Frequency-Based Relevant Grammar Features of the Caucasian Languages

Veronika Yurievna Danilova¹, Elena Andreevna Makarova², Vladimir Nikolaevich Polyakov³* and Valery Dmitrievich Solovyev⁴

¹Department of Theoretical and Applied Linguistics, Philological Faculty, Moscow State University, Moscow, Russia; verniknet@list.ru
²Institute of Linguistics, Russian Academy of Sciences, Moscow, Russia
³NUST, MISIS, Moscow, Russia; pvn-65@mail.ru
⁴Kazan Federal University, Kazan, Russia

Abstract

Background/Objectives: The article regards the areal community of the Caucasian languages aiming to reveal relevant features of each family and suggest hypotheses on the development of separate representatives of each family. Methods/Statistical Analysis: The main tool for the research is the database “Languages of the World” of Institute of Linguistics of Russian Academy of Sciences. A relatively new method of feature detection built on contrast queries was applied. Particularly, the study separately compared the features of each group of the Caucasian languages to the Altaic language family and revealed a set of features characteristic only of the language family under study. Findings: The languages from the areal community of the Caucasian languages were regarded from the point of view of their relevant features – features that occur in most languages of a family under study, but are very rare in other families. For each group of the Caucasian languages a core of relevant features was found. Based on these results, as well as on the results of the applied variation method, we made an attempt to trace back the structural evolution of the Caucasian languages. Moreover, the variation method showed that the relevant features of a language family are not necessarily genealogic, i.e. they could be absent in the parent languages. Several hypotheses on the development of separate Caucasian languages from the groups of languages were suggested. Applications/Improvements: The research is a good basis for further inquiries on the development of the Caucasian languages. Moreover, it presents an example of the method for contrast queries application in studying the evolution of language families.

Keywords: Contrast Query, Relevant Features, The Caucasian Languages

1. Introduction

The Caucasian languages are an areal community of languages of peoples who live in Caucasia¹. The genealogic nature of these languages is not very clear for the modern science that is why they are studied in the light of an areal community. Comparative and historical linguistics, as well as statistical linguistics, paid great attention to the issue of genealogy and areal contacts of the Caucasian languages, e.g. works of S.A. Starostin² and G. Silnitsky³. Nevertheless, the problem of detecting and studying the core of relevant grammar features of genealogic communities included in the area of the Caucasian languages, has not been studied thoroughly enough.

One of the aims of the present research is to study the structural evolution of linguistic communities of the Caucasian language area. Three aspects play the key role in modeling of the structural evolution of languages:

* Author for correspondence
genealogical tree of the languages, borrowing of structural features and development of new features.

The study uses the typological database "Languages of the World" of Institute of Linguistics of Russian Academy of Sciences (IL RAS) - one the biggest digital resources in the world, containing a unified description of language grammars.

2. Relevant Features

Relevant features occur in most languages of the group under study, but are very rare in languages of other groups. They are the features necessary for differentiation of linguistic units (families, genera). The definition of relevant features can be attributed not only to particular linguistic units, but also to the whole families of language. Thus, the relevant features of a language family are a set of features which are characteristic of most languages from this family and which occur as a whole set only in it.

The detection of relevant features is of great importance for the typology of languages, for ascertaining their genealogic structure and areal contacts. If we have a set of particular grammar features of languages from two families of languages and some general knowledge about typology and genealogic structure of the world languages, we can formulate and develop hypotheses on genealogic affinity of these families, possible contacts and borrowing (or losing) of features, which are not present (or were present) in the structure of the family’s parent language, etc., i.e. suggest hypotheses on the linguistic evolution.

The research presented in this paper is based on the most complete digital grammar information pool on the Caucasian languages from the database “Languages of the World” of IL RAS.

The aim of the present paper is to detect and to study the relevant features of three genealogic communities of the Caucasian languages: Northwest Caucasian, Northeast Caucasian and Kartvelian.

