMB= 0.22) is as low as that of the Turkic (MV= 18.77%, MTMB= 0.21) family. If we state that the Turkic taxon is compact, than we should agree that the Tungus-Manchurian taxon is also compact. Actually, its compactness is almost two times higher than the compactness of the Paleo-Asiatic language family (MV= 33.89%, MTMB= 0.46). Let us check the reaction of the Tungus-Manchurian level of compactness (MV= 18.60%, MTMB= 0.22) to the introduction of some isolated languages on the one hand and on the other hand - the language, which belongs to the Tungus-Manchurian family, i.e. the Orok language. It turns out that the introduction of all isolated languages makes compactness less, but the introduction of Oroch makes it higher. Yukaghir (MV= 18.61%, MTMB= 0.24), Ket (MV= 18.75%, MTMB= 0.23), Korean (159.22%, TMB=2.01), Japanese (MV= 21.15%, MTMB= 0.28), Sweet Grass Cree (MV= 24.53%, MTMB= 0.33), Nivhi (MTMB= 22.88%, MTMB= 0.37). As a matter of fact, the introduction of the Orok langauge sharply diminishes the dispersion to (MV= 17.41%, MTMB= 0.20). It tells us that Orok is very much typologically similar. It also means that Orok surely belongs to the Tungus-Manchurian taxon. It is interesting to find out that the Chinese language also diminishes the dispertion of the Tungus-Manchurian taxon (MV= 18.05%, MTMB= 0.22), though very little. No wonder Chinese is similar because Tungus-Manchurian languages were in a tight contact with Chinese for many centuries.

Let us analyse the Paleo-Asiatic family in the same way. It is much less compact (MV= 33.89%, MTMB= 0.465) than the Tungus-Mancurian (MV= 18.60%, MTMB= 0.22) family. At least it is true if one takes 5 Paleo-Asiatic languages: Chuckchee, Koriak, Itelmen, and two dialects of the Eskomo language: Naukan and Imaklin. These dialects sometimes are regarded as separate languages because they have drifted too far away from each other and communication is not possible. Introducing the isolated languages, we receive the following ordered series: Japanese (V= 263.67%, TMB=4.21; MV= 32.96%, MTMB= 0.51), Korean (275.58%, TMB=4.21; MV= 34.45%, MTMB= 0.53), Chinese (304.55%, TMB=5.24), Sweet Grass Cree (V= 305.38%, TMB=4.99; MV= 38.17%, MTMB= 0.62), Ket (V= 319.83%, TMB=8.22; MV= 39.98%, MTMB= 1.03), Yukaghir (V= 342.61%, TMB=7.46; MV= 42.83%, MTMB= 0.93), Nivhi (V= 359.74%, TMB=9.28; MV= 44.97, MTMB= 1.16).

Japanese and Korean undoubtedly show some typological similarity to the Paleo-Asiatic languages in question.

A linguist often encountes a problem if the parent language should be included into the certain language taxon (subgroup, group, family, etc). It is solved differently in every case. It is well known that Latin is the parent language for all the Romanic languages (Italian, French, Spanish, etc.). Neverthless, Latin is not included in this group of the Indo-European family. It is said to belong to the Italic group of this family. At the same time Old Persian is included in the group of Iranian languages and Old Russian in the group of the Slavonic languages of the Indo-European family. Old Turkic is included into the family of the Turkic languages (JaDm, 1982: 18 - 21). The same problem is seen with the dead languages. Very often they are quite different to the modern language of the certain group. Therefore, we had a problem with the Old Russian, Old Turkic and Old English. Our metrico-typological method allows us to compare any language to any language taxon. Neverthless, it is reasonable to compare Latin with the Romanic languages, Old Russian to Slavonic languages, Old English to the Germanic languages, and Old Turkic to the modern Turkic languages. We can do it by measuring the compactness of the language taxons before and after the introduction of the certain language.

Let us analyse the compactness of the Germanic group of the Indo-European family with the Old English and without it. The dispertion of the Germanic group, which consists of English, Dutch, German, Norwegian, and Swedish (MV= 24.51%, MTMB= 0.29), is rather great. It is greater than that of the taxons of the higher hierarchy: e.g. Tungus-Manchurian (MV= 18.60%, MTMB= 0.22), the Samoyedic family (MV= 18.29%, MTMB= 0.16) Turkic (MV= 18.77%, MTMB= 0.21), or Mongolian (MV= 10.78%, MTMB= 0.08). These values may speak for the Germanic group to be a sort of a language family on the one hand, or Mongolian, Turkic, Samoyedic and Tungus-Manchurian families not to be families, but groups of some vaster family.

The frequency of occurrence of the labial consonants in the sound chain of the Old English language is typologically close to those of the other modern Germanic languages. One can construct the following ordered series: Modern English (13.05%) - Dutch (12.03%) - Danish (11.95%) - Old English (11.81%) - Norwegean (10.60%) - Swedish (11.00%) - German (9.88%). So, one can see that Old English is similar to Danish and Norwegean.

In the ordered series of the front consonants Old English takes place close to Norwegean: German (42.60%), Swedish (41.20%) - Modern English (40,79%) - Dutch (39.98%) - Old English (33.88) - Norwegean (32,70%) - Danish (29.40%).

It is possible to process all the rest of the features: palatal, back, sonorant, occlusive, fricative and voiced. The total of all eight features by the coefficient of fariance is 196.09% without the Old English, thus the mean is 24.51%.

