HOW SHOULD THE HUNGARIAN LANGUAGE BE CORRECTLY CLASSIFIED WITHIN THE UGRIC LANGUAGE TAXON?

Yuri TAMBOVTSEV

Department of English, Linguistics and Foreign Languages of KF,
Novosibirsk Pedagogical University,
28, Vilyuiskaya Str., Novosibirsk, 630126, Russia
yutamb@mail.ru

The aim of this article is to consider the correct classification of the set of languages called the Ugric languages. The Ugric taxon includes Hungarian and two Ob-Ugric languages: Hanty and Mansi. However, Hungarian is so different from the Ob-Ugric languages that Hanty and Mansi should be put into a separate Ob-Ugric branch[1]. Actually, that is only half of the step towards the correct classification of the Ugric languages. We must insist that Hungarian be put into a separate subgroup called the Hungarian subgroup of the Ugric group, or perhaps organize a separate group inside the Finno-Ugric family.

Key words: Ob-Ugric languages, Hungarian, Hanty, Mansi, Finno-Ugric languages, Turkic languages, Uralic languages, Manchu-Tungus languages

Majtinskaja points out that the differences between Hungarian on the one hand, and the Ob-Ugric languages on the other, is so great that she thinks it possible that the split them came earlier than 1000 B.C.[2]. Many Finno-Ugric linguists agree with Majtinskaja that the split was that long ago. There are some consequences due to the long period of the separation. Everyone who has studied Hungarian and the Ob-Ugric languages has felt that they are too different. Kalman, who knew Mansi and Hanty and could fluently speak both

---

[1] *Jazyki Mira. Ural'skie jazyki,* p. 256.

[2] Majtinskaja, Obsko-Ugorskie jazyki [The Ob-Ugric languages]. In *Jazyki narodov SSSR. T.3. Finno-Ugorskie i Samodijskie jazyki.* [Languages of the Peoples of the USSR. Finno-Ugric and Samoyedic Languages], p. 316.

languages, told me in 1985 that it is a myth that Hungarian is similar to either Hanty or Mansi. When I myself began studying Hungarian and Mansi in 1973, I was surprised to find out how different they were. Later it was proved by the methods of experimental phonetics that the spectral as well as the articulatory and combinability characteristics of their phonemes are quite different[3].

Kalman believes Hungarian to be either slightly Ugric or not Ugric at all. Therefore, he does not consider it to be correct to put Hungarian and the Ob-Ugric languages into one language taxon. He found Mansi and Hanty to be closer to the Permic or Samoyedic languages than to Hungarian[4]. Criticizing those Finno-Ugrists who support the existence of the Ugric taxon, Kalman, who was an excellent specialist in Mansi and Hanty, claims that the so-called Ugric features are true to some other Permic and Volgaic languages[5]. Using the 100 list created by Swadesh, Raun received only 34% of common words between Hungarian and Mansi and 27% between Hungarian and Hanty. Therefore, 66% of words are different in Hungarian and Mansi. Hanty is less similar with 73% of its words being different. Mansi and Hanty share 48% of their words in commonality. However, really related languages like the Permic Udmurt and Komi-Zyrian have 70% of their words in common[6]. It is possible to state that we can talk about two languages as related only when they have not less than 60% of their words in common. In the natural sciences two objects are considered similar if, and only if, they have at least 70% of their elements in common[7]. Honti states that the differences between Mansi and Hanty are greater than between the Permic or Finnic languages[8]. So, the information as discussed above represents clues for further investigations to determine if the several dialects of Mansi and Hanty should be put into the same subgroup at all.

---

[3] TAMBOVTSEV, Y. Spektral'nye harakteristiki udarnyh tsentral'nozadner'adnyh glasnyh literaturnogo mansijskogo jazyka. [Spectral characteristics of the stressed central-back vowels of the Mansi literary language]. In *Fonetika sibirskih jazykov.* [Phonetics of the Siberian languages]; Sochetanie fonem v mansijskom jazyke. [Phonemic combinations in the Mansi language]. In *Teoreticheskie voprosy fonetiki i grammatiki jazykov narodov SSSR.* [Theoretical problems of phonetics and grammar of the languages of the peoples of the USSR]; Glasnye mansijskogo jazyka po dannym rentgenografii. [Mansi vowels by the X-ray data]. In *Issledovanie zvukovyh system sibirskih jazykov.* [Investigations of the sound systems of Siberian languages].
[4] KALMAN, B. The history of the Ob-Ugric languages. In *The Uralic Languages.* (Edited by Denis Sinor), p. 396.
[5] Ibid., p. 397.
[6] RAUN, A. Ueber die sogenannte lexikostatistische Methode oder Glottochronologie und ihre Amwendung auf Finnisch-Ugrische und Tuerkische. In *Ural_Altaische Jarbucher*, 28, p. 152.
[7] TAMBOVTSEV, Y. Klassifikatsija jazykov na osnove novyh metodov opredelenija mezhjazykovyh rasstojanij (na primere jazykov samodijskoj i finno-ugorskoj semej). [Classification of languages on the basis of new methods of measuring distances between languages (on the sample of Samoyedic and Finno-Ugric families]. In *Fenno-Ugristica.*
[8] *Jazyki Mira. Ural'skie jazyki*, p. 280.

After calculating the distances between Hungarian and the Ob-Ugric languages, Taagepera came to the same conclusion as Majtinskaja and Kalman, i.e. Hungarian severed its connections with the Ob-Ugric languages earlier than the Finno-Ugric language taxon divided into the Permic and Volgaic groups[9]. Marcantonio describes in details the battle of the linguists in the late 19th century concerning two competing interpretations of the origin of the Hungarians and their language. One party supported the theory of the Turkic origin of Hungarians and the other – the Ugric origin[10].

Estimating the degrees of closeness of related languages, Jahontov insists that only close languages have to be put into one group in any linguistic classification. Unfortunately, he is quite correct to point out that usually subgroups and groups are not defined properly. It means that the degrees of closeness are not taken into consideration[11]. Strictly speaking, subgroups should include more close groups. In their own turn, groups should include more close languages than families, and so on. Therefore, every stage in classifications fixes more and more distant relatives.

Taagepera and Kuennap analysed the distances among the Uralic and other Northern Eurasian languages basing on 46 structural features. In their study, Hungarian was also the most distant from the other Finno-Ugric languages, while the Ob-Ugric languages were closer to Permic languages and Mari[12]. Their Ob-Ugric taxon was the closest, i.e. the most compact while Hungarian stands quite far away from them. They measured the distance between Hungarian and Finnic languages. This distance looks approximately the same as the distance between Hungarian and Ob-Ugric languages. It is vividly seen in their Figure 3 on page 173. Hungarian forms a separate taxon by their data. It stands apart not only from the Ob-Ugric languages, but also from Finnic and Volgaic languages. The greatest distance is between Hungarian and Samoyedic languages[13]. In fact, their investigation also speaks for constructing a new separate group for Hungarian within the Finno-Ugric family. It is interesting to see that Taagepera and Kuennap actually verified our finding that Ob-Ugric languages are typologically closer to the Permic languages. It is necessary to remark that the results of the statistical investigation of Taagepera and Kuennap

---

[9] TAAGEPERA, R. The linguistic distances between Uralic languages. In *Linguistica Uralica*, pp. 166 – 167.

[10] MARCANTONIO, A. The Uralic Language Family. Facts, Myths and Statistics, pp. 35 – 37.

[11] JAHONTOV, S. E. Otsenka stepeni blisosti rodstvennyh jazykov. [Estimation of the degree of closeness of the related languages]. In *Teoreticheskie osnovy klassifikatsii jazykov mira.* [Theoretical Fundamentals of Classification of World Languages], p. 148.

[12] TAAGEPERA, R., KUENNAP, A. Distances among Uralic and other Northern Eurasian languages. In *Linguistica Uralica*, p. 161.

[13] Ibid., p. 172.

seem quite solid and reliable since they were based on the 46 phonetic, morphological and grammatical features[14].

Taking into consideration all these doubts, it is important to reconsider the Ugric language taxon. It is necessary to understand if the Ugric language taxon is natural or artificial. It depends on the place of the Hungarian language. If Hungarian is too different from the Ob-Ugric languages (Mansi and Hanty), then the Ugric taxon is artificial. In its turn, it helps to solve the problem if Hungarian is similar enough to enter into one group with Mansi and Hanty. We will consider the Ugric taxon further, but before that we must understand which taxon is natural and which is artificial.

Natural and Artificial Taxa

Before tackling the problem of the Ugric taxon, we must understand that some language taxa are natural and some are artificial. An example of a natural set of languages is the East Slavonic language taxon, which includes Belarussian, Russian and Ukrainian. It is possible to prove it because they have similar cognates. However, they are also typologically close and the direct communication of speakers is possible. It is a fact that the native speakers of Russian can understand both Ukrainians and Belarussians, while Ukrainians can understand Belarussians. Even if one takes Eastern, Southern and Western Slavonic languages, one can say that some sort of communication is possible. It is not so obvious with some other language taxa. Communication within some other language groups is not possible. Let us take the example of the languages and dialects which compose the Ugric group of languages: Hungarian, Mansi and Hanty. Hungarians do not understand either Mansi, or Hanty. The speakers of the languages of the Ob-Ugric branch (Mansi and Hanty) of the Ugric subgroup of the Finno-Ugric group of the Uralic family usually cannot understand each other, too. Even communication between the native speakers of different dialects of Mansi is often not possible. The Konda and Sosjva (i.e. Northern) 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 must understand each other. However, it 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 it is, since they are said to be separate languages. Perhaps, in the case of the Mansi and Hanty dialects, one should speak in terms of languages, and not dialects. At the same time, Mansi, Hanty and Hungarian are united in one group of languages.