The study is based on the combination of genealogic, geographic and typological data on families of languages presented in the database "Languages of the World" of IL RAS. We used the quantitative method of typological and areal studies, considered the data on 38 Caucasian languages and 76 Altaic languages, analyzed over 3800 grammar features. For map creation we used Google Maps.

We collected information on the genealogic and areal relations of the Caucasian languages in order to formulate the hypotheses on the linguistic evolution of genealogic communities of the languages in question.

3. The Caucasian Languages: Genealogic Structure, Ancestral Home and Areal Contacts

The Caucasian languages are a geographical community of languages of Caucasia, which are not included in any of the known families of languages spoken beyond Caucasia. There are three families that are traditionally distinguished among the Caucasian languages: Northwest, Northeast and Kartvelian. At the present time comparative linguistics has several significant theories on the genealogic relations between the families of the Caucasian languages and on their external relations.

The North Caucasian hypothesis was first suggested in. Supposedly, the disintegration of the North Caucasian community took place in the middle of the 6th or early 5th cc. BC. The location of the ancestral home of the North Caucasian community is not known, there is a hypothesis on the South Caucasian/West Asian location. Studies on phylogenetic trees (ASJP11) did not cast light on this problem.

Another famous hypothesis is the Iberian-Caucasian hypothesis on the affinity of all three groups of the Caucasian languages first suggested in 1888 by the end of the 20th century most linguists rejected this point of view.

An important hypothesis on external genealogic connection of the Caucasian languages is a hypothesis on the Sino-Caucasian macro-family suggested by S. Starostin. East Asia is a probable location of the ancestral home of the Sino-Caucasian language community.

On the other hand, the Kartvelian languages are assigned to the Nostratic macro-family along with Altaic, Dravidian, Indo-European and sometimes also Afro-Asiatic and Eskimo-Aleut. S. A. Starostin defined the time of the decay of the Protonostratic language to have happened in 12-10 cc. BC the most probable ancestral home of the Nostratic languages are the forests of South Ural and neighboring regions (Volga Region and Western Siberia).

Another theory on the affinity of the Hurro-
Urartian and Northeast Caucasian languages (Alarodian hypothesis) belongs to I. M. Diakonov and S. A. Starostin17. The Ibero-Caucasian hypothesis was popular in the late 19 – early 20 cc. (H. Schuchardt, A. Trombetti, N. Ya. Marr, K. Boud, R. Lafont), but now it is considered inconsistent.

The data stated above suggest that it is convenient to study the North Caucasian and Kartvelian groups separately. Supposedly, the disintegration of the North Caucasian parent language into North Caucasian and East Caucasian took place around 3800 BC. The disintegration of the parent East Caucasian language occurred in the 26-24 cc. BC. The developing of Caucasia and Transcaucasia was going in two directions: through the territory of modern Georgia (Nakh and Avar-Andic-Tsezic) and Dagestan (the other languages). The disintegration of the parent West Caucasian language into Abkhaz-Abaza, Ubykh and Adyghe branches took place around 640 BC18. The collapse of Hattic and Urartian Empires brought about the development of Kartvels, who began to settle in the foothills of North Caucasus and Colchis. Around the 20-19 cc. BC the Kartvelian parent language divided into Svan and common Georgian-Mingrelian-Laz.

The problem of areal relations of the Caucasian languages remains the less studied in the modern linguistics. If we speak about the areal relations between the Caucasian language families, the following point of view is taken as granted: The Caucasian languages present a sole linguistic area.

The Georgian influence on the Northeast Caucasian languages began in the Middle Ages, during the establishment of Georgian state. The zone of historical contacts between the Kartvelian and Northeast Caucasian languages is located near the Northern and Eastern borders of Georgia. These contacts are mainly one-sided: Limited lexical borrowings into the East Georgian dialects compared to the serious influence of Georgian.

