

# NORTH-CAUCASIAN LANGUAGES: MEASURES OF SIMILARITY<sup>86</sup>

Valery Solovyev<sup>87</sup>

Ewgeniya Bulygina<sup>88</sup>

Venera Bayrasheva<sup>89</sup>

## KEYWORDS

*Evolution, genetics, linguistic databases, peoples of Caucasus, similarity*

Studying the process of mankind settlement over the planet needs the knowledge from different fields of science. New big databases, such as grammar database “Jazyki mira” [4] and lexical database ASJP [3] let us study some issues more thoroughly than it was possible before. The paper suggests the comparison of linguistic and genetic data, as well as the data on geographical area of peoples’ living. The territory of Caucasus was chosen for the research as the previous studies of languages and peoples of this area turned out to be contradictory.

For the study we took the following languages: Abkhaz, Adyghe, Kabardian, Chechen, Ingush, Avar, Andi, Chamalal, Dargwa, Lezgian, and Tabassaran, that represent all the main groups of languages of the Northern Caucasus, namely Abkhaz-Adyghe, Nakh, and Daghestanian. Genetic data (the frequencies of haplogroups of Y-chromosome) were taken from O. Balanovsky database [1].

We formed five matrixes of distances between languages/peoples: genetic, lexical, grammatical, geographical and genealogical. The latter shows the affinity of the languages, which was established in historical linguistics, and is formed on the genealogy tree in the following way: a certain rank is given to every top of the tree: 0 goes to leaves and the highest rank of descendants greater by one is given to the other tops. Then

---

<sup>86</sup> © Valery Solovyev, Ewgeniya Bulygina, Venera Bayrasheva, 2011

<sup>87</sup> Kazan Federal University, Russia, [maki.solovyev@mail.ru](mailto:maki.solovyev@mail.ru)

<sup>88</sup> Kazan Federal University, Russia

<sup>89</sup> Kazan Federal University, Russia

distances between languages in the genealogical tree are defined as the rank of their closest common ancestor.

The Mantel correlation coefficient [2] was calculated for every pair of matrixes. The results are given in Table 1 (the upper part). The validity coefficients are given in the lower part of the table. If the validity coefficient is lower than 0,001, the result is statistically significant.

**Table 1. Mantel correlation coefficient**

	Genealogy	Geography	Genetics	Lexis	Grammar
Genealogy	–		0,726926	0,418453	0,516739
Geography	= 0,000	–	0.726926	0.418453	0.516739
Genetics	= 0,000	= 0,000	–	0.402573	0.471359
Lexis	<0,001	<0,001	<0,001	–	0.386565
Grammar	0,0059	0,0065	0,0023	< 0,001	–

The data from Table 1 show that genealogic, genetic and geographical distances strictly correlate with each other. Lexical distances and the other ones are also correlated. The obtained data prove co-evolution of languages and genes in the Northern Caucasus and the absence of considerable peoples' resettlement in the region as well. Grammar correlates with the other parameters in a lesser degree, probably, because of the greater dependence on borrowings as the result of cultural influence.

The scale of the region under study was varied during the research. The relations were analyzed more carefully for separate subgroups of Northern-Caucasus languages. The influence of additional geographical factor on the degree of similarity, i.e. the existence of the

common boundary between regions of peoples living, has been found. In general, correlations that have been described above remain.

We study separately the region of the whole Caucasus together with the Southern part. Several relative languages from other parts of Asia were taken for comparison. We used methods of multidimensional scaling and Principal component analysis in order to check the results obtained and for the purpose of visualization. The example of representing languages in the space of 3 main components on the basis of grammatical data is given in Figure 1. It displays the existence of a certain correlation between grammar, geography and genealogy.

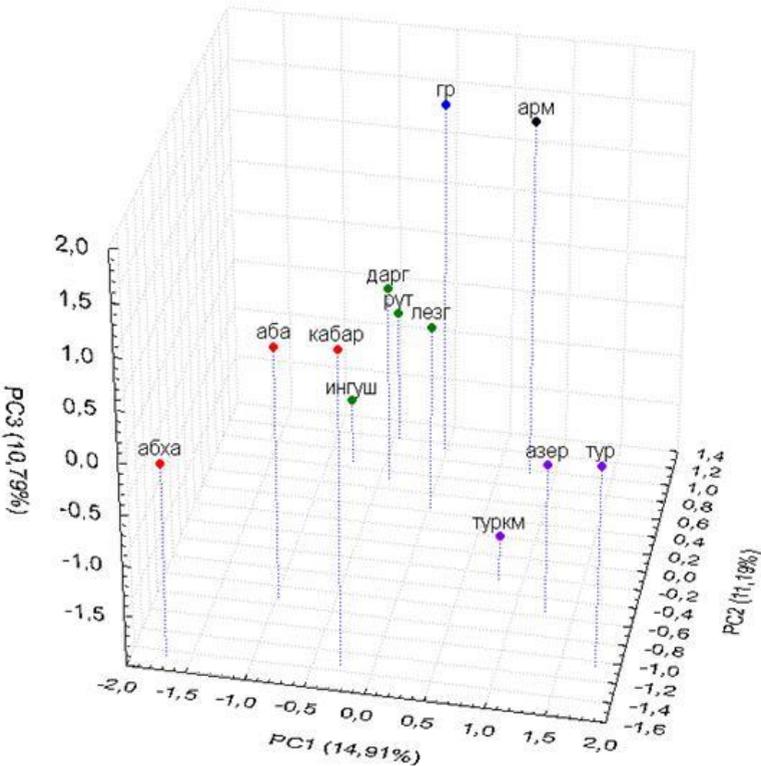


Fig. 1. Grammar Similarity of Languages in the Space of 3 principal components

## REFERENCES

- Balanovsky O.P., Psbenichnov A.S., Sychyov R.S., Balanovskaya Ye V.* Y-base: chastoty haplogrup Y khromosomy u narodov mira. 2010, <http://www.genofond.ru>.
- Mantel N.* The detection of disease clustering and a generalized regression approach. *Cancer Research*. V. 27 (2), 1967. p. 209–220.
- Müller A., Wichmann S., Velupillai V., Brown C.H., Brown P., Sauppe S., Holman E.W., Bakker D., List J.-M., Egorov D., Belyaev O., Mailhammer R., Urban M., Geyer H., Grant A.* ASJP World Language Tree of Lexical Similarity: Version 3. 2010. <http://email.eva.mpg.de/~wichmann/>
- Polyakov V.N., Solovyev V.D.* Kompyuternye modeli v tipologii i komparativistike. Kazan: KSU, 2006.