In this article we will try to construct a sort of taxonomy for different groups of languages. Taxonomy is always a sort of classification. So, we can say that classifications create taxa. Natural classifications create natural taxa and

---

[14] Ibid. pp. 178 – 179.

articicial classifications create artificial taxa. The example of an artificial taxon may be a set of languages beginning with the letter "m" in the alphabetic catalogue of 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, etc. One can see that these languages are from different genetic groups and families. Therefore, it is not a genetically-based classification. Thus, it is not natural, but artificial. Specialists in the theory of classifications usually think it quite essential to define first of all two types of classifications: natural and artificial[15]. Summing up all the points of view on the constructions of the 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 but to agree with Rozova, Kondakov, Strogovich, and other specialists in the field of theoretical classifications who analysed many classifications in many sciences and humanities. They come to the 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[16]. They point out that usually scholars try to build a natural classification because they consider natural classifications most important and "good". However, they end up with a sort of artificial classification. More often than not a natural classification is a kind of ideal. Genetic classifications are said to be natural. Rozova shows us that usually genetic classifications, which were built at the early stages of development of some science and were thought to be natural at the early stage, are not really natural. In fact, they turn out to be artificial at the later stages of the development and should be reconsidered and changed[17]. Possibly, that is the case with the Ugric group now. Perhaps the Ugric group should be analysed again and then reconsidered, or left as it is if it proves to be natural. Rozova warns against the consideration of a hypothesis as fact[18].

Sometime ago, the Finno-Ugric and Samoyedic languages were considered to be separate language families (JNSFU, 1966). However, now it is fashionable to unite them into one genetic family[19]. Though some linguists believe that the united set of Finno-Ugric and Samoyedic languages called the "Uralic family"

[15] ROZOVA, S. S. *Klassifikatsionnaja problema v sovremennoj nauke.* [Classification Problem in Modern Science], p. 45.

[16] KONDAKOV, N. I. *Logicheskij slovar'-spravochnik* [Logical Dictionary-Reference Book], p. 151; Ibid. pp. 46 – 49.

[17] ROZOVA, S. S. *Klassifikatsionnaja problema v sovremennoj nauke.* [Classification Problem in Modern Science], pp. 84 – 98.

[18] Ibid. pp. 87 – 92.

[19] AUSTERLITZ, R. Uralic languages. In *The World's Major Languages.* (Edited by Bernard Comrie), p. 569.

is a natural taxon of languages, other linguists (e.g. Kuennap, Marcantonio, Wiik, etc.) do not believe them to be a family. By a family linguists usually mean a genetically related language taxon, wich can be called natural. It it necessary to remark, of course, that it depends on how a language family is defined. One may call a family a set of languages, which are not genetically related. However, it is more correct to understand a family as a genetically related language entity, that is a set of closely related genetic languages. It is supposed to form a natural taxon. Many linguists believe that the Turkic languages form a natural taxon since these languages are very similar and direct communication is usually possible between them. Some specialists in the Finno-Ugric and Samoyedic studies are quite sceptical that all the Uralic languages, especially Finno-Ugric and Samoyedic, are genetically related. That is they do not believe the Uralic taxon of languages to be a natural language taxon. The demonstration of a genetic relationship depends on finding words of similar phonological shape that have equivalent meaning, called cognates. That means that if languages are related, their speech sound chains are similar.

Usually the languages of the world are classified into some taxa 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 taxa by a new method called typologo-metrical. Here, we shall touch upon the taxon of Uralic languages. The taxon of the Uralic languages is known to include the Finno-Ugric and the Samoyedic languages. We should analyse the typological similarity of the sound chains of the Finno-Ugric and the Samoyedic taxa to find out if they are similar enough to belong to one and the same language family. If they are not similar, then one should come to the conclusion that their combination into one language family is artificial.

Let us consider one point, wich may be the same for natural and artificial classifications: it is the usefulness of these classifications. Sometimes this point, especially in the initial stages of the development of some sciences or humanities, leads the scholars astray. Thus, a useful artificial classification may be taken for a natural classification. Actually, even though it may sound strange, both natural and artificial classifications are quite useful. A list of the Finno-Ugric, Samoyedic, Turkic or other world languages in alphabetic order is a fair example of a useful classification, which is at the same time is artificial. The order of the languages in these classifications and thus, the neighbouring languages, 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, for the purposes of organising the Turkic languages we used the 'alphabetic order' principle because there are at least fifteen classifications of

Turkic languages which may be called natural since they take into account some important and essential typological-genetic features. At the same time, the artificial language classifications select arbitrary features, which are not important or essential for this or that set of languages[20]. In this case, an artificial classification is more correct because a natural classification may be misleading. Gadzhieva does not believe it is possible to yield one classification of the Turkic languages, which should be true of all the aspects. On the contrary, she emphasizes 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[21].

We shall study different language taxa on the basis of the new method in linguistics, i.e. by examining the degrees of compactness of the main phonetic features. Firstly, let us discuss the notion of compactness and how to measure the degree of compactness of the different language taxa. 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 certain important and essential articulatory features. Several criteria of mathematical statistics were used to measure the compactness[22].

Establishing a strict Hierarchy of Language Taxa

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

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

[20] TAMBOVTSEV, Y. Is Kumandin a Turkic language? In *Dilbilim Arashtirmalari* (Istanbul).
[21] GADZHIEVA, N. Z. K voprosu o klassifikatsii t'urkskih jazykov i dialektov. In *Teoreticheskie osnovy klassifikatsii jazykov mira*, p. 125.
[22] TAMBOVTSEV, Y. *Konsonantnyj koeffitsijent v jazykah raznyh semej.* [The Consonant Coefficient in the Languages of Different Language Families].

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 receive a sort of taxonomy, and in linguistics that can be understood as a sort of classification. 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 concept of compactness and this is stated by various scholars who deal with taxonomy[23].

One has to define a set of languages as a branch, i.e. the smallest language taxon. One of the options is to define the 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. Thus, 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. So, one can see that the definitions of language taxa 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 a general term like "taxon". We propose by a language taxon to mean some sort of a set of languages. Actually, by our typological-metrical method we try to construct a kind of typologo-metric classification for the Finno-Ugric and Samoyedic languages known as the 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 convenience to classify them. Thus, in this case, one would call it an artificial classification. If they are sufficiently close from the phono-typological point of view, then they should be called a natural classification. It is highly probable that a natural classification is apt to be a genetic one. After calculating the Uralic compactness on the one hand, and Finno-Ugric and Samoyedic compactness on the other, one can draw certain conclusions. So, we can receive the values of compactness for these taxa: 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, Manchu-Tungus and other taxa of the world languages. Thus, we are

---

[23] ZAGORUJKO, N. G. *Metody raspoznavanija i ih primenenie*. [Methods of Recognition and their Application], p. 90.

trying to build up some new sort of system of organization for the Finno-Ugric, Samoyedic and other languages defined in accordance with their presumed or natural relationships based on a certain set of the selected features.

Recent Developments in Uralistics

It is obvious that not only the Ugric taxon should be reconsidered. It seems nowadays that the developments in Uralistics are a sort of crisis involving a scientific paradigm 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 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[24]. Kuhn is quite correct to stress that the old scientific paradigm never goes away peacefully. Usually, the scholars strongly and negatively react to new theories and those scholars who introduce them. Kuhn points out that what the scholars never do: it is to rush to the support of the new theory[25]. We can see the similar negative reaction of the majority of the specialists in Uralistics to the new theories of Kuennap, Marcantonio, 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 to 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 the inconsistencies and drawbacks of it. 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 are working on some other linguistic problems, which do not concern the classification of languages. Probably, 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 other linguists who think that with the 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 there are more and more new linguistic facts, which contradict the old classification, then it has to be reconsidered on the basis of the new level of linguistic knowledge. The linguists with modern linguistic thinking argue that the old linguistics classifications must be verified and checked again and again,

---

[24] KUHN, T. *Struktura nauchnyh revoll'utsij.* [Kuhn T. S. The structure of scientific revolutions], pp. 96 – 109.

[25] Ibid., pp. 110 – 119.

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 of development of linguistics, may be neither good, nor logical at the advanced level of 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 an abundance of new linguistic facts are discovered.

Some outstanding linguists such as Serebrennikov urged linguists to return to the established language taxa (classifications) again in order to verify them on the basis of certain laws of logic. He stressed that each established genetical language family (i.e. a language taxon or a classification) is not a fact but a hypothesis[26].

Compactness of Language Taxa

We built our definitions and ideas about compactness on the fundamentals of pattern recognition 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 linguistic 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, though we omitted mathematical formalism. We understand "compact" as "neatly fitted, firmly put together, closely united or packed, not gangling, 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 concept of compactness in philosophy, science, technology and everyday life[27]. In linguistics it was not used in the way we use it. It seems that we introduced it 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

[26] SEREBRENNOKOV, B. A. Problema dostatochnosti osnovanij v gipotezah, kasajushchihs'a geneticheskogo rodstva jazykov. [The problem of substantiated foundations in the hypothesis concerning the genetic relatedness of languages]. In *Teoreticheskie osnovy klassifikatsii jazykov mira. Problemy rodstva*, p. 6.