The interaction between the Kartvelian and Northwest Caucasian languages can be divided into three parts. The early period has not been studied, though there are hypotheses on Svan and Northeast Caucasian interaction on the territory of Abkhazia. The time of Georgian and Abkhaz contacts, which began before the 7th century is more definite. The results of intensive Abkhaz-Mingrelian interference are especially evident. As a result, there are numerous structural parallelisms due to the bilingualism of the adjoining zone.

Thus, the linguistic contacts that lasted for years resulted in an Intercaucasian lexical fund – a few dozens of common Caucasian words of the local origin.

A big layer of areal contacts of the Caucasian languages is made by their interaction with the ancient Indo-European languages19. There are several dozens of Old Persian-Asian isoglotic lines covering the Kartvelian and part of Northeast Caucasian language areas, but, apparently, laying aside the Northwest Caucasian area. The most perceptible contribution to the contents of the cultural vocabulary of the Kartvelian languages was made by Old Indo-European and Asian languages20. The direct contact of the Kartvelian languages (Colchian group and Georgian) and the Armenian languages in Transcaucasia, which lasted for some 2.5 thousand years, resulted in a vast structural adjustment. It embraced almost all sides of both components of comparison (lexicon, phonetics, morphology and even syntactic models). Another aspect of the Indo-European areal contacting is stipulated by the long-lasting contact of the Ossetian and Kartvelian languages (Georgian and Svan) of Northwest Caucasian (Adyghe)21 and Northeast Caucasian (little developed).

4. Contrast Queries for Detection of the Core of Relevant Grammar Features

Relevant grammar features are used for building of reliable hypothesis on the evolution of the language. The method of contrast queries suggested by V. N. Polyakov2 is based on the frequency of features in families/branches presented in the database. A good match of relevant features in two families means that they are genealogically kindred and vice versa, the mismatch guarantees their difference, i.e. this method can be used in disputable cases. It consists in the following: The user requests features, the frequency of which in the studied family/branch is over 0.5, for example and in the compared family, that is known to have no affinity with the first one, this frequency is less than 0.05. The frequency of occurrence is a relative number of the occurrence of a feature in a particular language community. The frequency can have values from 0 (does not occur in any language) to 1 (occurs in all languages of the community). Contrastive queries are performed using a special table, which includes frequencies of feature occurrence for all languages from the database “Languages of the World” of IL RAS.
Each group of the Caucasian languages is compared to the Altaic language family separately. Altaic languages were chosen as the most distant from the Caucasian languages, from the languages that are described well in the database “Languages of the World” of IL RAS. All results of the contrast queries were double-checked manually using the printed encyclopedia “Languages of the World.”

We received the following results:

• The Northwest Caucasian group (Table 1).
  
  Each line in the query shows one grammatical feature that is present in the contrast query to the stated subset of languages. As the database is organized in a hierarchal way, each line shows the grammatical feature at the end of the chain. The first feature is the name of the section (e.g., 2.5.3. Simple Sentence). Other features are shown for information on the branch of the tree the feature belongs to. The number in brackets means the inventory number of the feature in the database. Inventory numbers are assigned to a grammar feature when it is introduced in the database and does not change in new version. The vertical line (|) separates an inferior branch from the superior branch; the number of dots before the name of the feature means the level in the tree. An example of one line of the query is shown in Table 2.

Table 2. One line of the query divided into categories

| Feature (with complete path)                      | Inventory number of the feature | Relative frequency of occurrence among Altaic languages (76) | Relative frequency of occurrence among Northwest Caucasian languages (5) |
|--------------------------------------------------|---------------------------------|-------------------------------------------------------------|--------------------------------------------------------------------------|
| SIGN (1765) 2.5.3.SIMPLE SENTENCE                | 1776                            | 0.013                                                       | 1                                                                         |
| SIGN (1682) 2.5.1. WORD FORM STRUCTURE           | 1695                            | 0.013                                                       | 1                                                                         |
| SIGN (1484) 2.4.0.PARADIGMS                      | 1493                            | 0                                                           | 1                                                                         |
| SIGN (1290) 2.3.6.DEICTIC CATEGORIES              | 1371                            | 0.013                                                       | 1                                                                         |
| SIGN (959) 2.3.5.VERBAL CATEGORIES                | 1254                            | 0                                                           | 1                                                                         |
| SIGN (2) 2.1.1.PHONEMIC STRUCTURE                 | 156                             | 0.013                                                       | 1                                                                         |
| SIGN (959) 2.3.5.VERBAL CATEGORIES                | 1244                            | 0                                                           | 1                                                                         |