After the introduction of Old English the total became less, that is 188.48%, thus the mean is 23.56%. Comparing the means (c.f. 24.51% and 23.56%) we can see that the dispersion became less. This speaks about the less variation. It means that Old English is similar to the Germanic languages in the Germanic taxon. It tells us that Old English naturally belongs to the Germanic taxon. Sometimes, it is more conveient to observe the tendency by the T-coefficient (2.32 and 2.30).

Let us check if the Indo-European isolated languages Armenian, Greek and Albanian. Introducing them one by one into the Germanic taxon we receive the following ordered series: Greek (V= 196.99%, TMB=2.64; MV= 24.62%, MTMB= 0.33), Albanian (V= 202.30%, TMB=2.69; MV= 25.29%, MTMB= 0.34). All these 3 languages make the compactness of the Germanic group of the Indo-European family less. Consiquently, they are not typologically similar to the Germanic languages.

Romanic languages are known to have originated from Latin. Therefore, it is quite interesting if their parent language naturally belongs to the Romanic taxon. Actually, the introduction of Latin into the Romanic taxon makes its compactness higher, thus the dispersion diminishes from MV= 27.81%, MTMB= 0.36 (V= 222.49%, TMB=2.84) to MV= 26.25%, MTMB= 0.33 (V= 209.97%, TMB=2.66). This means that Latin is typologically similar to Romanic language, and therefore naturally belongs to the Romanic taxon.

Let us consider the artificial language called "Esperanto". It is mainly constructed from the roots of Latin, though some roots of Greek, Italian, Spanish, Italian, French, English, Russian and Polish are also included. It is strictly defined in the use of special word modifyers. It sounds like an unknown Romanic language. The introduction of Esperanto into the group of Romanic languages diminishes its dispersion from MV= 27.81%, MTMB= 0.36 (222.49%, TMB= 2.84) to MV= 25.56%, MTMB= 0.32 (204.48%, TMB=2.54). The introduction of Esperanto makes the compactness of the Romanic taxon more than the introduction of Latin. This means that Esperanto typologically is more similar to the Romanic languages than Latin. Therefore, Esperanto naturally belongs to the Romanic language taxon.

Analysing the isolated languages, one can notice that Greek (V = 214.78%, TMB= 2.75; MV= 26.85%, MTMB= 0.34) is closer to the Romanic languages in general than Albanian (V= 217.08%, TMB= 2.93; MV= 27.13%, MTMB= 0.37), or Armenian (V= 218.56%, TMB= 2.89; MV= 27.32, MTMB= 0.36).

It was possible to get the data on the frequency of the occurrence of phonemes in the texts of Hindi, Bengali, Gudjarati and Marathi from different publications (for details see Tambovtsev, 2001-a: 5 - 14). The phonostatistic Gypsy data were obtained on the material of the folklore songs of the Kolderarey Gypsy (Tambovtsev, 2001-a: 15 - 16).

The received dispertion of the Indic grpoup (V= 163.20%, TMB= 1.83; MV= 20.40, MTMB= 0.23) was tested by the introduction of the ancient Indic languages and iseolated languages one by one. As a result we receive the following ordered series: Albanian (V= 156.59%, TMB=1.97; MV= 19.57%, MTMB= 0.25), Greek (V= 167.13%, TMB=2.14; MV= 20.89%, MTMB= 0.27), Maharashtri (V= 168.33%, TMB=2.25; MV= 21.04%, MTMB= 0.28), Sanscrit (V= 168.77%, TMB=2.44; MV= 21.10%, MTMB= 0.31), Armenian (V=

169.08%, TMB= 2.68; MV= 21.14%, MTMB= 0.34), Vedic (V= 193.11%, TMB=3.10; MV= 24.14%, MTMB= 0.39). It is vivid from these data that Albanian enters typologically in the group of Indic languages. Greek has more similar sound chains than Sanskrit or the Middle Indian Prakrit Maharashtri.

One could notice in the ealier discussion that Old English and Latin typologically fit into their corresponding groups: Germanic and Romanic. Probably the modern Indic languages diverged from their original typology of distribution in the sound chain to a greater degree.

Our Iranic group of the Indo-European family includes the following languages: Iranian, Tadjic, Osetian, and Kurdish. One can observe that it is very compact (V= 105.64%, TMB=0.68; MV= 13.21%, MTMB= 0.09). It is about 2 times more compact than the Indic group. All isolated Indo-European languages make their compactness higher, thus, the dispersion less: Albanian (V= 125.60%, TMB=1.22; MV= 15.70%, MTMB= 0.15), Armenian (V= 138.18%, TMB= 1.49; MV= 17.27%, MTMB= 0.19), Greek (V= 139.74%, TMB= 1.58; MV= 17.46%, MTMB= 0.20). Therefore, typologically they are quite different.

Let us consider the Slavonic group of the Indo-European family. We managed to include in it nearly all the Slavonic languages. All the main Slavonic languages are presented for the analysis. Our Slavonic groups includes Eastern Slavonic languages (Russian, Belorussian and Ukrainian), Western (Sorbian, Czech, Slovak and Polish), and Southern (Serbo-Croatian and Bulgarian). Its dispersion (V= 121.71%, TMB= 1.34; MV= 15.21%, MTMB= 0.17) is greater than that of the Iranic group, but less than that of Indic (MV= 20.40%), Germanic (MV= 24.51%) or Romanic (MV= 27.81%).

Some linguists think the Balto-Slavonic language unity to have existed for not so long a time. This is why it is advisable to consider the united taxon of the Baltic (Lithuanian and Latvian) and Slavonic languages.

