THE PHONO-TYPOLOGICAL DISTANCES BETWEEN AINU AND THE OTHER WORLD LANGUAGES AS A CLUE FOR CLOSENESS OF LANGUAGES

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

The article deals with one of the genetically isolated languages — Ainu. It is usually a common practice in linguistics to provide a genetic identification of a language. The generic identity of a language is the language family to which it belongs. Therefore, is advisable to find a family for every isolated language.

The new method of phonostatistics proposed here allows a linguist to find the typological distances between Ainu and the other languages of different genetic language families. The minimum distances may be a good clue for placing Ainu in this or that language family. The result of the investigation shows the minimum typological distance between Ainu and the Quechuan family (American Indian languages).

Key words: consonants, phonological, distance, typology, frequency of occurrence, speech sound chain, statistics, closeness

Ainu is a genetically isolated language.² There are many different theories as to the origin and ethnic development of Ainu. Some scholars think they belong to the Manchu-Tungus tribes while others link them to the Palaeo-Asiatic peoples. A. P. Kondratenko and M. M. Prokofjev point out that some

¹ WHALEY, L.J. Introduction to Typology: The Unity and Diversity of Language, pp. 18-23.

² CRYSTAL, D. An Ecyclopedic Dictionary of Language and Languages, p. 11 and also Jazyki i dialekty mira. [Languages and Dialects of the World], p. 23.

scholars connect them to the American Indians.³ Some anthropological data show that Ainu are close enough to the American Indians.4

There are some other languages, which have not been placed into any language family: Japanese, Korean, Nivkhi, Yukaghir and Ket (Yug). However, for the latter, a new language family - Yenissey has been invented. Nevertheless, it is not a solution of the problem.

The aim of this article is to find the typological closeness of Ainu to the language families, to which it may be supposed to enter. Usually, genetically close languages are also typologically close. However, the typologically close languages may or may not be genetically close. Nevertheless, in the majority of cases typologically close languages are genetically close. We can find the phonostatistical closeness, which can give a good clue for the genetic relatedness.5

Why should one use quantitative methods in studying languages? A great philosopher and scientist Emmanuel Kant (1724 - 1804) in his well-known works explaining the structure of the world stated that everything in this world possesses quantity and quality. Actually, quantity may go over into quality when it is great enough.6

Long ago, in 1935, George Kingsley Zipf stated that it was necessary to introduce the so-called "Dynamic Philology" to achieve fruitful results in studying the structure and entity of language. As George A. Miller correctly put in the introduction to Zipf's book, one who wishes to study a rose should count its petals, not just enjoy it. G. K. Zipf believed that it is necessary to study the massive statistical regularity of every linguistic unit or phenomenon.8

Quantitative research needs the use of mathematical statistics. One cannot help agreeing with Christopher Butler, who requires a quantitative treatment in

⁴ KONDRATENKO, A. P., PROKOFJEV, M. M. O meste ainov v sisteme rasovoj klassifikatsii narodov mira. [About the place of Ainu in the system of the racial classification of the peoples of the world], pp. 38-44.

⁶ FROLOV, I. T. (ed.). Filosofskij slovar' [A Dictionary of Philosophy], p. 144.

⁸ ZIPF, G. K. The Psycho-Biology of Language. An Introduction to Psycho-Biology of Language, pp. 5-6.

³ KONDRATENKO, A. P., PROKOFJEV, M. M. O meste ainov v sisteme rasovoj klassifikatsii narodov mira. [About the place of Ainu in the system of the racial classification of the peoples of the world], pp. 3-5.

⁵ TAMBOVTSEV, Y. (2001d); TAMBOVTSEV, Y. (2001e); TAMBOVTSEV, Y. (2002a); TAMBOTSEV, Y. (2002b); TAMBOTSEV, Y. (2002c); TAMBOTSEV, Y. (2002d); TAMBOTSEV, Y. (2003a); TAMBOTSEV, Y. (2003b); TAMBOTSEV, Y. (2004).

⁷ ZIPF, G. K. The Psycho-Biology of Language. An Introduction to Psycho-Biology of Language, p. 12.

any linguistic research because it is difficult otherwise to understand and evaluate how relevant are the linguistic results.9

Establishing genetic language families, linguists compare every language with some other language or a group of languages. Jiří Krámský is correct to remark that one can establish a typology of languages based on the quantitative data received after comparing languages. The quantitative data gives a clearer vision of the differences and similarities between languages. The quantitative load of particular language phenomena is different in different languages. J. Krámský is quite right to observe that in linguistics there is a very close relation between quality and quantity, even if the conditions of the transition of quantity into quality are not established so safely as they are in natural sciences. Nevertheless J. Krámský assumes that in linguistics qualitative changes are asserted with the help of quantitative factors.

Our method measures distances between languages on the phonological level. It gives a vivid picture of the typological similarity of the sound pictures of the languages under investigation. It allows us to find out the archetype of this or that language family. The mean values the frequency of the consonantal groups.