As shown in Table 1, there are 9 features found as a result of the query.

• The Kartvelian language group.
  
  In the following way we found 16 relevant features for the Kartvelian language group (Table 3).

• The Northeast Caucasian language family.
  
  The query detected 12 relevant features (Table 4).

5. Search for Query Variations

The query master of the database “Languages of the World” of IL RAS allows defining a set of features according to the presence (or absence) of which the languages that satisfy the search criteria can be found. The sets of features received in the section 4 were used in the further queries. The further study of the set of languages and relevant features consists in a sequential elimination of one of the features from the initial set. This method is called a variation method. It allows suggesting hypotheses on the development of the studied family/branch.
| Occurrence of feature in the whole database | Feature (with complete path) | Inventory number of the feature | Relative frequency of occurrence among Altaic languages (76 languages) | Relative frequency of occurrence among Kartvelian languages (4 languages) |
|------------------------------------------|-------------------------------------------------|-------------------------------|-------------------------------------------------|-------------------------------------------------|
| 22                                      | SIGN (2) 2.1.1.PHONEMIC STRUCTURE|(90) .CONSONANTS|(91) .NOISY|(92) ...LARYNGEAL FEATURES|(96) ...VOICED/VOICELESS/ABRUPTIVES | 96.039 | 96.039 |
| 37                                      | SIGN (409) 2.1.4.SYLLABLE|(411) .THE ELEMENT PRECEDING THE VOWEL|(416) .CONSISTS, AS A RULE, OF|(420) .FROM ONE TO THREE CONSONANTS | 420.026 | 420.026 |
| 15                                      | SIGN (959) 2.3.5.VERBAL CATEGORIES|(1244) .VERSION | 1244 | 0 | 1 |
| 5                                       | SIGN (959) 2.3.5.VERBAL CATEGORIES|(1244) .VERSION|(1245) .BASIC MEANINGS|(1251) .SUBJECTIVE/OBJECTIVE/NEUTRAL | 1251 | 0 | 1 |
| 13                                      | SIGN (959) 2.3.5.VERBAL CATEGORIES|(1244) .VERSION|(1253) .METHOD OF EXPRESSION|(1254) .MORPHOLOGICAL CATEGORY | 1254 | 0 | 1 |
| 24                                      | SIGN (1290) 2.3.6.DEICTIC CATEGORIES|(1352) .GRAMMATICAL CATEGORY OF PERSON|(1357) .IN VERBS|(1360) .TWO-PERSON CONJUGATION | 1360 | 0 | 0.75 |
| 26                                      | SIGN (1290) 2.3.6.DEICTIC CATEGORIES|(1369) .ORIENTATION OF ACTION IN SPACE|(1370) .EXPRESSION|(1371) .VERBAL PREVERBS | 1371 | 0 | 0.75 |
| 13                                      | SIGN (1484) 2.4.0.PARADIGMS|(1490).VERB|(1491) .FINITE FORMS|(1492) .INFLEXION|(1493) .VERSION | 1493 | 0 | 1 |
| 100                                     | SIGN (1682) 2.5.1. WORD FORM STRUCTURE|(1688) .METHODS OF INFLEXION|(1695) .PREFIXES | 1695 | 0.013 | 1 |
| 122                                     | SIGN (1735) 2.5.2.WORD DERIVATION|(1736) .DERIVATION|(1737) .AFFIXATION|(1739) .PREFIXES | 1739 | 0.013 | 1 |
| 7                                       | SIGN (1735) 2.5.2.WORD DERIVATION|(1736) .DERIVATION|(1737) .AFFIXATION|(1743) .CIRCUMFIXES | 1743 | 0 | 0.75 |
| 46                                      | SIGN (1765) 2.5.3.SIMPLE SENTENCE|(1810) .AGREEMENT|(1811) .ACTANT DEFINES THE FORM OF PREDICATE|(1818) .OBJECTIVE AGREEMENT | 1818 | 0.013 | 1 |
| 17                                      | SIGN (1765) 2.5.3.SIMPLE SENTENCE|(1810) .AGREEMENT|(1811) .ACTANT DEFINES THE FORM OF PREDICATE|(1818) .OBJECTIVE AGREEMENT|(1820) .BY PERSON | 1820 | 0.013 | 1 |
| 36                                      | SIGN (1765) 2.5.3.SIMPLE SENTENCE|(1810) .AGREEMENT|(1811) .ACTANT DEFINES THE FORM OF PREDICATE|(1818) .OBJECTIVE AGREEMENT|(1822) .BY NUMBER | 1822 | 0.013 | 0.75 |
5.1 The Northwest Caucasian Group