First of all, let us introduce Lithuanian into the Slavonic taxon. Lithuanian makes the compactness of the united taxon higher, while the dispersion diminishes (V= 121.55%, TMB= 1.35; MV= 15.19% MTMB= 0.17). The introduction of Latvian makes the compactness less and dispersion greater (V= 125.77%, TMB= 1.43; MV= 15.73%, MTMB= 0.18). It means that Lithuanian typologically belongs to the Slavonic group, while Latvian does not belong to it. At the same time Latvian is closer to the Slavonic languages than the isolated Ind-European languages: Greek (V= 134.37%, TMB=1.67; MV= 16.80%; MTMB= 0.21), Albanian (V= 143.63%, TMB= 1.77; MV= 17.95%, MTMB= 0.22) and Armenian (V= 155.59%, TMB=2.09; MV= 19.45%, MTMB= 0.26).

Some linguists think that the Indo-European genealogical tree should be reconsidered in order to make it more correct, though they still think it is possible to construct such a tree. The majority of the linguists in the field of Finno-Ugristics and Uralistics seem to be quite happy with their old classical tree. However, the majority of the Turkologists do not think it is possible to construct a true Turkic tree, because many features in Turkic languages interface and because the Turkic languages are too close to each other.

Besides, one should bear in mind the sceptical attitude of many linguists to the theory of the genealogic language tree in general, whose number has grown over the last decades (Zhirmunskij, 1964: 104; Chemodanov, 1982: 94-98). However, other linguists think that the rejection of the genealogical scheme in the form of the genealogical tree must not cause the rejection of the genealogical relatedness of the languages, though the foundations of the language classifications should undergo a closer attention and stricter verification (Sharedzenidze, 1982: 71). In fact, Ago Kuennap and Angela Marcantonio believe that it is high time to reconsider some of the language families. They consider it quite wrong to call the Uralic set of languages "a family", since their genetic relationship has not been properly proved (Marcantonio, 2002). One can hope that our phonostatistical typological data may give lots of new material to reconsider different language taxons. It may allow linguists to verify some language taxons and to reject others.

It is necessary to point out that the compactness of language taxons, even by one parameter, is a good and fine tool to measure the similarity of languages. Let us take, for instance, the value of the consonantal coefficient, which is measured by the frequency of consonants and vowels in the sound chain. In fact it is the ratio of consonants to vowels Tambovtsev, 1986). Let us analyse the Ural-Altaic language union. We must again underline that it should not be considered a family, but a taxon of the higher level. It should be a taxon of the higher level since Fiino-Udric, Samoyedic, Turkic, Tungus-Manchurian and Mongolian families, which are united in this language taxon are themselves families. It is necessary to distinguish between language families and language unions or other taxons (Barannikova, 1973: 299-306; Shirokov. 1985: 100 - 187).

Let us make some experiments with the Turkic and Finno-Ugric languages. Let us suppose that once they were one language taxon. Let us construct this conditional language taxon. The dispersion of the united Turkic-Finno-Ugric taxon is presented without Old Turkic.

Let us recall that the dispersion of the Turkic taxon is less (MV= 18.77%) than the dispersion of the Finno-Ugric taxon (24.14%). On the contrary, the compactness of the Turkic family is higher than that of the Finno-Ugric family. The compactnes of the united Turkic-Finno-Ugric mixture is 12.39, while that of the Turkic family is 32.16 and Finno-Ugric family only 8.32, that is about 4 times less. The mixture does not prove natural, because it makes the dispertion greater. One can see that Finno-Ugric languages make the compactness of the Turkic taxon much less. It is possible to conclude that the Finno-Ugric languages cannot be mixed with the Turkic languages, because they are typologically different.

Let us use the same procedure for the Turkic (V= 150.16%, MV= 18.77%) and Mongolian (V= 86.24%. MV= 10.78%) families. Let us unite the Turkic and Mongolian families into one taxon. The dispersion of the mixture of the Turkic and Mongolian languages is much greater (V= 170.90%), than the dispersion of each of the families in question.

The united group of the Samoyedic and Finno-Ugric languages is called the Uralic language unity. The compactness of the Uralic (V= 226.49%, TMB= 4.58; MV= 28.31%, MTMB= 0.57) taxon is less than those of the Samoyedic (V= 128.04%, TMB= 1.10; MV= 18.29%, MTMB= 0.16) and Finno-Ugric taxons (V= 193.13%, TMB= 3.77; MV= 24.14%, MTMB= 0.47). This means that they are a sort of artificial language taxon. It may also mean that those linguists (e.g., Kuennap, 1998, 2003; Marcantonio, 2002) who want to reconsider the Uralic taxon are right. They do not believe the Uralic taxon to be a genetic family.

Though some linguists believe the Altaic languages to be a genetic taxon, that is, a language family (e.g. Crystal, 1992: 14), it is better to be on the safe side and to consider it a language unity. The Altaic language unity is said to consist of the Turkic, Mongolic and Tungus-Manchurian families.

The value of the compactness of Turkic and Tungus-Manchurian families is more or less on the same level. The dispersion of the Altaic taxon is greater (V= 207.76%), than those of the Turkic, Mongolic and Tungus-Manchurian families. It means that the languages of these 3 families are not typologically similar. This does not support the Altaic theory.

Robert Austerlitz calls the Ural-Altaic taxon as "a superstructure, a unit larger than a family". He proposes to call it a "stock or phylum" (Austerlitz, 1990: 569). The Ural-Altaic taxon embraces the Uralic and Altaic languages. It is more disperse (V= 247.82%) than each of them. It may mean that Ural-Altaic is quite an artificial taxon.