The use of quantitative data ensures that the languages are similar if the frequency of occurrence of certain linguistic units is similar. It takes into account both cases when the units are used very frequently or very seldomly. However, in classical linguistics, where the frequency is not taken into consideration, it is more often than not that the usual elements are compared with the rare elements. J. Krámský is correct to point out that the language units which are in the centre of one language system should not be compared to those from the periphery of another. The quantitative analysis shows us the units, which are in the centre from a language system and those which are at the periphery of it. Therefore, the typology of languages based on the quantitative data may add much to the established language families. In the centre from the established language families.

Ainu, like any other human language, has a specific structure of the speech sound chain. It can be distinguished by its structure from any other language. Every language has a unique structure of distributions of speech sounds in its phonemic chain. The distribution of Ainu vowels will not be considered till the

⁹ BUTLER, C. Statistics, pp. 255-264.

KRÁMSKÝ, J. On Some Problems of Quantitative Typology of Languages on Acoustic Level, pp. 15-18 and also KRUPA, V. On Quantification of Typology, pp. 31-36.

¹¹ KRÁMSKÝ, J. On Some Problems of Quantitative Typology of Languages on Acoustic Level, pp. 15-26.

¹² TAMBOTSEV, Y. (2001a); TAMBOTSEV, Y. (2001b); TAMBOTSEV, Y. (2001c); TAMBOTSEV, Y. (2003).

second stage of the investigation. The frequency of occurrence will be considered if and only if the frequency of occurrence of different groups of consonants will not differentiate Ainu from the other world languages. Let us point out that consonants bear the semantic load in the word, not vowels. Therefore, it is more possible to understand the meaning of the message by consonants, rather by vowels. However, if we fail to recognize and distinguish two languages, then we resort to the structure of occurrence of vowels in the speech sound chain. While comparing languages, it is necessary to keep to the principle of commensurability. Having it in mind, it is not possible to compare languages on the basis of the frequency of occurrence of separate phonemes, because the sets of phonemes in languages are usually different. The articulatory features may serve as the basic features in phono-typological reasoning.

Before the computer measures the phonological distances, one has to choose the phonological features, which are necessary and sufficient. One has to select the system of the informative features. In pattern recognition such features are called basic.¹³ Therefore, we have chosen all the features basic for the articulation of any speech sound. In the first stage we shall deal with consonants.

First of all, it is the classification of consonants according to the work of the active organ of speech or place of articulation (4 features). Secondly, it is the classification from the point of view of the manner of articulation or the type of the obstruction (3 features). Thirdly, it is the classification according to the work of the vocal cords (1 feature). In this way, 8 basic features are obtained: 1) labial; 2) fore lingual or front; 3) mediolingual or palatal; 4) guttural or back or velar; 5) sonorant; 6) occlusive non-sonorant; 7) fricative non-sonorant; and 8) voiced non-sonorant consonants. One should take the values of the frequency of occurrence of these 8 features in the speech chain of Ainu and compare them to those of the other languages. On the basis of the "chi-square" test and Euclidean distance, we have developed our own method of measuring the phonotypological distances between languages. It takes into account the frequency of occurrence of the 8 consonantal groups mentioned above and builds up the overwhelming mosaic of the language sound picture.

It is very important to find some typological characteristics in order to endeavour to place it in some defined language family. Some linguists consider it impossible to put Ainu in any of the known language families because it is still insufficiently studied. Actually, it is considered here that it is possible to

¹⁴ TAMBOTSEV, Y. (1994a); TAMBOTSEV, Y. (1994b); TAMBOTSEV, Y. (2004).

¹³ ZAGORUIKO, N. G. Prikladnye metody analiza dannyh i znanij. [Applied Methods of Data and Knowledge Analysis], pp. 54-75.

put Ainu in a language family if its phonostatistical characteristics are studied better.

Some ethnographers (e.g. S. A. Arutjunov) believe Ainu to be the direct offsprings of the ancient Neolithic population, which migrated from the continent to the Japanese islands, especially Hokkaido. It is the most Northern of the Japanese islands. It is still unknown if the Ainu people used to live on the lands around the Amur River or on the shore of the Okhotsk Sea. However, it is known that in the 18 – 19th century Ainu did live on the Sakhalin Island and the Kuril Islands. Unfortunately, nowadays Ainu are completely assimilated by the Japanese. Even at home they speak Japanese.

Only some of them can still speak Ainu. Ainu physical appearance differs from that of the Japanese. They look very much like the population of South-East Asia. Ainu did not keep the domestic animals since they were fishers and hunters. ¹⁵

The first scientifically reliable linguistic data were introduced by the publication of the Ainu-Russian Dictionary by M. M. Dobrotvorski (Kazan, 1875). A great contribution was made by B. Pilsudski, who published his materials later (Pilsudski B. Materials of the Ainu language and folklore. Krakow, 1912).