The result of the contrast query for the Northwest Caucasian group showed that the frequency of occurrence of all relevant features equal 1 (Table 1), which means that all five languages of the group – Abkhaz, Abaza, Adyghe, Kabardian and Ubykh – have all relevant features. Thus, the variation method is not applicable in this case.

5.2 The Kartvelian Group

Only one of four Kartvelian languages has all relevant features – Laz. It is not difficult to explain the fact that only one of four languages described in the database met the requirements. The frequency of occurrence was calculated for each feature separately, consequently, for one feature the frequency could be over 0.5 and more than two languages could possess this feature. But for different features the sets of languages that have them could intersect only partially. As a result, the final list of languages with a complete set of features included only 1 language – Laz.

The variation method helped us find three other languages of the Kartvelian group. Sequential elimination of relevant features showed that Mingrel does not have preverbs. Georgian lacks two relevant features: It can have more than three vowels before a consonant and it does not have objective agreement by number. And Svan is different from the other languages of the Kartvelian group in three features: Besides the possible number of consonants before a vowel being over three, it does not have two-person conjugation and lacks in circumfixes.

Table 4. List of features from the contrast query (Northeast Caucasian vs. Altaic)

| Occurrence of feature in the whole database | Feature (with complete path) | Inventory number of the feature | Relative frequency of occurrence among Altaic languages (76 languages) | Relative frequency of occurrence among Northeast Caucasian languages (28 languages) |
|---------------------------------------------|------------------------------|---------------------------------|-------------------------------------------------|-------------------------------------------------|
| 156                                         | SIGN (639) 2.3.2.SUBSTANTIVE CLASSIFICATIONS(641).CLASSES OF AGREEMENT | 641                             | 0,039                                           | 0.893                                           |
| 129                                         | SIGN (639) 2.3.2.SUBSTANTIVE CLASSIFICATIONS(641).CLASSES OF AGREEMENT(646).MEANS OF EXPRESSING CLASSES OF AGREEMENT(653) ...SYNTACTIC | 653                             | 0                                                | 0.893                                           |
| 70                                          | SIGN (639) 2.3.2.SUBSTANTIVE CLASSIFICATIONS(641).CLASSES OF AGREEMENT(646).MEANS OF EXPRESSING CLASSES OF AGREEMENT(653) ...SYNTACTIC(656) ...IN VERB | 656                             | 0                                                | 0.893                                           |
| 97                                          | SIGN (639) 2.3.2.SUBSTANTIVE CLASSIFICATIONS(641).CLASSES OF AGREEMENT(646).MEANS OF EXPRESSING CLASSES OF AGREEMENT(653) ...SYNTACTIC(661) ...IN ADJECTIVES | 661                             | 0                                                | 0.75                                            |
| 35                                          | SIGN (1484) 2.4.0.PARADIGMS(1490).VERB(1491).FINITE FORMS(1492).INFLEXION(1510) ...CLASS OF AGREEMENT | 1510                            | 0                                                | 0.893                                           |
| 23                                          | SIGN (1484) 2.4.0.PARADIGMS(1619).ADJECTIVE(1620).INFLEXION(1628) ...CLASS OF AGREEMENT | 1628                            | 0                                                | 0.786                                           |
| 29                                          | SIGN (1484) 2.4.0.PARADIGMS(1638).NOUN(1661).CLASSIFYING CATEGORIES(1665) ...CLASS OF AGREEMENT | 1665                            | 0                                                | 0.679                                           |
| 104                                         | SIGN (1682) 2.5.1.WORD FORM STRUCTURE(1688).METHODS OF INFLEXION(1695).PREFIXES | 1695                            | 0.013                                           | 0.964                                           |
| 42                                          | SIGN (1765) 2.5.3.SIMPLE SENTENCE(1766).STRUCTURE(1776).ERGATIVE | 1776                            | 0                                                | 0.964                                           |
5.3 The Northeast Caucasian Group