[27] *Entsyklopedija kibernetiki.* [Encyclopedia of Cybernetics], p. 457; HORNBY, A. S. (with assistance of Christina Ruse). *Oxford student's dictionary of current English*, p. 115; KONDAKOV, N. I. *Logicheskij slovar'-spravochnik* [Logical Dictionary-Reference Book], p. 254; OZHEGOV, S. I. *Slovar' russkogo jazyka.* [Dictionary of the Russian language], p. 280; Petrova, 1964, p. 314; *Webster's seventh new collegiate dictionary*, p. 168.

used in the acoustical sense, as one of the features of a vowel or consonant sound. It was invented and 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 idea in acuostic phonetics and is generally accepted[28].

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 taxa: 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. The taxonomic approach was taken from biosystems and pattern recognition. This approach requires us to select certain linguistic features. We have selected the frequency of the 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 Martynov, Gamkrelidze and Ivanov. Unfortunately they do not 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, Martynov speaks about "a compact language massive", meaning that the native speakers of a certain language live together[29]. Gamkrelidze and Ivanov speak about "the compactness of territorial language spread"[30].

Sometimes this term is used in the sense close to ours in archaeology. So, Tret'jakov speaks of "compact heaps of ancient relics"[31], the "compactness of the territories of the peoples"[32], or "compact settling"[33].

---

[28] CRYSTAL, D. An Encyclopaedic Dictionary of Language and Languages, p. 72.

[29] 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], p. 6

[30] GAMKRELIDZE, T.V., IVANOV, V. *Indoevropejskij jazyk i Indoevropejtsy.* [The Indo-European Language and the Indo-Europeans], p. 44.

[31] TRET'JAKOV, P. N. *U istokov drevnerusskoj narodnosti.* [At the Source of the Old Russian people], p. 81.

It is necessary to emphasize that in our work we used a more precise definition of compactness[34]. It was possible to receive a number of 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[35].

Some scholars speak about "condensation", which is in fact the same as compactness[36]. Apresjan uses the term "condensation" in linguistics. Basing on condensations his algorithm constructs classes of verbs in Russian[37]. We consider that in this case both of the terms "compactness" and "condensation" mean the same.

Let us consider in detail what is compactness or condensation. Mirkin considers condensation to be such a set of objects in which any inner link is shorter than any outer link[38]. Further we shall see that the definition of compactness is more or less the same in the opinion of different scholars.

The first scholars who defined compactness were Arkadjev and 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 rather vast neighbourhood[39].
    They developed the idea of compactness in their subsequent works. Actually, Arkadjev and Braverman remark that 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[40].

---

[32] Ibid., p. 3.

[33] TRET'JAKOV, P. N. *Po sledam drevnih slavjanskih plemen.* [Along the Traces of the Ancient Slavonic Tribes], p. 118.

[34] TAMBOVTSEV, Y. *Konsonantnyj koeffitsijent v jazykah raznyh semej.* [The Consonant Coefficient in the Languages of Different Language Families].

[35] Arkadjev et al., 1964, pp. 29-34; Bongard, 1967, pp. 30-31; Eliseeva et al., 1977, pp. 9-14; Mirkin, 1976, pp. 114-116; Vasil'jev, 1969, pp. 16-18; Zagorujko, 1972, pp. 21.

[36] MIRKIN, B.G. *Analiz kachestvennyh priznakov (matematicheskije metody i modeli).*[Analysis of quantitative features (mathematical methods and models), pp. 114-146.

[37] APRESJAN, Y. D. Algoritm postroenija klassov po matritse rasstojanij. [Algorithm of Constructing Classes by the Matrix of Distances]. In *Mashinnyj perevod i prikladnaja lingvistika,* pp. 141-144.

[38] MIRKIN, B.G. *Analiz kachestvennyh priznakov (matematicheskije metody i modeli).*[Analysis of quantitative features (mathematical methods and models), p. 235.

[39]ARKADJEV, A. G., BRAVERMAN E. M. *Obuchenije mashiny raspoznavaniju obrazov.* [Teaching a Machine Pattern Recognition], p. 28.

[40] ARKADJEV, A. G., BRAVERMAN, E. M. *Obuchenije mashiny klassifikatsii objectov.*

Another pattern recognizer, 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[41]. However, further Vasiljev cites all the requirements of Arkadjev and Braverman, which we discussed above. Therefore, they are not provided here.

Bongard remarks that if the set of points is compact, then they are all situated in a space closely[42].

Zagorujko believes that one often operates with the notion of compactness in pattern recognition, by which one means that the realization of one and the same image is reflected in geometrically close points of the sample space[43].

Eliseeva and Rukavishnikov speak about compactness and condensation as about the places where the points concentrate[44].

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 recognizers agree to the definition of compactness given by Glushkov and his co-authors in the *Encyclopaedia of Cybernetics*, who regard the hypothesis of compactness as a supposition that the subset of the images to be recognized is simple in a certain sense. The notion of simplicity may be realized 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[45]. 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

[Teching a Machine Object Classification], p. 27.

[41] VASILJEV, V.I. *Raspoznajushchije sistemy.* [Recognition Systems], pp. 16-17.

[42] BONGARD, M.M. *Problema uznavanija.* [The Problem of Recognition], pp. 30-31.

[43] ZAGORUJKO, N.G. *Metody raspoznavanija i ih primenenie.* [Methods of Recognition and their Application], p.21.

[44] ELISEEVA, I. I. et al. *Gruppirovka, korreljatsija, raspoznavanie obrazov.* (Statisticheskie metody, klassifikatsija i izmerenie svjazej). [Grouping, Correlation, Pattern Recognition. (Statistical Methods, Classification and Connection)], p. 40.

[45] *Entsyklopedija kibernetiki.* [Encyclopedia of Cybernetics], p. 229.

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 analyse this object and define some features on the basis of which we shall form taxa. The features must be basic. Kordonskij states that there are two forms of theoretical descriptions: a classificatory description introduces the notion of a taxon, while the experimental description introduces the notion of a type. Taxa fix the objects which exist. Types fix functioning of objects, i.e. the way the objects exist[46]. Languages may be regarded as objects and as sets of functions. This is why languages may be both taxa and types.

Shreider understands languages as inner systems, which can be classified. Therefore he understands the schemes of language classifications as the outer systems. In fact, by the outer system he means a class of objects, which have some common features. Moreover, he thinks that these objects may be united because – and only because – of the fact that they form a natural system. Developing his ideas further, 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[47]. It is quite important for linguists to understand whether existing language taxa are real (i.e. natural) or imaginary (i.e. artificial), 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 taxa 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 of the world languages. As 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 homogeneous property of the quality[48]. 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 the chi-square criterion is greater than its theoretical value. One can see that this certain language does not belong to this

---

[46] KORDONSKIJ, S. Tsikly dejatel'nosti i ideal'nyje objekty, p. 19.
[47] KORDONSKIJ, S. Taksony i analiticheskie objekty kak sistemy. [Taxa and analytical objects as systems]. In *Sistemnyj metod i sovremennaja nauka*, p. 141.
[48] MELJNIKOV, G. P. *Sistemologija i jazykovye aspekty kibernetiki.* [Systemology and the Language Aspects of Cybernetics], p. 22.

certain taxon[49]. Mel'nikov calls it the borders of the allowed diffusion of the functional states. Transferring his remark onto a language taxon as a holistic object, one can speak about the limits of the changes under which the given object remains in the limits of the stable functioning[50].

In other words, the stability of a taxon as a holistic object aims at the influence on those 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 not a group, but a mechanical mixture, i.e. a 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 as an indicator of the measure of how systemic this or that group is, if this group is a holistic object or not. In the opinion of Mel'nikov, any object may be regarded as a system, if it is defined how systemic it is[51]. Mel'nikov underlined the importance of treating the languages of the world as a kind of system, which should be studied by the methods of mathematical statistics[52].

As 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 it is called the place of articulation;
b) manner of articulation: sonorant, occlusive and fricative;
c) work of the vocal cords: voiced.

It is the most complete consonantal classification from the point of view of articulation. This is why we can call these features the most basic and essential for any human language. Therefore, we obtained 8 features for consonants: 1) labial; 2) front; 3) palatal; 4) velar; 5) sonorant; 6) occlusive; 7) fricative; 8) voiced. They have equal typological weights.

The frequency of occurrence of the consonantal groups mentioned above is calculated in percent to all the phonemes in the speech sound chain. This is

---

[49] TAMBOVTSEV, Y. *Dinamika funktsionorovanija fonem v zvukovyh tsepochkah jazykov razlichnogo stroja.* [Dynamics of Functioning of Phonemes in the Sound Chains of the Languages of Different Structure], pp. 23-69.

[50] MELJNIKOV, G. P. *Sistemologija i jazykovye aspekty kibernetiki.* [Systemology and the Language Aspects of Cybernetics], p. 55.

[51] Ibid., p. 68.

[52] MEL'NIKOV, G. P. *Sistemnaja tipologija jazykov: Printsypy, metody, modeli.* [Systemic Typology of Languages: Principles, Methods, Models], pp. 278 - 281; 347, etc.

why, the frequency of vowels is tacitly present, though in this work it is not explicitly used. If it is not possible to find out the differences in the compactness of language taxa based only on consonants, then the actual frequencies of vowels should be introduced.

The standard for comparing typological similarities of speech sound chains of different languages is the value of compactness of different language taxa. 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 did it before. We have developed several 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 cannot use it for the reasons of commensurability. Nevertheless, one can use standard deviation if one studies the variability of a certain 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 emphasize that it is quite necessary to keep to the principle of commensurability. Usually it is not discussed in linguistics. Nevertheless, it is one of the most basic principles in any scientific investigation[53].