A. A. Leontjev believes that Ainu lacks the opposition between the voiced and voiceless consonants. Thus, [b, d, g] are only the positional variants of [p, t, k] between two vowels. In his opinion, the vowel system is rather poor. The Ainu language belongs to the agglutinative languages, though it has some affixation.¹⁶

We fed into the computer the Ainu myths, songs and stories published in the book by Nikolai Aleksandrovich Nevski. N. A. Nevski collected the Ainu folklore on Hokkaido Island under the guidance of the well-known specialist of Ainu Prof. K. Kindaiti.

According to N. A. Nevski, the tribal name "Ainu" means "a noble man" or "a worthy man". Speaking about the ethnic history of Ainu, N. A. Nevski remarks that it is full of the fight with the Japanese. Earlier the Ainu tribes lived much further to the South, but the Japanese ousted them to the most Northern Japanese Island – Hokkaido. Under Japanese pressure the Ainu tribes moved further to the North, namely to the Kuril Islands and to the Sakhalin Island. ¹⁸

When in 1915 N. A. Nevski came to the Hokkaido Island, he found the Ainu living in polygamous families. It allowed Ainu men not to work. N. A. Nevski

¹⁵ Bol'shaja Sovetskaja Entsiklopedija [Great Soviet Encyclopaedia], vol. 1, pp. 306-307.

Bol'shaja Sovetskaja Entsiklopedija [Great Soviet Encyclopaedia], vol. 1, p. 307.

¹⁷ NEVSKI, N. A. Ainskij Fol'klor [Ainu Folklore], p. 175.

¹⁸ NEVSKI, N. A. Ainskij Fol'klor [Ainu Folklore], p. 9.

found Ainu villages on the banks of the Sarapet River in Hidaka province on Hokkaido Island.

N. A. Nevski could speak Ainu and Japanese very well. Nevertheless, he did not think that Ainu is close to Japanese. He thought it closer to the Palaeo-Asiatic languages. We will try to verify this idea by our method in this article. Though the Ainu were few, the Ainu language had several dialects. N. A. Nevski defined 3 dialects in Ainu: 1) the dialect of the Northern-Eastern part of Hokkaido; 2) the dialect of the Northern-Western part with the valley of the Isikari River in the centre. It is close to Sakhalin Island; 3) the dialect of the South-Eastern part, which is very different from the Sakhalin dialect, so that communication is hard.

After N. A. Nevski, we define the following Ainu phonemes:

Vowels: [i, u, e, o, a]

Consonants: [p, w, m, t, (d), ts, s, n, r, j, k, (g), ŋ, h]

The classification of the Ainu consonants by the work of the active organ of speech (i.e. place of articulation):

Labial: [p, w, m]

Forelingual (front): [t, (d), ts, s, n, r]

Mediolingual (palatal): [j]

Guttural (velar or back): [k, (g), ŋ, h]

The classification by the manner of articulation (the character of the

obstruction):

Sonorant: $[w, m, n, r, j, \eta]$

Occlusive non-sonorant: [p, t, (d), ts, k, (g)]

Fricative non-sonorant: [s, h]

The classification by the work of the vocal cords:

Voiced non-sonorant consonants: none. Some spontaneous facultative voicing of [t, k] is possible.

After computing the Ainu text, we received the following frequencies of the phonemic occurrence in the sound chain:

	Frequency	% of all ph.	% of cons.
Labial:	1,057	9.28	18.26
Forelingual (front):	2,724	23.91	47.05
Palatal (mediolingual):	583	5.12	10.07
Guttural (back):	1,425	12.51	24.62
Sonorant:	2,932	25.74	50.65
Occlusive non-sonorant:	2,177	19.11	37.60
Fricative non-sonorant:	680	5.97	11.75
Voiced non-sonorant:	3	0.03	0.05

¹⁹ NEVSKI, N. A. Ainskij Fol'klor [Ainu Folklore], p. 10.

The total of consonants: 5,789 - 50.82%The total of vowels: 5,603 - 49.18%

The value of the consonantal coefficient (i.e. the ratio of consonants to vowels): 1.03 Sample volume of the Ainu text: 11,392 phonemes.

In pattern recognition many measures of distances between two objects are used. Nikolai G. Zagoruiko recommends use of Euclidean distances when the value of the features is equal.²⁰ We consider all our features to be equal since we cannot claim that the frequency of occurrence of labials is more important than the frequency of occurrence of sonorants or the frequency of occurrence of palatals is more important than the frequency of occurrence of the fricatives and so on.

It is necessary to introduce some system of references when dealing with the distances between Ainu and the other languages. Such point may be the distance between two texts in some language. We calculated the distances between two texts in the Marquesan language. It is 0.505. The closest to Ainu was one of the Austronesian languages – Tagalog with the distance 9.310. Let us remember the words of N. A. Nevski that Ainu is close to the Palaeo-Asiatic languages. Indeed, one of the Palaeo-Asiatic languages, namely the Chukchi language with the distance 10.954 is rather close. The next closest language is also a Palaeo-Asiatic language – Koryak with the distance 12.781. Korean is a bit closer – 12.636. Japanese is further away – 15.269. As we can see from the tables below the other languages are also rather far away. So, the closest Manchu-Tungus language is Ul'ch with the distance 13.464.