Sixteen of twenty-eight Northeast Caucasian languages have a complete set of relevant features: Andi, Archi, Bezhta, Botlikh, Godoberi, Hunzib, Dargwa, Karata, Tindi, Khwarshi, Chamalal, Akhvakh, Ingush, Chechen, Tsez, Avar.

The variation method helped us detect numerous deviations from the initial set of relevant features: We detected languages that do not have all relevant features:

- The noun is not inflected by the class of agreement in: Bats, Tsakhur, Hinuq, Lak and Tabasaran.
- Classes of agreement are not expressed in the adjective and consequently, the adjective is not inflected by the class of agreement, in: Kryts, Khinalug and Budukh.
- The following languages do not have classes of agreement at all: Aghul, Lezgian.
- Classes of agreement are not expressed in the adjective and in the noun in Rutul.

6. Visualization on a Map for Building of Hypotheses on the Borrowings of Grammar Features

The query to the database “Languages of the World” of IL RAS showed that all five Northwest Caucasian languages have a complete set of relevant features (Figure 1). Besides the name of the language, here and in below figures show the geographic coordinates of the administrative center of the linguistic area.

![Figure 1. The Northwest Caucasian languages (all have a complete set of relevant features).](image1)

The Kartvelian languages are shown in Figure 2. Laz, which has a complete set of relevant features, is marked with a drop. Mingrel (lacks preverbs), Georgian (can have more than three vowels before a consonant and lacks objective agreement by number) and Svan (can have more than three vowels before a consonant, lacks two-person conjugation and circumfixes) are marked with squares.

![Figure 2. The Kartvelian languages: Laz (has all relevant features); Mingrel, Georgian and Svan (lack some relevant features).](image2)

Svan, which is classified with the Kartvelian languages, could probably borrow its peculiarities from the Northwest Caucasian languages, as it is likely to have had a linguistic contact due to their areal proximity (Figure 3).

![Figure 3. Location of Svan relatively to the other Kartvelian languages and the Northwest Caucasian languages.](image3)

The study of the Northeast Caucasian languages is especially interesting from the point of view of linguistic evolution, that’s why they will be regarded in more detail. The distribution of all Northeast Caucasian languages according to the presence/absence of relevant features is presented in Figure 4.

All deviations from the initial set of relevant features of the Northeast Caucasian languages is connected with
classes of agreement: Some languages do not have their syntactic expression in adjectives, others do not have their expression in nominal inflection and still other languages completely lack classes of agreement. Noteworthy, all languages that lack one or more types of inflection by the class of agreement (marked by a circle, square, rhomb, star – Figure 4) bent for the borders with Georgian and Azerbaijan.

Figure 4. The Northeast Caucasian languages that have all relevant features and that lack some of them.