However, we may look upon the Altaic taxon from the other angle. That is, from the point of view of the degree of the dispersion of the Indo-European family, whose dispersion is greater (V= 223.79%), than that of the Altaic (V= 207.76%) taxon. This speaks for the Altaic theory.

Really, if the majority of linguists support the idea of the Indo-European family, then one should think twice before rejecting the Altaic taxon as a family, because the compactness of Altaic is greater.

Let us test our method on the material of the taxon, which is called the Balkan language union or Sprachbund. Usually, it includes 3 languages of different groups of the Indo-European family, together with 2 isolated languages. Bulgarian and Serbo-Croatian are from the Slavonic group. Rumanian is from the Romanic group. Greek and Albanian are isolated Indo-European languages. It is a problem if Armenian should also be included in the Balkan language union.

Some linguists insist on including Armenian in the union, others are against it. Let us try to solve this problem by our method. It is also interesting to see how compact the group made up of different groups of the Indo-Eropean family is. Likewise it is interesting to find out if Turkish is similar to them typologically. Surely, close contacts between Turkish and the Balkan languages must have some typological similarity as a result. For several centuries Turkish heavily influenced the languages of the conquered Balkan peoples. The dispersion of the Balkan language union (V= 237.89, TMB= 2.90; MV= 29.74%, MTMB= 0.36) is not much greater than that of Uralic, or Altaic taxons. Ural-Altaic taxon is more disperse than the Balkan taxon.

Let us put Armenian and Turkish into the Balkan taxon. The introduction of Armenian makes its dispersion less (V= 234.23%, TMB=3.13; MV= 29.28%. MTMB= 0.39). It may indicate that one is correct to put Armenian into it since Armenian is typologically similar. The introduction of Turkish into the the Balkan language taxon makes its dispersion even lower (V= 224.99%, TMB=2.89; MV= 28.12%, MTMB= 0.36), It means that the Balkan languages are typologically very similar to Turkish.

Let us consider the Austronesian family, which includes Tagalog, Indonesian, Dayak, Hawaiian and Samoan. Its dispersion is rather high (V= 369.65%, TMB= 7.22; MV= 46.21%, MTMB= 0.90). Maybe because of its high dispersion Japanese naturally enters into it, making its dispersion less (V= 340.33%, TMB= 6.49; MV= 42.54%, MTMB= 0.81).

Speaking about the Aboriginal languages of Australia, one should bear in mind that thet do not have the class of the fricative consonants. This is why it is advisable to compare only the means, not the totals. The other warning is that the distribution of the voiced consonants is very diverse, which may mean that there is no proper opposition by voiced consonants. So, if we consider only 6 features (out of 8) of the taxon of the Australian aboriginal languages, then it turns rather compact. The mean dispersion without the data on the voiced consonants is rather low (MV= 15.84%, MTMB= 0.29). With voiced consonants, its dispersion is greater (MV= 21.51, MTMB= 0.55). Anyway, it is much more compact than the taxon of the languages of American Indians (V=346.95%, TMB=8.59; MV= 43.37%, MTMB= 1.07). If we put the Ket (Yug) language into this taxon, we find out that its dispersion diminishes (V= 345.81%, TMB= 9.21; MV= 43.23%, MTMB= 1.15). It may speak for the typological closeness of Ket with the languages of American Indians. The introduction of Nivhi (V= 355.87%, TMB= 9.71; MV= 44.48%, MTMB= 1.21) or Yukaghir (V= 349.49%, TMB= 9.08; MV= 43.69%, MT+ 1.14) makes the dispersion higher. Thus, neither Nivhi, nor Yukaghir naturally belong to the taxon of American Indian languages.

Unfortunately, the journal space is limited, therefore, we could not show all the details. The details can be found elsewhere (Tambovtsev, 1994-a; 1994-b; 2001-a; 2001-b; 2001-c; 2003).

Conclusions

Our data show that it is possible to make the following conclusions:

1. The sound chains of the Uralic languages show that Uralic taxon is not natural from the typological point of view, but rather artificial, i.e. created by linguists.

2. Labial, front, palatal, back, sonorant, occlusive, fricative and voiced consonants have rather strict limits of occurrence not only in Finno-Ugric, but also in world languages.
3. The group of front consonants is used in different language taxons more compactly than the other 7 consonantal groups. The next compact group is occlusive consonants.
4. The frequency of occurrence of the voiced consonants is the most changable and unstable feature among the eight chosen features. It may be absent in some languages.
5. By the total of the 7 chosen features (without voiced consonants) the most compact is the Mongolic family. Then comes Samoyedic, Turkic, Tungus-Manchurian and Finno-Ugric family. The Paleo-Asiativ family is demonstrates such dispersion that it looks more like a language unity than a family.
6. Usually, a language subgroup is more compact than a group, and in its own turn a group is more compact than a family. A language family is more compact than a language unity, and so on. The least compact is the language taxon that includes all the languages of the world. The mixture of two taxons (i.e. subgroups, groups, familes, etc) is usually less compact than each of these taxons.
7. It is high time to reconsider the place of Hungarian in the Ugric language taxon.

One can see that the main conclusion of this article is that a typologo-metrical approach on phonological level shows that that the Uralic taxon is not natural, but looks like a mechanical conglomeration of different languages of Finno-Ugric and Samoyedic origin, put in one set by linguists for classification purposes. However, this artificial language taxon is quite useful for the students of linguistics to embrace all the scope of Finno-Ugric and Samoyedic language. Our data may give the Uralic linguists more impetus to reconsider the Uralic language taxon.