However, the closest to Ainu turned out to be the American Indian languages of the North and South America. So, Quechua has the distance of 5.451 and Inga 7.388. They both belong to the Quechua family of American Indian languages. Quechua and Inga Indians live in South America.

Having compared Japanese to some languages, we received the following phono-typological distances:

Japanese – Uyghur (6.77); Japanese – Nanay (8.12); Japanese – Yakut (8.26); Japanese – Sea Dayak (8.86); Japanese – Kazakh (9.02); Japanese – Turkish (9.05); Japanese – Ket (9.52); Japanese – Baraba Tatar (9.76); Japanese – Uzbek (10.63); Japanese – Hausa (10.98); Japanese – Georgian (11.05); Japanese – Kazan Tatar (11.07) and so on. One can see, that Uyghur, Yakut, Kazakh, Turkish, Baraba Tatar, Uzbek and Kazan Tatar are Turkic languages. Nanay is a Manchu-Tungus language. Therefore, one can notice that Japanese is

²⁰ ZAGORUIKO, N. G. Prikladnye metody analiza dannyh i znanij. [Applied Methods of Data and Knowledge Analysis], pp. 198-199.

closer to the so-called Altaic languages which include Turkic, Mongolian and Manchu-Tungus languages. Many world languages were compared to Japanese. We cannot show all the distances here for lack of space. However, the maximum distances were found for Japanese - German (22.24); Japanese -English (19.83); Japanese – Romanian (15.08) and Japanese – Swedish (17.03). As a conclusion, we can also state that speech sound picture of Japanese is rather far away from the languages, which are geographically close: Chinese, Nivkhi, Itelmen or Indonesian. It was a surprise to us. Our data state that the speech sound pattern of Japanese resembles that of Uyghur – one of the Turkic languages spoken in the Middle Asia. The Uyghur people are often linked to the Old Turkic tribes, who used to live in the steppes of Southern Russia before the Tatar-Mongols captured them in the 9th century A.D. We must point out that it is not a coincidence since the other native Altaic people have a very similar data of closeness to Japanese. Turkic and Manchu-Tungus tribes may have had a sort of common origin with Japanese. It may verify the Altaic hypothesis of Japanese origin. It is especially vivid, when the Austro-Oceanic and other languages do not show such a great closeness.

As a matter of fact, Ainu shows a great typological closeness to Quechua and Inga, which belong to the American Indian languages, geographically far away from Ainu. Explaining the close distances between Ainu and American Indian languages we must recall the original hypothesis put forward by some unknown Catholic monk and then picked up by the great mathematician Gotfried Wilhelm Leibnitz (1646 - 1716). In Russia it was developed by an outstanding archaeologist Aleksei Pavlovich Okladnikov. Actually, in 1938 he published an article in which he claimed that the people in the Americas originated from the Siberian tribes. According to his ideas the Neolithic people from Siberia migrated to the most Northern-Eastern point of Siberia. There they found the Bering ice bridge which allowed them to get to Alaska in Northern America.21 However, according to his theory the Neolithic peoples who used to live on the banks of the Angara and Lena Rivers and the Baikal Lake first moved towards to the East and got to the shores of the Pacific Ocean.²² I should guess part of these peoples moved eastward to the Japanese Islands. May be, ancient Ainu were in their number. Then other Neolithic tribes related to the Ainu moved further and got to South America but preserved their articulation basis. This is why, the distribution of the consonantal groups in Ainu and American Indian languages is typologically similar. A. P. Okladnikov points out that the anthropological features of American Indians and Siberian peoples are

²¹ OKLADNIKOV, A. P. Arheologicheskie dannye o drevneishej istorii Pribaikal'ja. [Archaeological data on the ancient history of the Lake Baikal area], p. 224.

OKLADNIKOV, A. P., VASIL'EVSKIJ, R. S. Po Al'aske i Aleutskim ostrovam. [Travels in the Alaska and the Aleut Islands], pp. 12-67.

similar. The other strong point in Okladnikov's reasoning is that South and North America never had any apes or monkeys from whom people may have developed. Actually, many animals from Siberia also crossed this ice Bering Bridge to the North America. This is why, not only the people but also the animals in Siberia and America are the same. In fact, the Bering ice bridge existed twice.

First, it was some 65-35 thousand years ago and then some 28-25 thousand years ago. It is supposed that each period during which it existed was not less than 18-15 thousand years. At least some 19 thousand years ago it existed. A. P. Okladnikov believed that the Americas were inhabited by two waves, i.e. in the middle and upper Palaeolithic period.²³ Our data support this theory. From the typological point of view, some American Indian languages (cf. Tab. 1-2) are very close to Ainu. At the same time, Ainu is not typologically close to Caucasian languages (cf. Tab.18) – especially, Kabardian (25.08) and Adyghe (30.546). One can find more details on the typological distances between Ainu and the other world languages in the tables (Tab.1 26).