In comparing objects (in this case – languages) one should keep to one and the same principle and the comparison must use universal parameters[54]. We believe compactness to be a universal parameter since we can measure it on the taxa of languages of different genetic origin and morphologic structure.

To measure compactness one should first measure its inverse, i.e. diffusion 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 the T-coefficient, let us consider the calculation of

[53] DRUJANOV, L. A. *Zakony prirody i ih poznanije*, pp. 28 - 52; PETROV, V. V. *Kriticheskij analiz problemy nesoizmerimosti v sovremennoj "filosofii nauki".* [Critical analysis of the problem of incommensurability in modern "philosophy of sciences"], p. 6; ZAGORUJKO, N. G. et al. Statisticheskoe raspredelenie udarenija v russkom slove. [Statistical Distribution of Stress in Russian Words]. In Materialy seminara *"Statisticheskaja optimizatsija prepodavanija jazykov i inzhenernaja lingvistika."*, pp. 32 – 34.
[54] KONDAKOV, N. I. *Logicheskij slovar'-spravochnik* [Logical Dictionary-Reference Book], p. 151.

the standard deviation, which is the basis for calculation of the coefficient of variance.

Very often linguists compare means of occurrence of some linguistic units and do not go further than that. Actually, the mean value is more reliable than the mode or median. One should take the mean because it represents the distribution better than the mode or median[55]. Nevertheless, to see the variance a linguist should use the standard deviation because it is the measure of dispersion and changeability. 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[56]. Standard deviation measures how the values of a variable are grouped around its 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 are the values around its mean. The greater the dispersion, the greater the changeability of some linguistic phenomenon. And on the contrary, if the dispersion becomes less, then the changeability is less.

Let us discuss the properties of standard deviation in detail. Compactness is really closely connected to variability. 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 another question before considering measures of variability in some detail: why do we bother to measure this property at all? We agree with him that variability may be of interest in its own right[57]. In this case, it is important because we are attempting 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 ranges, nor the means of some linguistic values is evident enough while comparing two language taxa. 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 in a simple case with only one dimension, though our conclusions are true for multi-dimensional cases, too. Theoretically, one can imagine several cases.

It is easy to see that if two taxa 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:

[55] PAVLOVSKIJ, Z. *Vvedenije v matematicheskuju statistiku*, p. 55.

[56] BUTLER, Ch. *Statistics in Linguistics*, p. 37.

[57] Ibid., p. 35.

A) [_[_[_[_[_[_[_[. B) [[[[_____[[[[

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

Practically, in linguistics one encounters this situation when the number of the members in the taxa 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 a taxon has the same number of members as the other one, then its compactness is less, if its range is greater:

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

On the contrary, if the range in both taxa 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 linguistical statistics[58].

Let us consider the properties of the coefficient of variation. As 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 less) than the standard deviation 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 the Indo-European family is 1.14, while the value of the standard deviation of the front consonants is 5.38. We shall see further that in fact the values of the standard deviations of labial and front consonants are incommensurable, that is, they cannot be compared directly. There is no basis for comparison since the mean of the labial consonants in the Ugric languages is 11.79, while the mean of the Ugric front consonants is 3 times greater (31.18). The same incommensurability one can observe in other language taxons. For instance, the labial mean in the 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

---

[58] Ibid., p. 37.

reliable relative measure[59]:

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

where V– the coefficient of variation

    S – the standard deviation

    M – the mean value

In fact, Butler, Piotrovskij, myself and others showed 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[60].

If we reconsider the compactness of the labial consonants on the basis of the coefficient of varience, then we can see that it is more similar to the actual facts. The coefficient of variance of the labial consonants in the Ugric taxon is 12.65%, while that of the front consonants is 6.09% (that is 2 times less). Their standard deviations are more or less the same at 1.49 and 1.90. Let us take another example. The coefficient of the variance of labial (V=10.01%) and front (14.25%) consonants of the Germanic group, which shows that the compactness is more or less the same (cf. 10.01% and 14.25%), though their standard deviations are quite different (cf. 1.14 and 5.38). One can see that the 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 taxa 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%). The coefficient of variation allows us to compare the labial campactness of the groups of the Indo-European family mentioned above with those of other families: Mongolian (V=7.55%), Manchu-Tungus (16.19%) or American Indian languages (29.89%).

The value of the coefficient of variation as well as the 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. In other words, stability is a property allowing for some distortion, not great, and within certain limits.

Actually, there arises a question in regard to how great these allowed limits should be. It is easier to find out such limits for the T – coefficient, as we shall see further. Now, it should be explained how it is possible to calculate the TMB

---

[59] RICHMOND, S. B. *Statistical analysis*, pp. 89 – 90.

[60] BUTLER, Ch. *Statistics in Linguistics*, pp. 37-43.

coefficient. We do it with the help of the theoretical values for the "chi-square" criterion, given in statistical tables. In order to calculate the T-coefficient, it is necessary to divide the obtained chi-square value by its table value taking into consideration the degree of freedom. If the value of the TMB coefficient is less than 1 (a unit), then the set under investigation should be considered homogeneous. If it is equal to 1 (a unit) or greater, then the set is not homogeneous. It is necessary to point out that we can measure different degrees of homogeneity or dispersion by the value of the TMB coefficient.

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 empirically, i.e. from practice. Unfortunately, different scholars allow for different limits. Some of them consider it tolerable to have its value at a range up to 50%[61]. Sis'kov is sure that the taxon is homogeneous and stable if the value of coefficient of variation is less than 33%[62]. 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%[63]. Piotrovskij calculates the stability of the use of the English definite article (V=7.12%) and the German word "power" in the texts of the former Federal Republic of Germany (V= 90.00%) and the former German Democratic Republic (160.00%): he thinks that the distribution of the article may be called stable.

Martynenko correctly points out that the greater the value of the coefficient of variation, the greater the probability that the set is not homogeneous. He calls it the criterion of fluctuation[64]. Sis'kov believes that the high value of this coefficient may indicate the mixture of two distributions with quite different means[65]. Venchikov and his colleagues think that the value of this coefficient should not be greater than 50%[66]. Zajtsev considers critical a value that is 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 large variation: 85% - 104%; 6) abnormal variation: 105% and more. In the limits of the normal variation he

---

[61] MARTYNENKO, G. J. *Osnovy stilemetrii*, p. 62.

[62] SIS'KOV, V. I. Ob opredelenii velichiny intervala pri gruppirovkah, p. 10.

[63] MARTYNENKO, G. J. *Osnovy stilemetrii*, p. 150 – 154.

[64] Ibid., p. 62.

[65] SIS'KOV V. I. *Korrel'atsionnyj analiz v ekononicheskih issledovanijah*, p. 101.

[66] VENCHIKOV et al. *Osnovnye prijomy statisticheskoj obrabotki rezul'tatov nabl'udenij v oblasti fiziologii,* p. 21.

defines the so-called lower norm: 5% - 24%[67]. 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. We will consider the fluctuation of the acoustic duration of speech sounds. The value of the coefficient of variation of the voiced part of English consonants is 26.56% - 27.49% and the tempo of speech 7.69% - 16.04%. The duration of Russian vowels is 14.97% - 27.83%. The value of the coefficient of variation of Kumandin vowel duration is in the limits of 5.49% - 18.04%, and in Ket the limit is 20.51% - 34.97%.

Compactness of the Ugric Taxon in Comparison to different Language Taxa

Further we shall use the value of compactness to judge if a taxonomy or cauterization is made correctly. It is shown in detail in Tables (cf. Tab 1 - 3). We will calculate the compactness of a language taxon, then combine it with another taxon and measure if the vaue of compactness becomes greater or less. We can notice that the value of the compactness of an artificial taxon which consisits of two or more groups is less than that of a natural taxon. 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 homogeneous a taxon is. One can measure it with the help of the value of the coefficient of variation or the TMB-coefficient. If the value of the coefficient of variation or the TMB-coefficient becomes smaller, then the language, which is introduced into 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 similar a language is 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 used 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 show the sound similarities of the genetically related languages. The Eastern subgroup of the Slavonic group of the Indo-European language family shows very close similarity:

	mother	brother	sister	children	grand-child
Russian	mat'	brat	s'estra	d'et'i	vnuk
Ukrainian	mati	brat	sestra	d'iti	vnuk

---

[67] ZAJTSEV, G. N. *Matematika v eksperimental'noj botanike.* [Mathematics in Experimental Botany], p. 39.

Belarussian matsi   brat   s'astra   dzetsi   unuk

We can take the compactness of the Eastern subgroup of the Slavonic taxon as a fair standard for the language compactness. Its coefficient of variance is 7.72% and the TMB = 0.03. One can see that these three languages are really very similar. Further, we'll compare these values to the values of the other language taxa.

Approximately the same similar sound picture we find in other genetically related languages. Genetical 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 the contact which they have over a long duration of time. This is why, it is important to trace the contact between languages. We shall compare the values of compactness of the groups in the language families where groups are well-defined (e.g., Finno-Ugric and Indo-European).

Compactness of the Ugric Language Taxon

The compactness of the Ugric language taxon allows us to find the answer to the question of how natural the Ugric language taxon is. In other words, the main aim of this article is to consider the set of languages called the Ugric languages. According to the modern state of the development of Finno-Ugric studies, the Ugric subgroup of the Finno-Ugric language family 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 the Turkic languages, because during its long history of development Hungarian had a lot of close contact with Turkic languages. Abondolo, Kalman and other linguists underline that Hungarian is not typical among the Uralic family[68].

