

Yuri Tambovtsev

Novosibirsk Pedagogical University, Novosibirsk

**Distances between Polish and other Slavonic languages:
A phono-typological comparison**

Abstract. The Slavonic languages have customarily been taken as a reference point for Polish. Measuring the phono-typological distance between Polish and the other Slavonic languages allows us to determine whether classifying Polish in this language group is sound. The distances are derived by measuring distances in multi-dimensional space; in this case, the number of dimensions is nine, as we take nine features for measure and comparison.

A new outlook on language classification is proposed here. The new method is based on both phonostatistical and metric analysis. It is founded on the structure of the frequency of occurrence of consonants in the speech sound chain. This is a good clue for defining the typological closeness of languages. It allows a linguist to find the typological distances between a language (in this case - Polish) and the other languages of different genetic groups of a language family (in this case - the Indo-European language family). This method can assign any language to a language taxon, i.e. a sub-group, a group or a family. The minimum distance may be a good clue for placing a language in this or that language taxon. This method of calculating distances between Polish and some other languages is used for the first time, and verifies that Polish belongs to the Western subgroup of the Slavonic group of Indo-European languages.

Key words: phonological, similarity and distance, articulatory, features, typology, frequency of occurrence, consonants, vowels, speech sound chain, statistics, artificial, natural, language similarity, comparative method, distances, closeness, language taxon, taxa of languages, classification.

Introduction

The goal of this paper is to measure the phono-typological distances between Polish and some Slavonic languages. In this case, these distances are calculated on the basis of the mathematical statistical criterion called Chi-criterion. In fact, the sound picture of Polish is compared to several Slavonic languages in order to understand how typical it is. In a language group some languages are more typical and some are less typical (Tambovtsev, 1986). The method which we developed investigates the distances between languages on phonetic level. The value of the distances indicates the degree of similarity of languages. The more similar the languages, the less the distance. And the more similar the languages in a taxon, the greater its compactness (Tambovtsev, 2004). A great value of dispersion in a language taxon means that the languages in it are not similar (Tambovtsev, 2006). The sound picture of the language is the total of the frequency of occurrence of the elements of the speech sound chain in this language.

We computed the frequency of occurrence of speech sounds in some Slavonic (Russian, Belarusian, Serbo-Croatian, Sorbian, Macedonian, Slovene) languages, while the data on the other languages were taken from publications (see the details in Tambovtsev, 2001). Thus, the data on the frequency of occurrence of the Polish phonemes in the text were taken from Segal (1972: 239 – 248). He calculated the frequency of occurrence of Polish phonemes in the texts of several Polish writers: 1) Ja.Ivashkevich “A girl and pigeons”; 2) L.Kruchkowski “The first day of freedom”; 3) S.Mrozhek “Goral”, “Interval”, “Nadzeja”, “Malyprzyjciel”, “Jak waiczylem”; 4) Ju.Shanywski “Profesor Tutka o zlodzieju”, “O slowie drukowanym”, “Sprawa Osobista”, “O przozolachi miodzie”, “O dwoch malowidlach”. The total Polish sample comprises 104,603 phonemes (Segal, 1972: 138), large enough for us to believe that the Polish data are reliable. Frequency data are as represented in Table 1.

Table 1

The Frequency of Occurrence of Consonants and Vowels in the Speech Sound Chain of Slavonic Languages, by per cent.