In conclusion, it is possible to state a great typological closeness between Ainu and some American Indian languages. We are far from stating that genetically isolated language – Ainu – is genetically close to the languages of the Quechuan family, namely Quechua and Inga. However, from the point of view of typology Ainu is very similar to these American Indian languages. Having this typological clue, linguists may have a closer look at them from the genetic point of view.

REFERENCES

- Bol'shaja Sovetskaja Entsiklopedija [Great Soviet Encyclopaedia]. Vol. 1 30. Moskva: Sovetskaja Entsiklopedija, 1978.
- BUTLER, C. Statistics. In *Projects in Linguistics*. A *Practical Guide to Researching Language*. Alison Wray, Kate Trott and Aileen Bloomer with Shirley Reay and Chris Butler. London-New York: Arnold-Hodder, 1998. 303 p.
- CRYSTAL, D. An Encyclopedic Dictionary of Language and Languages. Oxford: Blackwell, 1992. 428 p.
- FROLOV, I. T. (ed.). Filosofskij slovar' [A Dictionary of Philosophy] Moskva: Politizdat, 1980. 445 p.

²³OKLADNIKOV, A. P., pp. 224-260, also OKLADNIKOV, A. P., VASIL'EVSKIJ, R. S., pp. 17-39.

- Jazyki i dialekty mira. [Languages and Dialects of the World]. Moskva: Nauka, 1982. 208 p.
- KONDRATENKO, A. P., PROKOFJEV, M. M. O meste ainov v sisteme rasovoj klassifikatsii narodov mira. [About the place of Ainu in the system of racial classification of the peoples of the world]. Yuzhno-Sahalinsk: Institut Morskoj Geologii i Geofiziki, 1989. 51 p.
- KRÁMSKÝ, J. On Some Problems of Quantitative Typology of Languages on Acoustic Level. In *Prague Studies in Mathematical Linguistics*, no. 3. Prague, 1972. pp. 15-26.
- KRUPA, V. On Quantification of Typology. In *Linguistics*, no. 12, 1965, pp. 31-36.
- NEVSKI, N. A. Ainskij Fol'klor [Ainu Folklore]. Moskva: Nauka, 1972. 175 p.
- OKLADNIKOV, A. P. Arheologicheskie dannye o drevneishej istorii Pribaikal'ja. [Archaeological data on the ancient history of the Lake Baikal area]. In *Vestnik Drevnej Istorii*, no 1 (2). Moskva: Rossijskaja Akademija Nauk, Institut Istorii, 1938, pp. 224-260.
- OKLADNIKOV, A. P., VASIL'EVSKIJ, R. S. Po Al'aske i Aleutskim ostrovam. [Travels in Alaska and the Aleut Islands]. Novosibirsk: Nauka, 1976. 168 p.
- TAMBOVTSEV, Y. (1994a). Dinamika funktsionirovanija fonem v zvukovyh tsepochkah jazykov razlichnogo stroja. [Dynamics of functional phonemes in phonetic languages of various types]. Novosibirsk: NGU, 1994. 133 p.
- TAMBOVTSEV, Y. (1994b). Tipologija uporjadochennosti zvukovyh tsepej v jazyke. [Typology of the order of phonetic types in language]. Novosibirsk: NGU, 1994. 199 p.
- TAMBOVTSEV, Y. (2001a). Kompendium osnovnyh statisticheskih harakteristik funktsionirovanija soglasnyh fonem v zvukovoj tsepochke anglijskogo, nemetskogo, frantsuzskogo i drugih indoevropejskih jazykov. [Compendium of the basic statistical characteristics of functioning of consonants in the speech chain of English, German, French and other Indo-European languages]. Novosibirsk: Novosibirskij klassicheskij institut, 2001. 129 p.
- TAMBOVTSEV, Y. (2001b). Funktsionirovanie soglasnyh fonem v zvukovoj tsepochke uralo-altajskih jazykov. [Functioning of consonants in the sound chain of Ural-Altaic languages]. Novosibirsk: Novosibirskij klassicheskij institut, 2001. 132 p.
- TAMBOVTSEV, Y. (2001c) Nekotorye teoreticheskie polozhenija tipologii uporjadochennosti fonem v zvukovoj tsepochke jazyka i kompendium statisticheskih harakteristik osnovnyh grupp soglasnyh fonem. [Some theoretical fundamentals of the typology of orderliness of phonemes in the sound chain of language and the compendium of statistical characteristics