The initial map of the distribution of languages that have a complete or an incomplete set of relevant features proves the hypothesis that the Proto-Northeast Caucasian language split into two directions – Western (Nakh and Avar-Andi-Tsezic) and Eastern Central (other branches), see Figure 5.

The question of the development of such variability is very important. Is it a simplification of inflection and consequently, is it spreading from North to the borders of the area of Northeast Caucasian languages? Or is this variability a result of complication of the system of inflection, which is spreading northward from the borders?

Relevant features of a language family or genus are not necessarily genealogic. The ancestor of the Northeast Caucasian languages is considered to be the Hurro-Urartian group. Neither language of this group – Hurrian or Urartian – possessed any classes of agreement, which allows us to suggest that the Northeast Caucasian languages were gradually developing the features of inflection by the class of agreement. As shown in Section 3, populating of East Caucasus was taking place from Transcaucasia (from the South-West) to Georgia and Dagestan (to the North and North-East).

Figure 5. Split of the Proto-Northeast Caucasian languages into two directions: Western - Avar-Andi-Tsezic and Nakh (white), and Eastern – others (black).

The spread of the languages according to the number (or complete absence) of classes of agreement is shown in Figure 6. According to these data, we can suggest that the settlement of East Caucasian was happening in two directions: To the North and to the South.

Figure 6. Spread of the Northeast Caucasian languages according to the number of classes of agreement.
7. Discussion

Unfortunately, the method of contrast queries and further variations did not answer the questions concerning external connections of the Caucasian languages, but the nature of this method did not presuppose search for the answer to these questions. In order to give a proficient answer, it is necessary to use phonetic and grammar information on the World Languages (WALS).

8. Conclusion

The present paper studied three groups from the areal community of the Caucasian languages – Northwest Caucasian, Kartvelian and Northeast Caucasian. The languages from the groups in question were regarded from the point of view of their relevant feature found by contrast queries to the database “Languages of the World” of Institute of Linguistics of Russian Academy of Sciences.

The examination of feature variations allowed us to suggest several hypotheses on the development of separate languages or groups of the Caucasian languages. Thus, Svan, which lacks three relevant features of the Kartvelian group, was suggested to have borrowed some peculiarities from Northwest Caucasian, which borders with its area of distribution.

The variation method applied to the results of the contrast query for the Northeast Caucasian languages showed that relevant features of a language are not always genealogic. So, the system of classes of agreement, which initially lacked in the ancestor languages, was acquired and complicated during the settlement of the speakers of the Northeast Caucasian languages further to the North and to the South.

There are still questions on the history of the Caucasian languages that remain unanswered. Among them are the fact of Mingrel and Georgian lacking some of the relevant features of the Kartvelian group or the reason for distribution of presence/absence of inflection by classes of agreement in verb/adjective/noun for Northeast Caucasian languages. But this research can become a good basis for further studies in this direction.

9. Acknowledgments

We would like to thank A. Kibrik for his positive attitude and valuable commentaries. The research was supported by RFBR grant # 16-06-00187.