№	Phoneme class	Polish	Sorbn	Slov	Czec	Rusn	Ukrn	Belrsn	SrbCr	Blgrn	Slven	Macdn	OldRsn
1.	Labial	16.66	14.83	12.79	13.57	12.63	13.01	14.45	11.96	12.91	12.54	11.67	13.19
2.	Apical	27.42	32.12	30.03	32.16	35.18	34.60	33.84	31.75	30.24	35.53	36.95	33.30
3.	Palatal	9.31	6.47	7.21	6.20	4.27	3.91	4.63	5.53	1.45	4.40	0.89	2.96
4.	Guttural	5.75	6.68	6.39	6.87	5.74	6.28	7.12	5.63	6.68	6.28	6.20	5.98
5.	Sonorant	21.10	27.12	20.25	21.39	23.07	21.42	20.70	25.06	19.54	24.04	18.68	19.44
6.	Occlusive	21.16	18.79	20.82	20.47	20.14	21.18	21.91	20.37	14.96	21.08	24.12	20.17
7.	Fricative	16.88	14.19	15.35	16.94	14.61	15.20	17.43	9.44	16.78	13.63	12.91	15.82
8.	Voiced	13.90	11.51	9.69	13.28	11.69	15.19	12.53	9.89	12.51	13.72	13.27	11.72
9.	Vowels	40.86	39.90	43.58	41.20	42.18	57.80	39.96	45.13	48.72	41.25	44.29	15.60

Sample sizes of the Slavonic language objects (dialects and languages) under investigation in phonemes: Polish – 104,603 (Segal, 1972); Sorbian – 93,110 (Tambovtsev, 2001); Slovak – 20,000 (Bosak, 1965); Czech – 186,641 (Ludvikova et al., 1967); Russian – 20,189,475 (Tambovtsev, 2001); Ukrainian – 300,000 (Perebejnos, 1965); Belarusian – 37,487 (Tambovtsev, 2001); Bulgarian – 18,240 (Marinova et al., 1964); Slovene – 108,363 (Tambovtsev); Macedonian – 62,795 (Tambovtsev, 2001); Old Russian – 52,712 (Tambovtsev).

The data on the occurrence of the groups of consonants and vowels in different world languages are not the same as in Polish. The data on the frequency of occurrence of Lithuanian were obtained by B.I. Svecevichus and A. Girdenis. Their material of folklore and dialogues and the phonemic frequency of occurrence united in nine different groups are discussed elsewhere (Tambovtsev, 2001: 67 – 69). The Latvian data were obtained by Z. Baikov, E. Kuzina, A. Lorenz, Z. Nesaule, E. Piel and T. Jakubaitis. We have combined all their data in one sample to obtain the sound picture of Latvian (see the details in Tambovtsev, 2001: 64 – 67).

The Old Russian text material for computing was taken from the book by A.N. Kozhin who transcribed the Old Russian texts by modern Russian letters. The texts of the tenth – thirteenth

centuries were computed. They included the texts of Russkaja Pravda, Povest' Vremennyh let, Kievskaja letopis' and Pouchenija Vladimira Monomaha (Kozhin, 1974).

All the languages under investigation were compared by nine features. These features were derived on the basis of articulation. First of all, all the speech sounds were classified into the class of vowels and the class of consonants. Then by the active organ of speech the consonants were divided into four features: 1) labial; 2) front; 3) palatal; 4) guttural. Then, by the manner of the articulation the consonants were divided into 5) sonorant; 6) occlusive; 7) fricative. By the work of the vowel cords we received one more feature, 8) voiced consonants. The occurrence of all the vowels in the speech sound chain gives the ninth feature: 9) vowels. In fact, these nine features are typologically universal and can be found in all the world languages. Therefore, the results of the comparison are quite commensurable (Tambovtsev, 2003-a; 2003-b). The investigation is thus complete and systemic from the point of view of phonetics (Table 2).

Table 2

The Frequency of Occurrence of Consonants and Vowels in the Speech Sound Chain of Some Baltic, Turkic and Finno- Ugric Languages, by per cent.

№	Phoneme class	Lithn	Latvn	AKizh	Jakut	Mrqui	Wicht	Haida	Ngaan	Mong	MokM
1.	Labial	13,63	10,86	5.98	6.10	7.80	2.67	1.70	8.42	7.52	10.96
2.	Apical	33.37	35.89	38.25	32.77	15.99	23.67	33.20	29.57	36.57	40.09
3.	Palatal	3.59	3.07	2.54	5.22	0.00	1.38	3.37	10.16	1.48	1.81
4.	Guttural	7.39	5.55	11.71	13.27	18.79	30.23	22.36	10.33	11.32	7.79
5.	Sonorant	19.28	19.48	23.63	24.29	9.41	15.28	20.91	35.58	23.56	20.35
6.	Occlusive	22.25	20.78	29.12	22.14	27.35	27.45	29.88	22.90	22.66	22.77
7.	Fricative	16.45	15.11	5.73	10.92	5.82	15.22	9.84	0.00	10.67	17.53
8.	Voiced	12.63	11.10	11.94	9.17	0.00	0.00	13.84	4.98	18.57	11.31
9.	Vowels	42.02	44.63	41.52	42.64	57.42	42.05	39.37	41.52	43.11	39.35