- of the basic groups of consonants]. Novosibirsk: Novosibirskij klassicheskij institut, 2001. 130 p.
- TAMBOVTSEV, Y. (2001d). The phonological distances between Mongolian and Turkic languages based on typological consonantal features. In *Mongolian Studies: Journal of the Mongolia Society*. USA: The Mongolia Society, 2001, vol. 24, pp. 41-84.
- TAMBOVTSEV, Y. (2001e). Perednejazychnye soglasnye kak pokazatel' odnoj iz lingvisticheskih universalij jazykov mira [Front consonants as an indicator for one of the universals of world languages]. In Sibirskij Lingvisticheskij Seminar (Novosibirsk), no. 2 (2). Novosibirsk, 2001. pp. 18-27.
- TAMBOVTSEV, Y. (2002a). Comparative typological study of language distances based on the consonants in sound chains of various languages. In The 5th National Colloquium for Computational Linguistics in the UK. In *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. (2002b) Is Kumandin a Turkic language? In *Dilbilim Araştırmaları*. Istanbul, 2002, pp. 63-104.
- TAMBOVTSEV, Y. (2002c). Korean and Japanese as Members of the Altaic Language Family. In *Abstracts. Permanent International Altaistic Conference 45th Meeting, Budapest, June 23-28, 2002.* Hungary: Research Group for Altaic Studies. Hungarian Academy of Sciences, 2002, pp. 13-14.
- TAMBOVTSEV, Y. (2002d). Structure of the frequency of occurrence of consonants in the speech sound chain as an indicator of the phonotypological closeness of languages. In ALL-ACH 2002. New Directions in Humanities Computing. The 14th Joint International Conference, University of Tübingen, 24-28 July, 2002. Conference Abstracts. Tübingen: Universität Tübingen, 2002, pp. 138-139.
- TAMBOVTSEV, Y. (2003a). Tipologija funktsionirovanija fonem v zvukovoj tsepochke indoevropejskih, paleoaziatskih, uralo-altajskih i drugih jazykov mira: kompaktnost' podgrupp, grupp, semej i drugih jazykovyh taksonov. [Typology of functioning of phonemes in sound chain of Indo-European, Palaeo-Asiatic, Ural-Altaic and other world languages: compactness of subgroups, groups, families and other language taxons]. Novosibirsk: Sibirskij nezavisimyj institut, 2003. 143 p.
- TAMBOVTSEV, Y. (2003b). The phonological similarity between Turkic languages based on some phonological features of consonants. In *Linguistic and Oriental Studies from Poznań*. Poznań: UAM, 2003, vol. 5, pp. 85-118.

- TAMBOVTSEV, Y. Uralic Language Taxon: Natural or Artificial? (Typological Compactness of Uralic Languages and other Language Taxons: Branches, Subgroups, Groups, Families and Superfamilies). In *Fenno-Ugristica*, no. 26. Tartu: University of Tartu, 2004, pp. 200-246.
- WHALEY, L. J. Introduction to Typology: The Unity and Diversity of Language. London: SAGE, 1997. 321 p.
- ZAGORUIKO, N. G. Metody raspoznavanija i ih primenenie. [The Methods of Pattern Recognition and Their Application]. Moskva: Sovetskoe Radio, 1972. 206 p.
- ZAGORUIKO, N. G. Prikladnye metody analiza dannyh i znanij. [Applied Methods of Data and Knowledge Analysis]. Novosibirsk: Institute of Mathematics SORAN, 1999. 269 p.
- ZIPF, G. K. *The Psycho-Biology of Language. An Introduction to Psycho-Biology of Language.* Cambridge: Massachusetts Institute of Technology Press, 1935. 336 p.

APPENDIX

EUCLIDEAN DISTANCES between Ainu and other world languages, united in different genetic families.

Tab. 1 Phonostatistical EUCLIDEAN DISTANCES between Ainu and Quechuan family of American Indian languages. The mean of the distances -6.42.

#	Language	Distance
1.	Ainu	0
2.	Quechua	5.451
3.	Inga	7.388
	Mean	6.42

Tab. 2 Phonostatistical EUCLIDEAN DISTANCES between Ainu and Mayan family of American Indian languages. The mean of the distances – 12.03.

	Language	Distance
1.	Ainu	0
2.	Tz'utujil	9.922
3.	Pocomchi'	10.848
4.	Acateco	11.725

	Mean	12.03
7.	Quichean	14.295
6.	Kekchí	13.374
5.	Jacalteco	12.025

Tab. 3 Phonostatistical EUCLIDEAN DISTANCES between Ainu and Algic family of American Indian languages. The mean of the distances – 15.71.

#	Language	Distance
1.	Ainu	0
2.	Sweet Grass Cree	14.258
3.	Ojibwa	17.71
	Mean	15.71

Tab. 4
Phonostatistical EUCLIDEAN DISTANCES between Ainu and Tucanoan family of American Indian languages. The mean of the distances – 16.51.

#	Language	Distance
1.	Ainu	0
2.	Secoya	16.884
3.	Siriano	16.126
	Mean	16.51

Tab. 5
Phonostatistical EUCLIDEAN DISTANCES between Ainu and Tupi-Guarani family of American Indian languages. The mean of the distances – 16.88.

#	Language	Distance
1.	Ainu	0
2.	Guarani	15.115
3.	Kaiwá	18.640
	Mean	16.88

Tab. 6
Phonostatistical EUCLIDEAN DISTANCES between Ainu and Caddoan family of American Indian languages. The mean of the distances – 19.70.