REFERENCES

1. Abondolo, 1990 - Abondolo, Daniel. Hungarian. - In: The World's Major Languages (edited by Bernard Comrie). Oxford: Oxford University Press, 1990, p. 577 - 592.
2. Apresjan, 1966 – Apresjan, Yuri D. Algoritm postroenija klassov po matritse rasstojanij. [Algorithm of Constructing Classes by the Matrix of Distances]. – Mashinnyj perevod i prikladnaja lingvistika. Vol. 9. – Moscow: Moscow State Pedagogical Institute of Foreign Languages, 1966, p.72 - 79.
3. Arkadjev et al., 1964 – Arkadjev A. G., Braverman E. M. Obuchenije mashiny raspoznavaniju obrazov. [Teaching a Machine Pattern Recognition]. – Moscow: Nauka, 1964. – 112 pp.
4. Arkadjev et al., 1971 – Arkadjev A. G., Braverman E. M. Obuchenije mashiny klassifikatsii objectov. [Teching a Machine Object Classification]. – Moscow: Nauka, 1971. – 192 pp.
5. Austerlitz, 1990 - Austerlitz, Robert. Uralic languages. - In: The World's Major Languages. (Edited by Bernard Comrie). - Oxford / New York: Oxford University Press, 1990, p.567 - 576.
6. Barannikova, 1973 – Barannikova, Lidija Ivanovna. Vvedenie v jazykoznanie. [Introduction into Linguistics]. - Saratov: Saratov University, 1973. - 384 pp.
7. Baskakov, 1981 – Baskakov, Nikolaj Aleksandrovich. K istoriko-tipologicheskoj fonologii tjurkskih jazykov. [On Historico-typological Phonology of Turkic Languages]. - in: Voprosy Jazykoznanija, # 1, 1981, p. 60 - 69.
8. Bongard, 1967 – Bongard, Mihail Moiseevich. Problema uznavanija. [The Problem of Recognition]. – Moscow: Nauka, 1967. – 320 pp.
9. Butler, 1985 - Butler, Christopher. Statistics in Linguistics. - Oxford: Basil Blackwel, 1985. - 214 pp.
10. Chemodanov, 1982 – Chemodanov N. S. Germanskie jazyki. [Germanic languages]. - In: Sravniteljno-istoricheskoe izuchenie jazykov raznyh semej. - Moskva: Nauka, 1982, p. 91 - 115.
11. Crystal, 1992 - Crystal, David. An Encyclopedic Dictionary of Language and Languages. - Oxford: Blackwell, 1992. - 428 pp.
12. Dide, 1985 - Dide E. Metody analiza dannyh: podhod, osnovannyj na metode dinamicheskih sgushchenij. - Moskva: Finansy i ststistika, 1985.
13. Drujanov L. A. Zakony prirody i ih poznanije. - Moskva: Prosveshchenije, 1982.
14. Eliseeva et al, 1977 – Eliseeva I. I. and Rukavishnikov V. O. Gruppirovka, korreljatsija, raspoznavanie obrazov. (Statisticheskie metody, klassifikatsija i izmerenie svjazej). [Groupping, Correlation, Pattern Recognition. (Statistical Methods, Classification and Connection)]. – Moscow: Statistika, 1977. – 144 pp.
15. EK, 1975 - Entsyklopedija kibernetiki. [Encyclopaedia of Cybernetics]. - Kiev: Ukrainian Sovetsk. Entsiklopedija, 1975. - 607pp.
16. Gamkrelidze et al., 1984 – Gamkrelidze, Tomaz V. and Ivanov, Vjacheslav Vs. Indoevropejskij jazyk i Indoevropejtsy. [The Indo-European Language and the Indo-Europeans]. (Part 1 and 2). – Tbilisi: Tbilisi State University, 1984. - 1329 page.
17. Gadzhieva N. Z. K voprosu o klassifikatsii t'urkskih jazykov i dialektov. - In: Teoreticheskie osnovy klassifikatsii jazykov mira. - Moskva: Nauka, 1980, p. 100 - 126.
18. Hornby A. S. (with assistance of Christina Ruse). Oxford student's dictionary of current English. - Oxford: Oxford university press, 1984.
19. Hurgin Ja. I. Kak ob'jat' neob'jatnoje? - Moskva: Znanije, 1979.
20. Isaev, 1978 – Isaev M. I. O jazykah narodov SSSR. [On the Languages of the Peoples of the USSR]. - Moscow: Nauka, 1978. - 222 pp.