Sample sizes of the Baltic, Turkic and Finno-Ugric language objects (dialects and languages) under investigation in phonemes: Lithuanian (Lithn) – 100,000 (Svecevichus and Girdenis); Latvian (Latvn) – 100,000 (Baikov, Kuzina, Lorenz, Nesaule, Piel and Jakubaitis); Jakut – 236,245 (Tambovtsev, 2001); Altai-Kizhi (Akizh) 190,525 (Tambovtsev, 2001); Tatar-Baraba (Tbrb) – 67,569 (Tambovtsev, 2001); Wichita (Wicht – American Indian) – 10,316 (Tambovtsev, 2001); Haida (Haida – American Indian) – 3,708 (Tambovtsev, 2001); Ngaanyatjarra (Australian Aboriginal) – 21,259 (Tambovtsev, 2001); Mongolian – 87,625 (Tambovtsev, 2001); Moksha-Mordovian – 143,720 (Tambovtsev, 2001).

Labial. The data on the frequency of occurrence of the labial consonants (Tab.1) in the chosen Slavonic languages allow us to construct the following ordered series: 1) Macedonian – 11.67; 2) Serbo-Croatian – 11.96; 3) Bulgarian – 12.91; 4) Slovene – 12.54; 5) Russian – 12.63; 6) Slovak – 12.79; 7) Ukrainian – 13.01; 8) Old Russian – 13.19; 9) Czech – 13.57; 10) Belorussian – 14.45; 11) Sorbian – 14.83; 12) Polish – 16.66.

This indicates that Polish has the greatest concentration of the labial vowels among the Slavonic languages. The mean concentration of the labial consonants in Slavonic languages is 13.35. Polish uses more labial consonants in its speech sound chains than that. It is curious to note that the concentration of labials in Old Russian and Czech is around the mean value. So, they are in the centre of the Slavonic languages in the case of the use of labial consonants. It is possible to see this in the graph. It is necessary to emphasize that Polish is unique from the point of view of the occurrence of labial consonants in its speech sound chain.

Front. In order to understand the place of Polish according to the distribution of the front consonants, we must construct the following ordered series: 1) Polish – 27.42; 2) Slovak – 30.03; 3) Bulgarian – 30.24; 4) Serbo-Croatian – 31.75; 5) Sorbian – 32.12; 6) Czech – 32.16; 7) Old Russian – 33.30; 8) Belarusian – 33.84; 9) Ukrainian – 34.60; 10) Russian – 35.18; 11) Slovene – 35.53; 12) Macedonian – 36.95.

Polish is also original in the use of the front consonants. It does not use them as much as do the other Slavonic languages. Polish is at the edge of the Slavonic languages in this respect. The

mean value of the occurrence of the front consonants is 32.75. So, we see again that it is Old Russian and Czech. It is vividly seen on the graph.

Palatal. It is possible to see whether Polish lies at the centre of the cluster of the Slavonic languages by the distribution of the palatal consonants. The ordered series of the Slavonic languages with respect to the palatals: 1) Macedonian – 0.89; 2) Bulgarian – 1.45; 3) Old Russian – 2.96; 4) Ukrainian – 3.91; 5) Russian – 4.27; 6) Slovene – 4.40; 7) Belarusian – 4.63; 8) Serbo-Croatian – 5.53; 9) Czech – 6.20; 10) Slovak – 7.21; 11) Polish – 9.31; 12) Sorbian – 6.47. The mean value of the occurrence of the palatal consonants in the Slavonic languages is 4.77. Belarusian is close to it.