#	Language	Distance
1.	Ainu	0
2.	Arikara	12.030
3.	Oneida	21.841
4.	Wichita	25.224
	Mean	19.70

Tab. 7
Phonostatistical Euclidean distances between Ainu and Finno-Ugric languages.
The mean of the distances – 16.95.

#	Language	Distance
1.	Ainu	0
2.	Khanty (Kazym)	11.783
3.	Mansi (Nothern-T)	12.635
4.	Karelian (Tihvin)	12.669
5.	Karelian (Livvik)	14.876
6.	Finnish	15.188
7.	Mansi (Konda)	16.113
8.	Khanty (Eastern)	16.271
9.	Komi-Zyrian	16.331
10.	Vod	17.001
11.	Saami (Lapp)	17.283
12.	Estonian	17.488
13.	Karelian (Ludian)	17.599
14.	Udmurt	18.219
15.	Mari (Gorno)	18.417
16.	Hungarian	18.550
17.	Vepsian	19.092
18.	Mari (Lugovo)	20.331
19.	Mordvin (Erzya)	21.066
20.	Mordvin (Moksha)	21.071
	Mean	16.95

Tab. 8
Phonostatistical EUCLIDEAN DISTANCES between Ainu and Samoyedic Languages. The mean of the distances – 16.31

#	Language	Distance
1.	Ainu	0
2.	Nganasan	15.083
3.	Nenets (Tambov)	15.104
4.	Selkup	15.253
5.	Kamas	19.808
	Mean	16.31

Tab. 9
Phonostatistical EUCLIDEAN DISTANCES between Ainu and Turkic Languages. The mean of the distances – 18.08

#	Language	Distance
1.	Ainu	0
2.	Tatar (Baraba)	14.124
3.	Yakut	14.440
4.	Salar	14.671
5.	Uyghur	15.240
6.	Karakalpak	15.691
7.	Bashkir	15.768
8.	Turkish	16.286
9.	Sary-Uyghur	17.194
10.	Tatar (Crimean)	17.294
11.	Tatar (Kazan)	17.472
12.	Orkhon-Yeniseian	17.624
13.	Dolgan	17.728
14.	Tatar (Chulym)	17.748
15.	Chuvash	18.028
16.	Uzbek	18.644
17.	Karachay	18.931
18.	Khakas	18.952
19.	Turkmen	18.983
20.	Kazakh	19.012
21.	Tofalar	19.462
22.	Shor	19.537

	Mean	18.08
28.	Altai (Kizhi)	23.984
27.	Azerbaijani	20.911
26.	Kirghiz	20.410
25.	Altai (Chalkan)	20.200
24.	Kumandin	20.050
23.	Tuvin	19.821

Tab. 10
Phonostatistical EUCLIDEAN DISTANCES
between Ainu and the Mongolian Family. The mean of the distances – 23.66

#	Language	Distance
1.	Ainu	0
2.	Kalmyk	22.982
3.	Mongolian	23.690
4.	Buriat	24.293
	Mean	23.66

Tab. 11 Phonostatistical Euclidean distances between Ainu and the Manchu-Tungus Family. The mean of the distances – 16.20

#	Language	Distance	
1.	Ainu	0	
2.	Ulch	13.464	
3.	Even (Lamut)	13.616	
4.	Evenki (Tungus)	13.835	
5.	Negidal	14.814	
6.	Orok	14.759	
7.	Udihe	15.466	
8.	Nanai	16.265	
9.	Oroch	20.787	
10.	Manchu	22.816	
	Mean	16.20	

Tab. 12 Phonostatistical Euclidean Distances between Ainu and the Yeniseian Family.

#	Language	Distance
1.	Ainu	0
2.	Ket (Yug)	20.245

Tab. 13
Phonostatistical EUCLIDEAN DISTANCES between Ainu and the Palaeo-Asiatic Family. The mean of the distances – 15.41

#	Language	Distance
1.	Ainu	0
2.	Chukchi	10.954
3.	Koryak	12.781
4.	Eskimo (Imaklin)	15.007
5.	Eskimo (Naukan)	16.149
6.	Itelmen	22.165
	Mean	15.41

Tab. 14 Phonostatistical Euclidean Distances between Ainu and Sino-Tibetan family. The mean of the distances -18.06

#	Language	Distance
1.	Ainu	0
2.	Burmese	9.144
3.	Thai	10.001
4.	Tibetan	22.441
5.	Chinese	23.764
6.	Dungan	24.947
	Mean	18.06

Tab. 15
Phonostatistical Euclidean Distances between Ainu and Bantu family. The mean of the distances – 12.40

#	Language	Distance
1.	Ainu	0
2.	Swahili	11.886
3.	Hanga	12.914
	Mean	12.40

Tab. 16 Phonostatistical EUCLIDEAN DISTANCES between Ainu and Afro-Asiatic family. The mean of the distances -16.72

#	Language	Distance
1.	Ainu	0
2.	Hausa	12.555
3.	Somali	14.788
4.	Assyrian	15.246
5.	Aramaic	17.371
6.	Soqotri	18.868
7.	Hebrew	18.920
8.	Arabic	19.289
	Mean	16.72