We take for our analysis the five Ugric languages and dialects: the Northern (Sos'va) dialect of Mansi, the Konda dialect of Mansi, the Northern (Kazym) dialect of Hanty, the Eastern dialect of Hanty, and Hungarian (Tab.1).

Tab. 1
The Frequency of Occurrence of the Groups of Consonants and the other statistical Characteristics of the 5 Ugric Languages and Dialects. Mean V%=27.66%; Mean TMB= 0.47.

---

[68] ABONDOLO, D. Hungarian. In *The World's Major Languages* (Edited by Bernard Comrie), p. 577.

Conson. Lang.	Labial	Front	Palat	Back	Sonor	Occlu	Fricat	Voice
Mansi North.	13.56	30.08	6.79	10.64	32.03	17.00	12.04	2.74
Mansi Konda	12.29	29.72	12.30	8.46	30.07	16.56	16.15	4.50
Hanty North.	12.60	30.83	7.60	8.61	30.97	17.19	11.48	0.00
Hanty East.	10.45	30.81	5.19	13.53	21.83	24.20	13.95	10.51
Hungarian	10.04	34.47	4.07	9.44	22.53	22.62	12.87	12.70
Mean	11.79	31.18	7.19	10.14	27.48	19.51	13.30	6.09
S	1.49	1.90	3.17	2.09	4.90	3.61	1.85	5.34
SI	2.22	3.61	10.05	4.35	24.01	13.03	3.41	28.52
V %	12.65	6.09	44.08	20.57	17.82	18.49	13.88	87.69
TMB	0.08	0.05	0.59	0.18	0.37	0.42	0.11	1.97
Chi-sq	0.75	0.46	5.59	1.72	3.49	4.00	1.03	18.73
1\2 CI	1.42	1.81	3.02	1.99	4.67	3.44	1.76	5.09

The data of Tab. 1 show that the value of the coefficient of variation of the labial consonants is 12.65% and TMB=0.08. Therefore, the front consonants of the five Ugric languages are dispersed less because their values are 6.09% and 0.05 respectively. The most dispersed are the palatal consonants: the MV=44.08% and the MTMB=0.59.

We can compare every group of the consonants of the five Ugric languages in this way not only to each other, but also to the other groups of languages. The labial consonants are dispersed in more or less the same way in the group of seven Finnic languages (c.f. MV= 12.41% and MTMB= 0.08). However, they are more dispersed in the four Volgaic (MV=17.10% and MTMB=0.12) languages. It takes much time and effort to compare each group of consonants. It is possible to find the cumulative value for all the groups. We should take the mean of the values in order to be comparable if the number of features (groups of consonants) becomes different (less or more). So, the mean value for the coefficient of variation is 27.66% and the mean TMB=0.47. They are greater than those of the Volgaic (V=17.90%; TMB=0.13) or Finnic (23.24%; TMB=0.35) groups. One must admit that the most compact is the Permic group (MV=11.65%; MTMB=0.07). In this group, Komi-Zyrian is very close to Komi-Permian (MV=3.16%; MTMB=0.01).

What can it mean? It can mean only one thing, i.e. the Ugric taxon is more dispersed and looks more like an artificial taxon. However, for this conclusion, we must consider Hungarian as a member of another language taxa. Let us, consider the point of view of those linguists who claim that Hungarian must be

considered a Turkic language. To prove it, we must introduce Hungarian into the taxon of the Turkic languages.

At first sight is 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 Hungarians during their history it is not so striking. It may be because of the ancient contacts of the Hungarians with Turkic peoples in Siberia and for the period of time when the Hungarians lived for more than one thousand years in the vicinity of the Volga river. Speech communication between Hungarians and the Ob-Ugrics, i.e. Mansi (Vogul) and Hanty (Ostjak), is not possible. We must point out 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 Marcantonio came to the conclusion that Hungarian is not a Uralic language. She writes about the history of attribution of Hungarian by linguists 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 that only 19% of the Budenz core lexicon is with any certainty Finno-Ugric[69].

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 more dispersed than homogeneous. 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 the different subgroups of the Finno-Ugric languages. Let us mix the Ugric and Permic taxa. 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 distribution of consonants in the Ugric and in the Permic languages. One can state that the Ugric and the Permic languages are typologically closer to each other than to the other Finno-Ugric languages. Austerlitz seems to have no solid foundation to put the Permic and the Volgaic languages into one taxon with the Baltic-Finnic languages[70]. Our data show quite vividly that the Ugric languages are closer to the Permic languages (cf. Table 2, 3). The united Ugro-Permic group is more compact (c.f. MV= 23.99%; MTMB= 0.47). Our statistical analysis confirms the conclusions of Budenz, Zsirai, Moor, Haidu,

---

[69] MARCANTONIO, A. The Uralic Language Family. Facts, Myths and Statistics, pp. 37 – 48.
[70] AUSTERLITZ, R. Uralic languages. In *The World's Major Languages*. (Edited by Bernard Comrie), p. 570.

Redei and Helimskij who believe that Finnic severed itself from the Ugro-Permic group rather than the converse.

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

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). This means that the 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 therein of 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), and 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 demonstrated with the Finno-Ugric languages and Sweet Grass Cree, an American Indian language of Canada (MV= 26.62%, MTMB= 0.56).

We agree with those linguists who think that the foundations of the language classifications should undergo a closer attention and stricter verification[71]. In fact, Ku"nnap and 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[72]. One can hope that our phonostatistical typological data may provide a large quantity of new material to reconsider different language taxa. It may allow linguists to verify some language taxa and to reject others.

It is possible to make the following conclusions:

1. The sound chains of the Ugric languages show that the Ugric taxon is not natural from the typological point of view, but rather artificial, i.e. created by linguists. It is high time to reconsider the place of Hungarian in the Ugric taxon.

2. The taxon of the Permic languages seems quite natural.

3. The Volgaic and Finnic taxa are less compact than the Permic taxon, but more compact than the Ugric taxon.

---

[71] SHARADZENIDZE, T. S. Rodstvo jazykov, protsessy divergentsii, konvergentsii i sootvetstvujushchie im klassifikatsii jazykov mira. [Genetical 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*, p. 71.

[72] MARCANTONIO, A. The Uralic Language Family. Facts, Myths and Statistics.

4.Labial, front, palatal, back, sonorant, occlusive, fricative and voiced consonants have rather strict limits of occurrence in the Finno-Ugric languages.

5.The group of front consonants is used in different language taxa more compactly than the other seven consonantal groups. The next compact group is occlusive consonants.

6.The frequency of occurrence of the voiced consonants is the most changeable and unstable feature among the eight chosen features. It may be absent in some languages.

7.Usually, a language taxon is compact if its languages are typologically and genetically close.

8.Our typologo-metrical approach at the phonological level shows which taxon is natural and which is not natural. The artificial taxon may be a mechanical conglomeration of different languages of different origins, put in one set by linguists for classification purposes. However, any artificial language taxon is quite useful at the early stage of investigation. Our phonological data and the phonostatistical method may give the linguists another impetus to reconsider the suspicious — i.e. disperse – language taxa.

Tab. 2

The Frequency of Occurrence of the Groups of Consonants and the other statistical Characteristics of the Permic Languages. The Compactness of the Permic Language Taxon by Eight Features. The Permic Subgroup of the Finno-Ugric Language Family. Mean coefficient of variation (V%) = 11.65. Mean TMB=0.01.

Consonant Language	Labia %	Front %	Palat %	Back %	Sonor %	Occlu %	Fricat %	Voice %
Komi-Zyrian	10.21	32.94	9.59	5.94	21.83	20.65	16.20	13.05
Komi-Permian	11.15	31.52	9.23	6.34	20.79	20.33	17.12	13.09
Mean	10.68	32.23	9.41	6.14	21.31	20.49	16.66	13.07
S	0.66	1.00	0.25	0.28	0.73	0.23	0.65	0.03
SI	0.44	1.00	0.06	0.08	0.53	0.05	0.42	0.00
V%	6.18	3.10	2.66	4.56	3.43	1.12	3.90	0.30
TMB	0.01	0.01	0.00	0.00	0.01	0.00	0.01	0.01
Chi-sq	0.04	0.03	0.01	0.01	0.02	0.00	0.02	0.02
1\2CI	8.39	12.71	3.18	3.56	9.28	2.92	8.26	0.38
Udmurt	13.66	29.47	6.94	8.71	25.10	21.98	11.70	12.90
Mean	11.67	31.31	8.59	7.00	22.57	20.99	15.01	13.01
S	1,78	1,74	1,44	1,50	2,25	0,88	2,89	0,10
SI	3,17	3,03	2,07	2,25	5,06	0,77	8,35	0,01
V%	15,25	5,56	16,76	21,43	9,97	4,19	19,25	0,77
TMB	0,09	0,03	0,08	0,11	0,08	0,01	0,19	0,00
Chi-square	0.54	0.19	0.48	0.64	0.45	0.07	1.11	0.00
1\2 CI	5.41	5.29	4.38	4.56	6.84	2.68	8.79	0.30

Tab. 3
The Frequency of Occurrence of the Groups of Consonants and the other statistical Characteristics of Five Ugric Languages and Komi-Zyrian. The Compactness of the Ugric and Komi-Zyrian Language Taxon by Eight Features. Mean coefficient of variation (V%) = 13.46. Mean TMB=0.46.