Guttural. The ordered series of the Slavonic languages by the guttural consonants in their speech sound chains:

1) Serbo-Croatian – 5.63; 2) Russian – 5.74; 3) Polish – 5.75; 4) Old Russian – 5.98; 5) Macedonian – 6.20; 6) Ukrainian – 6.28; 7) Slovene – 6.28; 8) Slovak – 6.39; 9) Bulgarian – 6.68; 10) Sorbian – 6.68; 11) Czech – 6.87; 12) Belarusian – 7.12.

The mean value of the guttural consonants is 6.30. Slovene is close to it. As usual Polish is far from the mean value.

Sonorant. The ordered series of the Slavonic languages by the distribution of the sonorant consonants is the following: 1) Macedonian – 18.68; 2) Old Russian – 19.44; 3) Bulgarian – 19.54; 4) Slovak – 20.25; 5) Belarusian – 20.70; 6) Polish – 21.10; 7) Czech – 21.39; 8) Ukrainian – 21.42; 9) Russian – 23.07; 10) Slovene – 24.04; 11) Serbo-Croatian – 25.06; 12) Sorbian – 27.12. The mean value of the sonorant consonants is 21.82. Ukrainian and Russian are close to the mean. Polish is rather far away from it.**

Occlusive. The ordered series of the Slavonic languages with respect to the distribution of the occlusive consonants: 1) Bulgarian – 14.96; 2) Sorbian – 18.79; 3) Ukrainian – 20.14; 4) Old Russian – 20.17; 5) Serbo-Croatian – 20.37; 6) Czech – 20.47; 7) Slovak – 20.82; 8) Slovene – 21.08; 9) Polish – 21.16; 10) Ukrainian – 21.18; 11) Belarusian – 21.91; 12) Macedonian – 24.12. The mean value of the occlusive consonants is 20.43, which fall close to Czech.

Fricative. The ordered series of the Slavonic languages with respect to the distribution of the fricative consonants: 1) Serbo-Croatian – 9.44; 2) Macedonian – 12.91; 3) Slovene – 13.63; 4) Sorbian – 14.19; 5) Russian – 14.61; 6) Ukrainian – 15.20; 7) Slovak – 15.35; 8) Old Russian – 15.82; 9) Bulgarian – 16.78; 10) Polish – 16.88; 11) Czech – 16.94; 12) Belarusian – 17.43. The mean fricative is 14.93. Thus, Russian and Ukrainian are closer to the mean than Polish.

Voiced. The ordered series of the Slavonic languages with respect to the distribution of the voiced consonants: 1) Slovak – 9.69; 2) Serbo-Croatian – 9.89; 3) Sorbian – 11.51; 4) Russian – 11.69; 5) Old Russian – 11.73; 6) Bulgarian – 12.51; 7) Belarusian – 12.53; 8) Macedonian – 13.27; 9) Czech – 13.28; 10) Slovene – 13.72; 11) Polish – 13.90; 12) Ukrainian - 15.19. The maximum of the voiced Ukrainian consonants may be explained by the fact that there is no devoicing at the end of the word in Ukrainian. The mean voiced is 12.41, which the value of Bulgarian is close to. Again, Polish is far away from the Slavonic mean.

Having analysed eight consonantal and one vowel feature we can come to the conclusion that Polish is quite a unique language which cannot be called a typical Slavonic language. The graphs clearly show it.

Using the TMB coefficient (Tambovtsev's coefficient) which is the ratio of the obtained value of the Chi-square criterion to its theoretical value taking into account the level of statistical significance and the number of the degrees of freedom (Tambovtsev, 2003-a; 2003-b), we

calculated the phono-typological distance between Polish and the other Slavonic languages, especially Belarusian. It is also necessary to measure the typological distances between Polish and the language with which it was in the closest contact during its historical development, that is, Lithuanian, which is not a Slavonic but a Baltic language.