10. References

1. Dirr A. Einführung in das Studium der kaukasischen Sprachen. Leipzig: Asia Major; 1928.
2. Starostin SA. Hypothesis of genetic connections of the Sino-Tibetan languages and Yenisein and North Caucasian languages. Linguistic reconstruction and the most ancient history of the East. Moscow. 1984; 4. 2015. Available from: http://starling.rinet.ru/Texts/cauccult.pdf
3. Silnitsky G. Quantitative grammatical and phonetic typology of languages and language features. Smolensk; 2004.
4. Languages of the World. Database of Institute of Linguistics of Russian Academy of Sciences. 2015. Available from: https://cloud.mail.ru/public/eaab329cb8c0b/DBlang/
5. Polyakov VN, Solovyev VD. Computational models and methods in typology and comparative linguistics. Kazan: Kazan State University, 2006. 2015. Available from: http://window.edu.ru/resource/068/59068
6. Polyakov VN, Solovyev VD, Wichmann S, Belyaev O. Using WALS and languages of the world. Linguistic Typology. 2009; 13:135–65.
7. Anisimov I, Polyakov VN, Solovyev VD. Database “Languages of the World”. New Version. New Research Horizons. Collection of Papers of the First International Forum on Cognitive Modeling 14-21 September, 2013, Italy, Milano Marittima”. In 2 parts. / Edited by S. Masalova V. Solovyev. - Part 1. Cognitive Modeling in Linguistics: Proceedings of the XIV International Conference « Cognitive Modeling in Linguistics. CML-2013». Rostov-on-Don: Southern Federal University Press; 2013. p. 27–34. ISBN 918-5-87872-731-0.
8. Polyakov VN. New method of search of the relevant features of a language family. Bulletin of Moscow State Linguistic University. 2014; 13(699):187–201. ISSN 1993-4750.
9. Google maps.2015. Available from: www. maps.google.com
10. Troubetzkoy N. Langues caucasiques septentrionales. – In : Les langues du monde. Paris: 1924.
11. ASPI, 2015. Available from: http://asjp.cld.org/
12. Uslar PK. Ethnography of the Caucasus: Linguistics. Tiflis. Chechen language. Part 1. Letters, 1888, 2. 2015. Available from: https://books.google.ru/
13. Dibo VA. “Language – ethnos – archaeological culture” Some thoughts concerning an Indo-European problem. Moscow. Globalization – ethno-cultural and ethno-linguistic processes. - Book 1. 2006. p., 75–94. Available from: http://www.philology.ru/linguistics1/dybo-06.html
14. Pedersen H. Türkische Lautgesetze, ZDMG, B., 57, HFT. III. Leipzig, 1903. 2015. Available from: http://menadoc. bibliothek.uni-halle.de/dmg/periodical/titleinfo/63445
15. Starostin SA. Nostratic and Sino-Caucasian. Works on linguistics. Moscow: Languages of Slavic cultures; 2007. p. 448–67. Available from: http://apsnyteka.org/file/starostin_s_trudy_po_jazykoznanyu.pdf
16. Yakhontov SE. Ancestral homeland of the Nostratic languages. Moscow: To 60th anniversary of V.A. Dibo; 1991.
17. D'yakonov IM, Starostin SA. Hurro-Urartian and East Caucasian languages. "Ancient East. Ethno-cultural ties" Moscow: Nauka; 1988. ISBN: 5-02-016792-4. 2015. Available from: http://apsnyteka.org/file/starostin_s_trudy_po_jazykoznanyu.pdf

18. Kassian A. Hattic as a Sino-Caucasian language. Ugarit-Forschungen. Internationales Jahrbuch für die Altertumskunde Syrien-Palästinas – Bd 41, 2009-2010. 2015. Available from: https://www.academia.edu/459794/Hattic_as_a_Sino-Caucasian_language

19. Avdiev VI. About cultural economic relations of the Caucasus and Forward Asia in the ancient time. Caucasus and the Culture of Ancient East. Archiv Orientalni. 1950; 18:1–2.

20. Klimov GA. Introduction to the Caucasian linguistics. Moscow: Nauka, 1986. 2015. Available from: http://reslib.org/#!book/Vvedenie_v_kavkazskoe_yazykoznanie/1123206/read/2

21. Balkarov BH. The Adyghe elements in the Ossetian language. Nalchik; 1965.

22. Alekseev ME, Klimov GA, Starostin SA, Testelets YG. (Eds.) Languages of the World: Caucasian languages. Moscow: Academia; 1999.

23. Makarova EA, Polyakov VN. The origin of the article in Indo-European Languages of Western Europe. Mediterranean Journal of Social Sciences. 2015; 6(5S4):61–75. ISSN 2039-9340.

24. Dryer MS, Haspelmath M. (Eds). The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013. 2015. Available from: http://wals.info