N	Consonant Language	Labia %	Front %	Palat %	Back %	Sonor %	Occlu %	Fricat %	Voice %
1	Mansi Sos	13.56	30.08	06.79	10.64	32.03	17.00	12.04	02.74
2	Mansi Kon	12.29	29.72	12.30	08.46	30.07	16.56	16.15	04.50
3	Hanty Kaz	12.60	30.83	07.60	08.61	30.96	17.19	11.48	00.00
4	Hanty East	10.45	30.81	05.19	13.53	21.83	24.20	13.95	10.51
5	Hungarian	10.04	34.47	04.07	09.44	22.53	22.62	12.87	12.70
m	Mean	11.79	31.18	07.19	10.14	27.48	19.51	13.30	06.09
s	Standard	01.49	01.90	03.17	02.09	04.90	03.61	01.85	05.34
sI	S-squared	02.22	03.61	10.05	04.35	24.01	13.03	03.41	28.52
V%	Coeff. Var.	12.65	06.09	44.08	20.57	17.82	18.49	13.86	87.69
χ^2	Chi-square	00.75	00.46	05.59	01.72	03.49	04.00	01.03	18.73
TMB	TMB coeff.	00.08	00.05	00.59	00.18	00.37	00.42	00.11	01.97
Hlf CI	Half Confid. Interval	01.42	01.81	03.02	01.99	04.67	03.44	01.76	05.09
	Ugric + Komi-Zyrian								
	Consonant groups	Labia	Front	Palat.	Back	Sonor	Occlu	Fricat	Voice
	Komi-Zyrian	10.21	32.94	09.59	05.94	21.83	20.65	16.20	13.05
m	Mean	11.53	31.48	07.59	09.44	26.54	19.70	13.78	07.25
s	Standard	01.48	01.84	03.00	02.54	04.95	03.26	02.03	05.56
sI	S-squared	02.19	03.40	09.00	06.42	24.50	10.63	04.13	30.91
V%	Coeff. Var.	12.85	05.86	39.51	26.84	18.66	16.54	14.75	76.66
χ^2	Chi-square	00.95	00.54	05.93	03.40	04.62	02.70	01.50	21.32
TMB	TMB coeff.	00.09	00.05	00.54	00.31	00.42	00.24	00.14	01.93
Hlf CI	Half Confid. Interval	01.22	00.95	05.02	02.05	01.65	00.54	00.43	03.82

Tab. 4
The Frequency of Occurrence of the Groups of Consonants and the other statistical Characteristics of Five Ugric Languages and Three Permic languages. The Compactness of the Mixed Ugric and Permic Taxon by Eight Features. Mean coefficient of variation (V%) = 23.99. Mean TMB=0.47.

N	Consonant Language	Labia %	Front %	Palat %	Back %	Sonor %	Occlu %	Fricat %	Voice %
1	Mansi Sos	13.56	30.08	06.79	10.64	32.03	17.00	12.04	02.74
2	Mansi Kon	12.29	29.72	12.30	08.46	30.07	16.56	16.15	04.50
3	Hanty Kaz	12.60	30.83	07.60	08.61	30.96	17.19	11.48	00.00
4	Hanty East	10.45	30.81	05.19	13.53	21.83	24.20	13.95	10.51

N	Language								
5	Hungarian	10.04	34.47	04.07	09.44	22.53	22.62	12.87	12.70
6	Komi-Zyrian	10.21	32.94	9.59	5.94	21.83	20.65	16.20	13.05
7	Komi-Permi	11.15	31.52	9.23	6.34	20.79	20.33	17.12	13.09
8	Udmurt	13.66	29.47	6.94	8.71	25.10	21.98	11.70	12.90
m	Mean	11.99	30.77	8.23	8.89	26.09	19.70	14.09	8.11
s	Standard	1.41	1.19	2.34	2.59	4.83	2.89	2.40	5.56
sI	S-squared	1.99	1.47	5.47	6.68	23.33	8.37	5.77	30.91
V%	Coeff. Var.	11.78	3.88	28.39	29.08	18.53	14.68	17.05	68.55
χ^2	Chi-square	0.99	0.28	3.98	4.51	5.37	2.55	2.46	26.68
TMB	TMB coeff.	0.08	0.02	0.32	0.36	0.43	0.20	0.20	2.12

Tab. 5

The Frequency of Occurrence of the Groups of Consonants and the other statistical Characteristics of Five Ugric and Four Volgaic Languages. The Compactness of the mixed Ugric and Volgaic Language Taxon by Eight Features.

N	Consonant Language	Labia %	Front %	Palat %	Sonor %	Occlu %	Fricat %	Voice %
1	Mansi Sos	13.56	30.08	06.79	32.03	17.00	12.04	02.74
2	Mansi Kon	12.29	29.72	12.30	30.07	16.56	16.15	04.50
3	Hanty Kaz	12.60	30.83	07.60	30.96	17.19	11.48	00.00
4	Hanty East	10.45	30.81	05.19	21.83	24.20	13.95	10.51
5	Hungarian	10.04	34.47	04.07	22.53	22.62	12.87	12.70
6	Mari Mnt	09.99	33.90	06.06	07.92	24.62	16.35	16.90
7	Mari Lawn	09.47	37.95	01.90	09.28	23.81	18.22	16.57
8	Mordo Erz	13.72	36.78	01.76	07.44	23.37	21.36	14.97
9	Mord Mok	11.26	36.70	01.71	08.90	20.78	20.86	16.93
m	Mean	11.49	33.47	05.26	25.56	19.37	14.65	07.75
s	Standard	01.61	03.21	03.46	04.27	02.93	02.15	04.31
sI	S-squared	02.59	10.30	11.97	18.23	08.60	04.60	18.58
V%	Coeff. Var.	14.00	00.58	65.66	16.72	15.13	14.64	55.64
χ^2	Chi-square	01.80	02.46	18.21	05.71	03.55	02.51	19.18
TMB	TMB coeff.	00.12	00.16	01.17	00.37	00.23	00.16	01.24
Hlf CI	Half Confid. Interval	00.99	01.99	05.26	02.65	01.82	01.33	02.67

Tab.6

The Compactness of the Volgaic Language Taxon by Eight Features. The Volgaic Subgroup of the Finno-Ugric Language Family.

N	Consonant Language	Labia %	Front %	Palat %	Back %	Sonor %	Occlu %	Fricat %	Voice %	Sum, %
1.	Mari Mnt	09.99	33.90	06.06	07.92	24.62	16.35	16.90	09.43	57.8

130

N	Language									
										7
2.	Mari Lawn	09.47	37.95	01.90	09.28	23.81	18.22	16.57	08.89	58.60
3.	Mordo Erz	13.72	36.78	01.76	07.44	23.37	21.36	14.97	11.42	59.70
4.	Mord Mok	11.26	36.70	01.71	08.90	20.78	20.86	16.93	09.52	58.57
m	Mean	11.11	36.33	02.86	08.39	23.15	19.20	16.34	09.82	
s	Standard	01.90	01.72	02.14	00.85	01.66	02.35	00.93	01.11	
s²	S-squared	03.61	02.96	04.58	00.72	02.76	05.52	00.86	01.23	
V %	Coeff. Var.	17.10	04.73	74.83	10.13	07.17	12.24	05.69	11.30	
χ²	Chi-square	00.97	00.24	04.80	00.26	00.36	00.86	00.16	00.38	
T M B	TMB coeff.	00.12	00.03	00.61	00.03	00.05	00.11	00.02	00.05	
Hl f CI	Half Confid. Interval	03.49	03.16	03.93	01.56	03.05	04.32	01.71	02.04	

Tab. 7

The Compactness of the Balto-Finnic Language Taxon by Eight Features. The Balto-Finnic Subgroup of the Finno-Ugric Language Family.

N	Consonant Language	Labia %	Front %	Palat %	Back %	Sonor %	Occlu %	Fricat %	Voice %
1.	Veps	11.11	24.87	10.46	11.52	19.30	24.71	13.95	13.97
2.	Vodian	11.95	33.62	02.68	07.66	20.71	21.93	13.26	08.50
3.	Estonian	10.21	35.18	01.62	07.69	22.45	20.45	11.80	09.82
4.	Karelian (Tihvin)	09.66	24.79	09.83	09.89	21.73	20.36	12.08	08.02
5.	Karelian (Livvik)	09.66	24.79	09.83	09.89	21.73	20.36	12.08	08.02
6.	Karelian (Ludik)	08.66	34.53	01.43	10.38	19.01	21.67	14.32	11.80
7.	Finnish	08.73	34.44	02.19	08.75	23.32	18.00	12.79	03.57
m	Mean	10.21	31.03	04.69	09.49	21.20	20.82	13.41	09.29
s	Standard	01.27	04.59	03.87	01.49	01.60	02.23	00.88	03.31
sI	S-squared	01.61	21.07	14.98	02.23	02.57	04.96	00.77	10.96
V%	Coeff. Var.	12.41	14.79	82.58	15.73	07.56	10.69	06.54	35.60
χ²	Chi-square	00.95	04.07	19.16	01.41	00.73	01.43	00.34	07.08
TM B	TMB coeff.	00.08	00.32	01.52	00.11	00.06	00.11	00.03	00.56
Hlf CI	Half Confid. Interval	00.93	03.36	02.85	01.09	01.18	01.63	00.64	02.49

Tab. 8

The Frequency of Occurrence of the Groups of Consonants and the other statistical Characteristics of the Volgaic and Balto-Finnic Languages. The

Compactness of the Mixed Volgo-Finnic Language Taxon by Eight Features.
Mean coefficient of variation (V%) = 23.22. Mean TMB=0.35.