Distances between Polish and other languages

The ordered distance series (TMB coefficient) between Polish and the other Slavonic languages on the basis of the nine phonetic features is the following:

Polish – Czech (9.32)
 Polish – Slovak (11.99)
 Polish – Sorbian (15.60)
 Polish – Belarusian (17.11)
 Polish – Ukrainian (23.72)
 Polish – Russian (24.99)
 Polish – Slovene (25.46)
 Polish – Old Russian (30.54)
 Polish – Serbo-Croatian (34.81)
 Polish – Bulgarian (53.60)
 Polish – Macedonian (66.29)

Everything is seen in comparison, as the ancient Greeks used to say. Therefore, we must compare Polish not only to Slavonic languages, but also to some languages of the other language groups and families.

Polish – Lithuanian (Baltic) (24.62)
 Polish – Latvian (Baltic) (39.66)
 Polish – Moksha-Mordovian (59.90)
 Polish – Jakut (Turkic) (77.59)
 Polish – Kabardinian (Caucasian) (82.22)
 Polish – Mongolian (Mongolian) (94.03)
 Polish – Altai-Kizhi (127.57)
 Polish – Haida (American Indian) (185.15)
 Polish – Ngaanyatjarra (Australian Aboriginal) 186.99

Polish – Wichita (American Indian) (313.96)

Polish – Marquiz (Austronesian) 321.18

It is obvious that Polish is farther away from the non-Slavonic languages besides Lithuanian and Latvian, which belong to the Baltic group of the Indo-European language family.

Lithuanian and the other Baltic language – Latvian, may give a good clue to phonetic language closeness. The distance between Lithuanian and Latvian comprised 6.45. The distances between Lithuanian and some other languages may serve as a standard of the distance. Thus, we must calculate the phono-typological distances between Lithuanian and some other languages. It is a commonplace to say that Baltic languages (Lithuanian and Latvian) used to belong to the so-called Balto-Slavonic language community. First of all, it is advisable to verify how close the Lithuanian language is to the Slavonic languages, especially Old Russian. The typological distance between Polish and Lithuanian is 24.62. Is this distance big or small?

We received the following ordered series of the distances between Lithuanian and some other languages on the basis of nine phonetic features (TMB coefficient): 1) Belarusian – 1.92; 2) Old Russian – 2.84; 3) Ukrainian – 3.64; 4) Russian – 6.07; 5) Czech – 6.14; 6) Latvian – 6.45; 7) Slovenian – 7.46; 8) Slovak – 12.99; 9) Macedonian – 17.11; 10) Sorbian – 18.22; 11) Bulgarian – 19.64; 12) Polish – 24.62; 13) Serbo-Croatian – 25.66.

One can see that the phono-typological distance between Lithuanian and Latvian (6.45) is greater than between Lithuanian and some Slavonic languages (c.f. Belarusian, Old Russian, Ukrainian, Russian and Czech). The closeness between Lithuanian and Slavonic languages may be ascribed to the common Balto-Slavonic language unity. It is believed that Slavonic and Baltic languages have the same Balto-Slavonic proto-language. Some linguists think that the Baltic and Slavonic tribes lived together and had very similar languages. Their languages were so similar that they could understand each other all right (Ambrazas et. al., 1966: 502).

The greater closeness of Lithuanian to Belarusian may be explained by the intensive contacts between them in the frame of the ancient Belarusian-Lithuanian common state. Krivitskij A.A., Mihnevich A.E., Podluzhnyj A.I., and other Belarusian linguists underline that in this common Belarusian – Lithuanian kingdom all the books were written in the Belarusian language. The upper crust of the feudal society of this kingdom spoke Belarusian because it was fashionable. The Belarusian language was used as the only language of instruction at schools. Before that the Belarusian tribes assimilated the Letto-Lithuanian tribes of the Jatv'ags (Krivitskij et al., 1973: 6-7).