21. Jartseva, 1980 - Jartseva V. N. "Jazykovoj tip" sredi sopredeljnyh ponjatij. ["Language type" among similar notions]. - In: Teoreticheskie osnovy klassifikatsii jazykov mira. [Theoretical Fundamentals of Classification of World Languages]. - Moscow: Nauka, 1980, p. 24 - 61.
22. JaDm, 1982 - Jazyki i dialekty mira. [Languages and Dialects of the World]. - Moskva: Nauka, 1982. - 208 pp.
23. Kalman, 1988 - Kalman, Bela. The history of the Ob-Ugric languages. - In: The Uralic Languages. (Edited by Denis Sinor). - Leiden\New York: E. J. Bril, 1988, p. 395 -412.
24. Kondakov, 1975 – Kondakov, N. I. Logicheskij slovar'-spravochnik [Logical Dictionary-Reference Book]. – Moscow: Nauka, 1975. – 720 pp.
25. Kordonskij, 1983 - Kordonskij Simon G. Taksony i analiticheskie objekty kak sistemy. [Taxons and analytical objects as systems]. - In: Sistemnyj metod i sovremennaja nauka. - Novosibirsk: Novosibirsk, 1983, c. 141 - 150.
26. Kordonskij S. Tsikly dejatel'nosti i ideal'nyje objekty. - Moskva: Pantori, 2001.
27. Kuennap, 1998 - Kuennap, Ago. Breakthrough in Present-day Uralistics. - Tartu: University of Tartu, 1998.
28. Kuennap, 2003 - Kuennap, Ago. Review on the book by Angela Marcantonio "The Uralic Language Family: Facts, Myths and Statistics. Oxford (UK) and Boston (USA): Blackwell Publishers, 2002, 335 pp. - In: Eurasian Studies Yearbook, # 75, 2003, p. 171 - 176.
29. Kun T. Struktura nauchnyh revoll'utsij. [Kuhn T. S. The structure of scientific revolutions]. - Moskva: Progress, 1977. - 300 pp.
30. Marcantonio, 2002 - Marcantonio, Angela. The Uralic Language Family. Facts, Myths and Statistics. - Oxford: Blackwell Publishers, 2002. - 335 pp.
31. Martynenko G. Ja. Osnovy stilemetrii. - Leningrad: Leningrad university press, 1988.
32. Martynov, 1983 – Martynov V. V. Jazyk v prostransve i vremeni. K probleme glottogeneza slavjan. [Language in Space and Time, To the Problem of Glottogenesis of the Slavs]. – Moscow: Nauka, 1983. - 108 pp.
33. Mayr, 1971 - Mayr E. Printsipy zoologicheskoj sistematiki [Principles of Zoological Systematics]. - Moscow: Mir, 1971. - 454
34. Meljnikov, 1978 - Meljnikov, Gennadij Prokopjevich. Sistemologija i jazykovye aspekty kibernetiki. [Systemology and the Language Aspects of Cybernetics]. - Moscow: Sovetskradio, 1978. - 368 pp.
35. Mel'nikov, 2003 - Mel'nikov, Gennadij Prokop'evich. Sistemnaja tipologija jazykov: Printsypy, metody, modeli. [Systemic Typology of Languages: Principles, Methods, Models]. - Moskva: Nauka, 2003. 395 pp.
36. Mirkin, 1976 - Mirkin Boris G. Analiz kachestvennyh priznakov (matematicheskije metody i modeli). [Analysis of quantitative features (mathematical methods and models)]. - Moskva: Statistika, 1976.
37. Mirkin, 1980 - Mirkin B. G. Analiz kachestvennyh priznakov i struktur. - Moskva: Statistika, 1980.
38. Musaev, 1984 – Musaev K. M. Leksikologija tjurkskih jazykov. [Lexicology of Turkic Languages]. - Moscow: Nauka, 1984. - 228 pp.
39. OALD, 1998 - Oxford Advanced Learner's Dictionary of Current English. - Oxford: Oxford University Press, 1998. - 1428 pp.
40. Ozhegov, 1970 - Ozhegov S. I. Slovar' russkogo jazyka. [Dictionary of the Russian language]. - Moskva: Sovetskaja entsiklopedija, 1970.
41. Pavlovskij, 1967 - Pavlovskij Z. Vvedenije v matematicheskuju statistiku. - Moskva: Statistika, 1967.
42. Petrov, 1974 – Petrov, Vasilij Viktorovich. Kriticheskij analiz problemy nesoizmerimosti v sovremennoj "filosofii nauki". [Critical analysis of the problem of

incommensurability in modern "philosophy of sciences". - Novosibirsk: Novosibirsk University, 1974. - 18 pp.