N	Consonant Language	Labia %	Front %	Palat %	Back %	Sonor %	Occlu %	Fricat %	Voice %
1.	Mari Mnt	09.99	33.90	06.06	07.92	24.62	16.35	16.90	09.43
2.	Mari Lawn	09.47	37.95	01.90	09.28	23.81	18.22	16.57	08.89
3.	Mordo Erz	13.72	36.78	01.76	07.44	23.37	21.36	14.97	11.42
4.	Mord Mok	11.26	36.70	01.71	08.90	20.78	20.86	16.93	09.52
5	Veps	11.11	24.87	10.46	11.52	19.30	24.71	13.95	13.97
6	Vodian	11.95	33.62	02.68	07.66	20.71	21.93	13.26	08.50
7	Estonian	10.21	35.18	01.62	07.69	22.45	20.45	11.80	09.82
8	Karelian (Tihvin)	09.66	24.79	09.83	09.89	21.73	20.36	12.08	08.02
9	Karelian (Livvik)	09.66	24.79	09.83	09.89	21.73	20.36	12.08	08.02
10	Karelian (Ludik)	08.66	34.53	01.43	10.38	19.01	21.67	14.32	11.80
11	Finnish	08.73	34.44	02.19	08.75	23.32	18.00	12.79	03.57
m	Mean	10.54	32.96	4.02	9.09	21.90	20.23	14.47	9.48
s	Standard	1.50	4.55	3.35	1.37	1.83	2.30	1.71	2.65
s²	S-squared	2.25	20.70	11.22	1.87	3.33	5.30	2.91	7.00
V%	Coeff. Var.	14.22	13.80	83.27	15.04	8.34	11.37	11.79	27.90
χ^2	Chi-square	2.13	6.28	27.91	2.06	1.52	2.62	2.01	7.38
TMB	TMB coeff.	0.12	0.34	1.52	0.11	0.08	0.14	0.11	0.40
Hlf CI	Half Confid. Interval	0.82	2.48	1.83	0.75	0.99	1.26	1.23	1.44

Tab. 9
The Total and the Mean Values of the Variance Coefficient and the TMB
Coefficient by Eight Features in the Ugric Language Taxon. The Introduction of
the Permic and Volgaic Languages into the Ugric Language Taxon.

Language Taxon	Total V%	Mean V%	Total TMB	Mean TMB
Ugric	221.27	27.66	3.77	0.47
Ugric + Permic	191.94	23.99	3.73	0.47
Ugric + Volgaic	210.78	26.35	3.63	0.45

Tab. 10
The Total and the Mean Values of the Variance Coefficient and the TMB
Coefficient by Eight Features in the Volgaic Language Taxon. The Introduction
of the Balto-Finnic Languages into the Volgaic Language Taxon.

Language Taxon	Total V%	Mean V%	Total TMB	Mean TMB
Volgaic	143.19	17.90	1.02	0.13
Volgaic + Balto-Finnic	185.73	23.22	2.82	0.35

Tab. 11
The total and the mean Values of the Variance Coefficient and the TMB
Coefficient by Eight Features in the Balto-Finnic Language Taxon.

Language Taxon	Total V%	Mean V%	Total TMB	Mean TMB
Balto-Finnic	185.90	23.24	2.79	0.35

Tab. 12
The total and the mean Values of the Variance Coefficient by Eight Features in
the Finno-Ugric Language Family in Comparison with the other Language
Taxa.

Language Taxon	Mean V%
Ugric	27.66
Balto-Finnic	23.24
Volgaic	17.90
Western Slavonic	10.07
Eastern Slavonic	7.72
North-West Iranian	6.48

REFERENCES

ABONDOLO, D. Hungarian. In *The World's Major Languages* (Edited by Bernard Comrie). Oxford: Oxford University Press, 1990. pp. 577 - 592.

APRESJAN, Y. D. Algoritm postroenija klassov po matritse rasstojanij. [Algorithm of Constructing Classes by the Matrix of Distances]. In *Mashinnyj perevod i prikladnaja lingvistika,* Moscow: Moscow State Pedagogical Institute of Foreign Languages, 1966, vol. 9, pp. 72 - 79.

ARKADJEV, A. G., BRAVERMAN E. M. *Obuchenije mashiny raspoznavaniju obrazov.* [Teaching a Machine Pattern Recognition]. Moscow: Nauka, 1964. 112 p.

ARKADJEV, A. G., BRAVERMAN, E. M. *Obuchenije mashiny klassifikatsii objectov.* [Teching a Machine Object Classification]. Moscow: Nauka, 1971. 192 p.

AUSTERLITZ, R. Uralic languages. In *The World's Major Languages.* (Edited by Bernard Comrie). Oxford / New York: Oxford University Press, 1990, pp.567 - 576.

BONGARD, M. M. *Problema uznavanija.* [The Problem of Recognition]. Moscow: Nauka, 1967 – 320 pp.

BUTLER, Christopher. *Statistics in Linguistics.* Oxford: Basil Blackwel, 1985.

CRYSTAL, David. *An Encyclopaedic Dictionary of Language and Languages.* Oxford: Blackwell, 1992. 428 p.

DRUJANOV L. A. Zakony prirody i ih poznanije. - Moskva: Prosveshchenije, 1982.

ELISEEVA I. I., RUKAVISHNIKOV V. O. *Gruppirovka, korreljatsija, raspoznavanie obrazov. (Statisticheskie metody, klassifikatsija i izmerenie svjazej).* [Grouping, Correlation, Pattern Recognition. (Statistical Methods, Classification and Connection)]. Moscow: Statistika, 1977. 144 p.

Entsyklopedija kibernetiki. [Encyclopedia of Cybernetics]. Kiev: Ukrainian Sovetsk. Entsiklopedija, 1975. 607p.

GAMKRELIDZE, T. V., IVANOV, V. *Vs. Indoevropejskij jazyk i Indoevropejtsy.* [The Indo-European Language and the Indo-Europeans]. (Parts 1 and 2). Tbilisi: Tbilisi State University, 1984. 1329 p.

GADZHIEVA N. Z. K voprosu o klassifikatsii t'urkskih jazykov i dialektov. In *Teoreticheskie osnovy klassifikatsii jazykov mira.* Moskva: Nauka, 1980, pp. 100 - 126.

HORNBY A. S. (with assistance of Christina Ruse). *Oxford student's dictionary of current English.* Oxford: Oxford University Press, 1984.

JAHONTOV, S. E. Otsenka stepeni blisosti rodstvennyh jazykov. [Estimation of the degree of closeness of the related languages]. In *Teoreticheskie osnovy klassifikatsii jazykov mira.* [Theoretical Fundamentals of Classification of World Languages]. Moscow: Nauka, 1980, pp. 148-157.

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, pp. 24 - 61.

Jazyki i dialekty mira. [Languages and Dialects of the World]. Moskva: Nauka, 1982. 208 pp.

Jazyki Mira. Ural'skie jazyki. Moskva: Nauka, 1993. 398 pp.

KALMAN, B. The history of the Ob-Ugric languages. In *The Uralic Languages.* (Edited by Denis Sinor). Leiden\New York: E. J. Bril, 1988, pp. 395-412.

KONDAKOV, N. I. *Logicheskij slovar'-spravochnik.* [Logical Dictionary-Reference Book]. Moscow: Nauka, 1975. 720 p.

KORDONSKIJ, S. G. Taksony i analiticheskie objekty kak sistemy. [Taxa and analytical objects as systems]. In *Sistemnyj metod i sovremennaja nauka.* Novosibirsk: Novosibirsk, 1983, pp. 141 - 150.

KORDONSKIJ, S. Tsikly dejatel'nosti i ideal'nyje objekty. Moskva: Pantori, 2001.

KUENNAP, A. *Breakthrough in Present-day Uralistics.* Tartu: University of Tartu, 1998.

KUENNAP, A. Review on the book by Angela Marcantonio "The Uralic Language Family: Facts, Myths and Statistics. Oxford (UK) and Boston (USA): Blackwell Publishers, 2002, 335 p. In *Eurasian Studies Yearbook,* # 75, 2003, pp. 171 - 176.

KUHN, T. *Struktura nauchnyh revoll'utsij.* [Kuhn T. S. The structure of scientific revolutions]. Moskva: Progress, 1977. 300 p.

MAJTINSKAJA, Obsko-Ugorskie jazyki [The Ob-Ugric languages]. In *Jazyki narodov SSSR. T.3. Finno-Ugorskie i Samodijskie jazyki.* [Languages of the Peoples of the USSR. Finno-Ugric and Samoyedic Languages]. Moskva: Nauka, 1966, p. 316 - 362.

MARCANTONIO, A. *The Uralic Language Family. Facts, Myths and Statistics.* Oxford: Blackwell Publishers, 2002. 335 p.

MARTYNENKO, G. *Osnovy stilemetrii.* Leningrad: Leningrad University Press, 1988.

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 p.

MAYR, E. *Printsipy zoologicheskoj sistematiki.* [Principles of Zoological Systematics]. Moscow: Mir, 1971. 454 p.

MELJNIKOV, G. P. *Sistemologija i jazykovye aspekty kibernetiki.* [Systemology and the Language Aspects of Cybernetics]. Moscow: Sovetskradio, 1978. 368 p.

MEL'NIKOV, G. P. *Sistemnaja tipologija jazykov: Printsypy, metody, modeli.* [Systemic Typology of Languages: Principles, Methods, Models]. Moskva: Nauka, 2003. 395 p.

MIRKIN, B. G. *Analiz kachestvennyh priznakov (matematicheskije metody i modeli).* Analysis of quantitative features (mathematical methods and models). Moskva: Statistika, 1976.

OALD. *Oxford Advanced Learner's Dictionary of Current English.* Oxford: Oxford University Press, 1998. 1428 p.