The distance measurements between Latvian and the 12 Slavonic languages allowed us to build the following ordered series by the value of the TMB coefficient:

Latvian – Lithuanian 6.45
 Latvian – Ukrainian 6.50
 Latvian – Slovene 8.83
 Latvian – 10.68
 Latvian – Macedonian 12.34
 Latvian – Czech 14.38
 Latvian – Bulgarian 15.89
 Latvian – Slovak 16.31
 Latvian – Sorbian 24.46
 Latvian – Polish 39.66

Distances from the theoretical phono-typological centre of the Slavonic languages to some Slavonic and Baltic languages

Obviously, the language which is in the centre of the cluster of Slavonic languages has the least distance to every Slavonic language. We can judge how close this or that Slavonic language is to the centre by the least mean distance between the language in question and the other Slavonic languages.

It is possible to find how close Polish is to the centre of the Slavonic cluster. This may show whether Polish is a typical Slavonic language. We must add all the distances between Polish and the other Slavonic languages. Then we must find the mean distance between Polish and the rest of the Slavonic languages. In this case, the mean is equal to 28.29. Let us consider the ordered series of the mean distances to the centre of the Slavonic cluster:

Russian 11.79
 Old Russian 12.77
 Slovene 12.81
 Czech 13.12
 Ukrainian 13.28
 Belarusian 13.83
 Lithuanian 14.35
 Slovak 17.56
 Sorbian 18.08
 Serbo-Croatian 23.76
 Polish 28.29
 Bulgarian 32.58
 Macedonian 34.74

Thus, one can see that Russian is in the centre of the Slavonic language group. It is not so for Polish. Bulgarian and Macedonian, however, are farther away from the centre.

The Polish mean distance indicates that Polish is not a typical Slavonic language by its sound picture. If we put it in the terms of the most and the least typical Slavonic language, then we must come to the conclusion that the most typical Slavonic language by its sound picture is Russian (11.79). One can see it on the graph. Lithuanian is close to the centre of the cluster of the Slavonic languages. This supports the idea of the Balto-Slavonic unity.

Let us consider the linguistic taxa of the Slavonic languages. They are divided into Eastern (Russian, Belarusian, Ukrainian), Western (Polish, Czech, Slovak, Sorbian) and Southern (Serbo-Croatian, Bulgarian, Macedonian, Slovene) groups. Therefore, it is interesting to check

into what group Polish belongs. We measure the distance of Polish to the members of these three taxa and measure its total. After that, we divide this total by the number of the members in the taxon. The total of distances between Polish and the other three members of the Western subgroup of the Slavonic group is 36.91. Thus, the mean distance is 12.30. In the same way we find the mean values between Polish and the Eastern subgroup (21.94) and the Southern subgroup (45.04). The least mean distance shows to which subgroup the Polish language belong. Obviously, it is the Western subgroup since it has the least total of the mean distances (12.30). Our typological investigation verifies the classical genetic taxa. Thus, one can see that measuring the typological distances helps to establish, verify, or improve the linguistic classifications (Tambovtsev, 2005; 2007; 2008-a; 2008-b; 2009).

Conclusions

According to the total of the distribution of certain phonemes in the speech sound, chain Polish is closest to Czech (9.32). The least mean distance (12.30) indicates that Polish belongs to the Western subgroup of the Slavonic group of the Indo-European language family. Having analysed eight consonantal and one vowel features we can come to the conclusion that Polish is a unique language which cannot be called typical of Slavonic languages.