43. Rassadin, 1988 - Rassadin Valentin I. Rol' kontaktov v obrazovanii t'urksko-mongol'skoj jazykovoj obshchnosti. [The role of contacts in the formation of the Turkic-Mongolian language unity]. - In: Voprosy Jazykoznanija, # 4, 1988, p.103-108.
44. Richmond, 1964 - Richmond S. B. Statistical analysis. - New York: The Ronald press company, 1964.
45. Rozova, 1986 - Rozova, Stalina Sergeevna. Klassifikatsionnaja problema v sovremennoj nauke. [Classification Problem in Modern Science]. - Novosibirsk: Nauka, 1986. - 222 pp.
46. Porhomovskij, 1982 - Porhomovskij V. Ja. Problemy geneticheskoj klassifikatsii jazykov Afriki. [Problems of genetic classification of African languages]. - In: Teoreticheskie osnovy klassifikatsii jazykov mira. [Theoretical Fundamentals of Classification of World Languages]. - Moscow: Nauka, 1982, p. 258 - 312.
47. Serebrennikov, 1982 - Serebrennokov, Boris Aleksandrovich. Problema dostatochnosti osnovanij v gipotezah, kasajushchihs'a geneticheskogo rodstva jazykov. [The problem of sunstantuated foundations in the hypothesis concerning the genetic relatedness of languahes]. - In: Teoreticheskie osnovy klassifikatsii jazykov mira. Problemy rodstva. - Moskva: Nauka, 1982, p.6 - 62.
48. Sharadzenidze, 1982 - Sharadzenidze T. S. Rodstvo jazykov, protsessy divergentsii, konvergentsii i sootvetstvujushchie im klassifikatsii jazykov mira. [Genetic relatedness of languages, the processes of divergence, convergence and the classifications of world languages corresponding to them]. - In: Teoreticheskie osnovy klassifikatsii jazykov mira. Problemy rodstva. - Moskva: Nauka, 1982, p. 63 - 107.
49. Shirobokova, 2001 - Shirobokova, Natalia Nikolaevna. Istoricheskoe razvitie Yakutskogo konsonantizma. [Historical Development of Yakut Consonant System]. - Novosibirsk: Nauka, 2001. - 151 pp.
50. Shirokov, 1985 - Shirokov, Oleg Sergeevich. Vvedenie v jazykoznanie. [Introduction into Linguistics]. - Moskva: Moscow University Press, 1985. - 263 pp.
51. Sis'kov V. I. Ob opredelenii velichiny intervala pri gruppirovkah. - Vestnik statistiki, # 12, 1971, p. 3 - 11.
52. Sis'kov V. I. Korrel'atsionnyj analiz v ekononicheskih issledovanijah. - Moskva: Statistika, 1975.
53. Tambovtsev, 1976 - Tambovtsev, Yuri. Raspoznavanie fonem chelovekom. [Recognition of phonemes by man]. - In: Empiricheskoe predskazanie i raspoznavanie obrazov. Vychisliteljnye sistemy. Volume 67, Novosibirsk: Institute of Mathematics of the Siberian Branch of Russian Academy, 1976, p. 161-164.
54. Tambovtsev, 1986 – Tambovtsev. Yuri. Konsonantnyj koeffitsijent v jazykah raznyh semej. [The Consonant Coefficient in the Languages of Different Language Families]. - Novosibirsk: Novosibirsk University, 1986. - 16 pp.
55. Tambovtsev, 1990 – Tambovtsev, Yuri. Kompaktnost' finno-ugorskoj jazykovoj sem'ji po dannym konsonantnogo koeffitsienta. [Compactness of the Finno-Ugric Language Family by the Data of the Consonantal Coefficient]. - In: Linguistica Uralica, Vol. 26, №1, 1990, p.13-20.
56. Tambovtsev, 1994-a - Tambovtsev, Yuri. Dinamika funktsionorovanija fonem v zvukovyh tsepochkah jazykov razlichnogo stroja. [Dynamics of Functioning of Phonemes in the Sound Chains of the Languages of Different Structure]. - Novosibirsk: Novosibirsk University, 1994. - 133 pp.
57. Tambovtsev, 1994-b - Tambovtsev, Yuri. Tipologija uporjadochennosti zvukovyh tsepej v jazyke. [Typology of the Orderliness of Sound Chains in Language]. - Novosibirsk: Novosibirsk University, 1994. - 199 pp.

58. Tambovtsev, 2001-a - Tambovtsev, Yuri. Kompendium osnovnyh statisticheskih harakteristik funktsionirovanija soglasnyh fonem v zvukovoj tsepochke anglijskogo, nemetskogo, frantsuzskogo i drugih indoevropejskih jazykov. [Compendium of the Basic Statistical Characteristics of Functioning of Consonants in the Sound Chain of English, German, French and Other Indo-European Languages]. - Novosibirsk: Novosibirskij klassicheskij institut, 2001. - 129 pp.
59. Tambovtsev, 2001-b - Tambovtsev, Yuri. Funktsioniravanie soglasnyh v zvukovoj tsepochke uralo-altajskih jazykov. [Functioning of Consonants in the Sound Chains of Ural - Altaic Languages]. - Novosibirsk: Novosibirskij klassicheskij institut, 2001. - 132 pp.
60. Tambovtsev, 2001-c - Tambovtsev, Yuri A. Nekotorye teoreticheskie polozhenija uporjadochennosti fonem v zvukovoj tsepochke jazyka i kompendium statisticheskih harakteristik osnovnyh grupp soglasnyh fonem. [Some Theoretical Foundations of the Typology of Orderliness of Phonemes in the Sound Chain of Language and the Compendium of the Statistical Characteristics of the Basic Groups of Consonants]. - Novosibirsk: Novosibirskij Klassicheskij Institut, 2001. - 130 pp.
61. Tambovtsev, 2001-d - Tambovtsev, Yuri A. The phonological distances between Mongolian and Turkic languages based on typological consonantal features. - In: Mongolian studies. Journal of the Mongolia Society (USA), Vol.24, 2001, p.41 - 84.
62. Tambovtsev, Yuri. Comparative typological study of language distances based on the consonants in sound chains of various languages. - In: The 5th National Colloquium for Computational Linguistics in the UK. Proceedings of the Conference. (Edited by John Elliot). 8-9 January, 2002. University of Leeds, UK. - Leeds: University of Leeds, 2002, p.77-80.
63. Tambovtsev, Yuri. Is Kumandin a Turkic language? - In: Dilbilim Arashtirmalari (Istanbul), 2002, p. 63 - 104.
64. Tambovtsev, Yuri. Korean and Japanese as Members of the Altaic Language Family. - In: Abstracts. Permanent International Altaistic Conference 45th Meeting, Budapest June 23-28, 2002. Budapest: Research Group for Altaic Studies. Hungarian Academy of Sciences, 2002, p. 13-14.
65. Tambovtsev, Yuri. Structure of the frequency of occurrence of consonants in the speech sound chain as an indicator of the phono-typological closeness of languages. - In: ALL - ACH 2002. New Directions in Humanities Computing. The 14th Joint International Conference, University of Tuebingen, 24 - 28 July, 2002. Conference Abstracts. - Tuebingen: Universitaet Tuebingen, 2002, p. 138 - 139.
66. Tambovtsev, 2002 – Tambovtsev, Yuri. Lingvisticheskaja taksonomija: kompaktnost' jazykovyh podgrupp, grupp i semej. [Linguistical taxonomy: Compactness of subgroups, groups and fasmilies]. - In: Baltistica. Journal of Baltic Linguistics (Vilnius), Volume 37, #1, 2002, p. 131 - 161.
67. Tambovtsev, 2003-a - Tambovtsev, Yuri A. Phonological similarity between Basque and other world languages based on the frequency of occurrence of certain typological consonantal features. - In: Abstracts of the 17th International Congress of Linguists. Prague, Czech Republic, July 24 - 29, 2003. - Prague: Charles University, 2003, p.358 - 359.
68. Tambovtsev, 2003-b - Tambovtsev, Yuri Alekseevich. Tipologija funktsionirovanija fonem v zukovoj tsepochke indoevropejskih, peleoaziatskih, uralo-altajskih i drugih jazykov mira: kompaktnost' podgrupp, grupp, semej i drugih jazykovyh taksonov. [Typology of Functioning of Phonemes in the Sound Chain of Indo-European, Paleo-Asiatic, Uralo-Altaic and Other World Languages: Compactness of Subgroups, Groups, families and Other Language Taxons]. - Novosibirsk: Sibirskij Nezavisimyj Instutut, 2003. - 143 pp.