OZHEGOV, S. I. *Slovar' russkogo jazyka.* [Dictionary of the Russian language]. Moskva: Sovetskaja entsiklopedija, 1970.

PAVLOVSKIJ, Z. *Vvedenije v matematicheskuju statistiku.* Moskva: Statistika, 1967.

PETROV, V. V. *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 p.

RASSADIN, V. 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, pp.103-108.

RAUN, A. *Ueber die sogenannte lexikostatistische Methode oder Glottochronologie und ihre Amwendung auf Finnisch-Ugrische und Tuerkische.* Ural-Altaische Jarbucher, 28, pp. 151 - 154.

RICHMOND, S. B. *Statistical analysis.* New York: The Ronald press company, 1964.

ROZOVA, S. S. *Klassifikatsionnaja problema v sovremennoj nauke.* [Classification Problem in Modern Science]. Novosibirsk: Nauka, 1986. 222 p.

PORHOMOVSKIJ, V. J. Problemy geneticheskoj klassifikatsii jazykov Afriki. [Problems of genetical classification of African languages]. In *Teoreticheskie osnovy klassifikatsii jazykov mira.* [Theoretical Fundamentals of Classification of World Languages]. Moscow: Nauka, 1982, pp. 258 - 312.

SEREBRENNOKOV, B. A. Problema dostatochnosti osnovanij v gipotezah, kasajushchihs'a geneticheskogo rodstva jazykov. [The problem of substantiated foundations in the hypothesis concerning the genetic relatedness of languages]. In *Teoreticheskie osnovy klassifikatsii jazykov mira. Problemy rodstva.* Moskva: Nauka, 1982, pp. 6 - 62.

SHARADZENIDZE, T. S. Rodstvo jazykov, protsessy divergentsii, konvergentsii i sootvetstvujushchie im klassifikatsii jazykov mira. [Genetical 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, pp. 63 - 107.

SIS'KOV, V. I. Ob opredelenii velichiny intervala pri gruppirovkah. In *Vestnik statistiki,* # 12, 1971, pp. 3 - 11.

SIS'KOV, V. I. *Korrel'atsionnyj analiz v ekononicheskih issledovanijah.* - Moskva: Statistika, 1975.

TAAGEPERA, R. The linguistic distances between Uralic languages. In *Linguistica Uralica,* 1994, XXX, pp. 161 - 167.

TAAGEPERA, R. KUENNAP, A. Distances among Uralic and other Northern Eurasian languages. In *Linguistica Uralica,* 2005, Vol. XLI, # 3, pp. 161.

TAMBOVTSEV, Y. 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 the Russian Academy, 1976, pp. 161-164.

TAMBOVTSEV, Y. Spektral'nye harakteristiki udarnyh tsentral'nozadner'adnyh glasnyh literaturnogo mansijskogo jazyka. [Spectral characteristics of the stressed central-back vowels of the Mansi literary language]. In *Fonetika sibirskih jazykov.* [Phonetics of the Siberian languages]. Novosibirsk: Institut Istorii, Filosofii I Filologii of the Siberian Branch of the Academy of Sciences of the USSR, 1979, pp. 87 – 93.

TAMBOVTSEV, Y. Sochetanie fonem v mansijskom jazyke. [Phonemic combinations in the Mansi language]. In *Teoreticheskie voprosy fonetiki i grammatiki jazykov narodov SSSR.* [Theoretical problems of phonetics and grammar of the languages of the peoples of the USSR]. – Novosibirsk: Novosibirsk University Press, 1979, pp. 30 – 39.

TAMBOVTSEV, Y. Glasnye mansijskogo jazyka po dannym rentgenografii. [Mansi vowels by the X-ray data]. In *Issledovanie zvukovyh system sibirskih jazykov.* [Investigations of the sound

135

systems of Siberian languages]. - Novosibirsk: Institut Istorii, Filosofii I Filologii of the
Siberian Branch of the Academy of Sciences of the USSR, 1979, pp. 78 – 89.

TAMBOVTSEV, Y. Konsonantnyj koeffitsijent v jazykah raznyh semej. [The Consonant Coefficient
in the Languages of Different Language Families]. - Novosibirsk: Novosibirsk University,
1986. 16 p.

TAMBOVTSEV, Y. 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, pp.13-20.

TAMBOVTSEV, Y. *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 p.

TAMBOVTSEV, Y. *Tipologija uporjadochennosti zvukovyh tsepej v jazyke.* [Typology of the
Orderliness of Sound Chains in Language]. Novosibirsk: Novosibirsk University, 1994. 199
p.

TAMBOVTSEV, Y. *Kompendium osnovnyh statisticheskih harakteristik funktsionirovanija
soglasnyh fonem v zvukovoj tsepochke anglijskogo, nemetskogo, frantsuzskogo i drugih
indoevropejskih jazykov.* [Compendium of the Basic Statistical Characteristics of the
Functioning of Consonants in the Sound Chains of English, German, French and Other Indo-
European Languages]. Novosibirsk: Novosibirskij klassicheskij institut, 2001. 129 p.

TAMBOVTSEV, Y. *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 p.

TAMBOVTSEV, Y. *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 p.

TAMBOVTSEV, Y. The phonological distances between Mongolian and Turkic languages based on
typological consonantal features. In *Mongolian studies. Journal of the Mongolian Society*
(USA), Vol.24, 2001, pp.41 - 84.

TAMBOVTSEV, Y. Comparative typological study of language distances based on the consonants in
the 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, pp.77-80.

TAMBOVTSEV, Y. Is Kumandin a Turkic language? In *Dilbilim Arashtirmalari* (Istanbul), 2002,
pp. 63 - 104.

TAMBOVTSEV, Y. 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, pp. 13-
14.

TAMBOVTSEV, Y. 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: Universitet
Tuebingen, 2002, pp. 138 - 139.

TAMBOVTSEV, Y. Lingvisticheskaja taksonomija: kompaktnost' jazykovyh podgrupp, grupp i
semej. [Linguistical taxonomy: Compactness of subgroups, groups and families]. In
Baltistica. Journal of Baltic Linguistics (Vilnius), Volume 37, #1, 2002, pp. 131 - 161.

TAMBOVTSEV, Y. 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, pp.358 - 359.

TAMBOVTSEV, Y. *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 Chains of Indo-European, Paleo-Asiatic, Uralo-Altaic and Other World Languages: Compactness of Subgroups, Groups, families and Other Language Taxa]. Novosibirsk: Sibirskij Nezavisimyj Instutut, 2003. - 143 pp.

TAMBOVTSEV, Y. Izmerenije fonostatisticheskih rasstojanij mezhdu ural'skimi jazykami. [Measuring phono-statistical distances between Uralic Languages]. - In *Fenno-Ugristica,* 25. Tartu: Tartu University Press, 2003, pp. 120 - 168.

TAMBOVTSEV, Y. Phonological similarity between Tofa and some other languages of Central Asia based on the frequency of occurrence of certain typological consonantal features. In Fenno-Ugristica, 25. - Tartu: Tartu University Press, 2003, pp. 169 - 179.

TAMBOVTSEV, Y. Klassifikatsija jazykov na osnove novyh metodov opredelenija mezhjazykovyh rasstojanij (na primere jazykov samodijskoj i finno-ugorskoj semej). [Classification of languages on the basis of new methods of measuring distances between languages (on the sample of Samoyedic and Finno-Ugric families). In *Fenno-Ugristica,* # 27, 2005 (Tartu), pp. 122 - 140.

TRET'JAKOV, P. N. *U istokov drevnerusskoj narodnosti.* [At the Source of the Old Russian people]. Leningrad: Nauka, 1970, 156 p.

TRET'JAKOV, P. N. *Po sledam drevnih slavjanskih plemen.* [Along the Traces of the Ancient Slavonic Tribes]. Leningrad: Nauka, 1982. 143 p.

VASILJEV, V. I. *Raspoznajushchije sistemy.* [Recognition Systems]. Kiev: Naukova Dumka, 1969. 392 p.

VENCHIKOV. A. I. et al. *Osnovnye prijomy statisticheskoj obrabotki rezul'tatov nabl'udenij v oblasti fiziologii.* Moskva: Meditsina, 1974.

VINOGRADOV, V. A. 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, pp. 224 - 256.

VINOGRADOV, Victor Alekseevich. Funktsionaljno - tipologicheskie kriterii i genealogicheskaja klassifikatsija jazykov. [Functional - typological criteria and the genealogical classification of languages]. - In: Teoreticheskie osnovy klassifikatsii jazykov mira. [Theoretical Fundamentals of Classification of World Languages]. - Moscow: Nauka, 1982, pp. 258 - 312.

Webster's seventh new collegiate dictionary. Springfield: Merriam company, 1965.

ZAGORUJKO, N. G. *Metody raspoznavanija i ih primenenie.* [Methods of Recognition and their Application]. Moscow: Sovetskradio, 1972. 208 p.

ZAGORUJKO, N. G. *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 p.

ZAGORUJKO, N. G., TAMBOVTSEV, Y. Statisticheskoe raspredelenie udarenija v russkom slove. [Statistical Distribution of Stress in Russian Words]. In *Materialy seminara "Statisticheskaja optimizatsija prepodavanija jazykov i inzhenernaja lingvistika."* Chimkent: Chimkent Pedagogical Institute, 1980, pp. 323 - 324.

ZAJTSEV, G. N. *Matematika v eksperimental'noj botanike.* [Mathematics in Experimental Botany]. Moskva: Nauka, 1990.