References

- Ambrazas et. al., 1966 – Ambrazas V., Vajtk'avichute V., Val'atskene V., Morkunas K., Sabal'auskas A., Ul'vidas K. Litovskij jazyk. - In: Jazyki narodov SSSR. T.1. Indoevropskie jazyki. - Moskva: Nauka, 1966, p. 500 – 527.
- Bosak, 1965 – Bosak J. *Frequency of Phonemes and Letters in Slovak and Numerical Expression of Some Phonemic Relations.* - Jazykovedny casopis 16, 1965, p.120 – 135.
- Kozhin, 1974 – Kozhin A.N. Hrestomatija po istorii russkogo literaturnogo jazyka. - Moskva: Vysshaja shkola, 1974, p.12 – 45.
- Krivitskij et al., 1973 – Krivitskij A.A., Mihnevich A.E., Podluzhnyj A.I. Belorusskij jazyk dl'a nebelorussov. - Minsk: Vysshaja shkola, 1973.
- Ludvikova et al., 1967 – Ludvikova M., Konigova M. Quantitative Research of Graphemes and Phonemes in Czech. In *The Prague Bulletin of Mathematical Linguistics*, N7, 1967, p.15-29.
- Marinova et. al., 1964 – Marinova M., Marinov A. Statisticheskie issledovanija na fonemite v balgarskija knozhoven jezik. - Balgarski jezik, 1964, 2 – 3, p.173 – 79.
- Perebejnos, 1965 – Perebejnos V. S. Chastota i sochetaemost' fonem sovremennogo ukrainskogo jazyka. - Kiev: Znanie, 1965. – 38 pp.
- Segal, 1972 – Segal D. M. Osnovy fonologicheskoj statistiki. - Moskva: Nauka, 1972. – 254 pp.
- Tambovtsev, 1986 – Tambovtsev Yu.A. Review. Arakin V.D. Sopostavitel'naja tipologija skandinavskix jazykov. - M., 1984, 136p. - In: *Language*, V.62, N2, 1986, p.459.
- Tambovtsev, 2001 – Yuri Tambovtsev. Kompendium osnovnyh statisticheskikh harakteristik funkcionirovanija soglasnyh fonem v zvukovoj tsepochke anglijskogo, nemetskogo, frantsuzskogo I drugih indoevropskix jazykov. - Novosibirsk: Novosibirskij Klassicheskij Institut. – 129 pp.
- Tambovtsev, 2003-a — Tambovtsev, Yuri. Tipologija funkcionirovanija fonem v zvukovoj tsepochke indoevropskix, paleoaziatskix, uralo-altajskix i drugih jazykov mira: kompaktnost' podgrupp, grupp, semej i drugih jazykovykh taksonov. - Novosibirsk: Sobirskij Nezavisimyj Institut, 2003. – 143 pp.
- Tambovtsev, 2003-b – Tambovtsev, Yuri. Phonological similarity between Basque and other world languages based on the frequency of occurrence of certain typological consonantal features. In *The Prague Bulletin of Mathematical Linguistics*, # 79-80, (Praha), 2003, p. 121 – 126.

- Tambovtsev, 2004 – Tambovtsev, Yuri. Uralic Language Taxon: Natural or Artificial? (Typological Compactness of Uralic Languages and other Language Taxons: Branches, Subgroups, Groups, Families and Superfamilies). In *Fenno-Ugristica*, # 26, 2004 (Tartu, Estonia), p. 200 – 233.
- Tambovtsev, 2005 - Tambovtsev, Yuri. Language Taxons and the Naturalness of their Classification. In *California Linguistic Notes* (USA), Volume XXX, # 2, Fall 2005, p. 1 – 25.
- Tambovtsev, 2006 - Tambovtsev, Yuri. Degrees of Dispersion of Various Language Taxa as the Index of Their Typological Similarity. In *Favete Linguis. Studies in Honour of Viktor Krupa*. Bratislava: Institute of Oriental Studies of Slovak Academy of Sciences, 2006, p. 219 – 260.
- Tambovtsev, 2007 - Tambovtsev, Yuri. How can typological distances between Latin and some Indo-European language taxa improve its classification? In *The Prague Bulletin of Mathematical Linguistics*, Number 88, December, 2007, p. 73 – 90.
- Tambovtsev, 2008-a — Tambovtsev, Yuri. Typological similarity between Ainu and other languages on the phonological level as a possible indicator of its genetic origin. In *Linguistic and Oriental Studies from Poznan'*, Volume 8, 2008, p. 81 – 96.
- Tambovtsev, 2008-b – Tambovtsev, Yuri. Typological Distance Between Korean and other Languages of Asia. In *Lingua Posnaniensis*, # 50 (L), 2008, p. 177 – 186.
- Tambovtsev, 2009 – Tambovtsev, Yuri. How Should the Hungarian Language Be Correctly Classified Within the Ugric Language Taxon? In *Asian and African Studies*, #18, 2009, p.103 – 137.