69. Tambovtsev, 2003-c - Tambovtsev, Yuri A. Izmerenije fonostatisticheskih rasstojanij mezhdu ural'skimi jazykami. [Measuring phono-statistical distances between Uralic Languages]. - In: Fenno-Ugristica, 25. - Tartu: Tartu University Press, 2003, p. 120 - 168.
70. Tambovtsev, 2003-d - Tambovtsev, Yuri. Phonological similarity between Tofa and some other lnaguages of Central Asia based on the frequency of occurrence of certain typological consonantal features. - In: Fenno-Ugristica, 25. - Tartu: Tartu University Press, 2003, p. 169 - 179.
71. Tret'jakov, 1970 – Tret'jakov P. N. U istokov drevnerusskoj narodnosti. [At the Source of the Old Russian people]. – Leningrad: Nauka, 1970, - 156 pp.
72. Tret'jakov, 1982 – Tret'jakov P. N. Po sledam drevnih slavjanskih plemen. [Along the Traces of the Ancient Slavonic Tribes]. - Leningrad: Nauka, 1982. - 143 pp.
73. Trubetzkoy, 1939 - Trubetzkoy Nikolai Sergeevich. Gedanken ueber das Indogermanenproblem. - In: Acta Linguistica, Vol.1, Copenhahague, 1939, p. 81 - 89.
74. Vasiljev, 1969 - Vasiljev, Vladimir Ivanovich. Raspoznajushchije sistemy. [Recognition Systems]. - Kiev: Naukova Dumka, 1969. - 392 pp.
75. Venchikov et al., 1974 - Venchikov A. I., Venchikov V. A. Osnovnye prijomy statisticheskoj obrabotki rezul'tatov nabl'udenij v oblasti fiziologii. - Moskva: Meditsina, 1974.
76. Vinogradov, 1973 - Vinogradov, Victor Alekseevich. Metody Tipologii. Jazykovye shodstva i metody ih issledovanija. [Methods of Typology. Language Similarities and the Methods of their Investigations]. - In: Obshchee Jazykoznanie. Metody Lingvisticheskih Issledovanij. - Moskva: Nauka, 1973, p. 224 - 256.
77. Vinogradov, 1982 - Vinogradov, Victor Alekseevich. Funktsionaljno - tipologicheskie kriterii i genealogicheskaja klassifikatsija jazykov. [Functional - tipological criteria and the genealogical classification of languages]. - In: Teoreticheskie osnovy klassifikatsii jazykov mira. [Theoretical Fundamentals of Classification of World Languages]. - Moscow: Nauka, 1982, p. 258 - 312.
78. Webster's seventh new collegiate dictionary. - Springfield: Merriam company, 1965.
79. Zagorujko, 1972 – Zagorujko, Nikolai Grigorievich. Metody raspoznavanija i ih primenenie. [Methods of Recognition and their Application]. – Moscow: Sovetskradio, 1972. - 208 pp.
80. Zagorujko, 1999 - Zagorujko, Nikolaj Grigorievich. Prikladnyje metody analiza dannyh i znanij. [Applied Methods of Data and Knowledge Analysis]. - Novosibirsk: Institute of Mathematics of the Siberian Branch of the Russian Academy, 1999. - 269 pp.
81. Zagorujko et al., 1980 – Zagorujko Nikolaj G. and Yuri Tambovtsev Statisticheskoe raspredelenie udarenija v russkom slove. [Statistical Distribution of Stress in Russian Word]. - In: Materialy seminara "Statisticheskaja optimizatsija prepodavanija jazykov i inzhenernaja lingvistika." - Chimkent: Chimkent Pedagogical Institute, 1980, p. 323 - 324.
82. Zagorujko et al., 2004 - Zagorujko N. and I. Borisova. The principles of natural classification. || Proceedings of the 7th International Conference on Pattern Recognition and Image Analysis. New Information Technologies. October 18 - 23, 2004. - Sankt-Peterburg: SPB Electr. Techn. University, 2004, p. 28 - 31.
83. Zajtsev G. N. Matematika v eksperimental'noj botanike. - Moskva: Nauka, 1990.
84. Zhirmunskij, 1964 - Zhirmunskij V. M. Vvedenie v sravniteljno-istoricheskoe izuchenie germanskih jazykov. [Introduction into Comparative-Historical Study of Germanic Languages]. - Leningrad