Tab. 17 Phonostatistical EUCLIDEAN DISTANCES between Ainu and Austronesian family. The mean of the distances -15.39

#	Language	Distance
1.	Ainu	0
2.	Tagalog	9.310
3.	Maori	10.513
4.	Indonesian	12.208
5.	Hawaiian	16.933
6.	Dayak	17.528

7.	Samoan	19.686
8.	Marquesan	21.568
	Mean	15.39

Tab. 18 Phonostatistical EUCLIDEAN DISTANCES between Ainu and Caucasian family. The mean of the distances – 22.64

#	Language	Distance
1.	Ainu	0
2.	Abkhaz	18.491
3.	Georgian	19.576
4.	Chechen	19.807
5.	Avar	22.325
6.	Kabardian	25.080
7.	Adyghe	30.546
	Mean	22.64

Tab. 19 Phonostatistical EUCLIDEAN DISTANCES between Ainu and Indo-European family. The mean of the distances – 20.12

#	Language	Distance
1.	Ainu	0
2.	Hindi	12.522
3.	French	14.767
4.	Serbo-Croatian	15.082
5.	Gujarati	15.280
6.	Marathi	15.309
7.	Sanskrit	15.317
8.	Armenian	15.681
9.	Prakrit	15.972
10.	Bengali	16.332
11.	Norwegian	16.744
12.	Romani (Gypsy)	16.815
13.	Spanish	16.899
14.	Moldavian	17.342
15.	Vedic	17.484

16.	Slovak	17.492
17.	Luzhits-Sorbian	18.313
18.	Persian	18.990
19.	Italian	19.375
20.	Russian	20.089
21.	Old Russian	20.622
22.	Bulgarian	20.712
23.	Osetin	20.818
24.	Czech	20.883
25.	Slovenian	20.925
26.	Latvian	21.136
27.	Gilani	21.189
28.	Lithuanian	21.351
29.	Portuguese	21.475
30.	Polish	21.614
31.	Belorusa	21.805
32.	Tajiki	22.379
33.	Swedish	22.535
34.	Dari	22.805
35.	Romanian	22.870
36.	Macedonian	23.062
37.	Kurdish	23.093
38.	Pashto	23.355
39.	Talysh	23.367
40.	Dutch	23.756
41.	Middle Persian (poetry by Firdousi)	25.432
42.	Ukrainian	25.572
43.	English	27.834
44.	German	30.595
	Mean	20.12

Tab. 20 Phonostatistical EUCLIDEAN DISTANCES between Ainu and Some Dead Indo-European languages. The mean of the distances – 17.31

#	Language	Distance
1.	Ainu	0

	Mean	17.31	
3. Gothic	15.968		
2.	Old English	18.654	

Tab. 21 Phonostatistical EUCLIDEAN DISTANCES between Ainu and Celtic family

#	Language	Distance
1.	Ainu	0
2.	Irish	17.272

Tab. 22 Phonostatistical EUCLIDEAN DISTANCES between Ainu and Italic family

#	Language	Distance	
1.	Ainu	0	
2.	Latin	18.803	

Tab. 23 Phonostatistical EUCLIDEAN DISTANCES between Ainu and Australian Aboriginal family. The mean of the distances -21.19

#	Language	Distance
1.	Ainu	0
2.	Ngaanytjaara	15.306
3.	Djingili	21.330
4.	Nyangumarta	20.931
5.	Nunggubuyu	21.627
6.	Mangarayi	21.976
7.	Ngandi	25.998
	Mean	21.19

Tab. 24 Phonostatistical Euclidean distances between Ainu and Isolated Languages

#	Language	Distance
1.	Ainu	0
2.	Albanian	21.334
3.	Basque	20.169
4.	Korean	12.636
5.	Japanese	15.272
6.	Yukaghir	15.304
7.	Nivkhi	22.679

Tab. 25 Phonostatistical EUCLIDEAN DISTANCES between Ainu and the Artificial Language - Esperanto.

#	Language	Distance
1.	Ainu	0
2.	Esperanto	17.09

Tab. 26
The Ordered Series of the Mean Phonostatistical Distances between Ainu and the Language Families.

#	Language Family	Mean Distance
1.	Quechuan (American Indian)	6.42
2.	Mayan (American Indian)	12.03
3.	Austronesian	15.39
4.	Palaeo-Asiatic	15.41
5.	Algic (American Indian)	15.71
6.	Manchu-Tungus	16.20
7.	Samoyedic	16.31
8.	Tucanoan (American Indian)	16.51
9.	Afro-Asiatic	16.72
10.	Tupi-Guarani (American Indian)	16.88
11.	Finno-Ugric	16.95
12.	Sino-Tibetan	18.06
13.	Turkic	18.08

14.	Caddoan (American Indian)	19.70
15.	Indo-European	20.12
16.	Yeniseian	20.24
17.	Australian Aboriginal	21.19
18.	Caucasian	22.64
19.	Mongolic	23.66