

Образец научного пруфридинга текста научной статьи по лингвистике для ведущего немецкого научного издательства De Gruyter Mouton

Выполненного русскоговорящим переводчиком

Было	Результат
<p>Abstract</p> <p>The relevance of the research is determined by the fact that the question whether genealogic relationship of languages can be defined basing on grammar data still remains unanswered. The objective of this article is to compare two phylogenetic trees: a tree built using data of the Automated Similarity Judgment Program (ASJP) project, and a tree built using the data of the World Atlas of Language Structures (WALS) project. Methods of the research: the method of two-objective optimization and the method of feature selection similar to data mining in the theory of artificial intelligence formed the methodological framework of this research. The material for the study includes 27 languages from the WALS, which meet the requirements to the grammatical description formulated in previous works. The research results: The comparison of the trees showed that each of them has its own advantages and disadvantages. Thus, it is possible to claim that the typological database of the WALS can provide some information on the similarity of languages. We also suggest a new variant of a phylogenetic tree that would include information both on the divergence (ASJP project) and the convergence (WALS project) of languages. The significance of the study: the present research can reveal the prospect of additional study of genealogic relationship of languages basing on large-scale description of their grammar structures.</p>	<p>Abstract</p> <p>The question whether genealogic relationship of languages can be defined based on grammar data remains unanswered. The objective of this article is to compare two phylogenetic trees: one built using the Automated Similarity Judgment Program (ASJP) project, and one using the World Atlas of Language Structures (WALS) project. The method of two-objective optimization and the method of feature selection similar to data mining in artificial intelligence formed the research framework. The material includes 27 languages from WALS that meet the requirements of the grammatical description formulated in previous works. A Hamming distance matrix was calculated for all languages under study, and, based on the matrix, a phylogenetic tree was built. The tree comparison showed that each has advantages and disadvantages. Thus, the typological database of WALS can provide some information on the similarity of languages. We also suggest a new variant of a phylogenetic tree that includes information on both the divergence (ASJP project) and the convergence (WALS project) of languages. The present research reveals prospects for additional study of languages' genealogic relationship based on large-scale descriptions of their grammar structures.</p>

Methods

Let us look closer at the main method used for the present research: linguistic phylogenetics (Nichols & Warnow, 2008). Phylogenetics has been widely used first in genetics (Edwards & Cavalli-Sforza, 1964) and then it was adopted by linguists as a method of defining the relationship of languages. One of the first works in comparative linguistics relying on phylogenetics are (Gray & Atkinson, 2003) and (Ringe et al., 2002).

Linguistics phylogeny is a universal and widely acknowledged method of comparison and classification of languages. Phylogenetic tree can reflect the quality of data if only a set of well-studied and well-described languages is used. The initial data can be lexical and phonetic data, as in the ASJP, grammatical data from the WALS or from the database “Languages of the World” IL RAS (Polyakov et al., 2009), or a combination of lexical and phonetic and grammatical data (Ringe, 2002). The received trees can further be used in future studies.

That is the reason linguistic phylogeny was chosen as a method for testing the new set of languages from the WALS in comparison with the ASJP.

According to the data (Polyakov, 2016), all languages in the WALS can be divided into sufficiently described and insufficiently described from the point of view of grammar. Thus, out of all languages presented in the WALS the authors selected 27 languages, the result of the pairwise comparison of which lies above line *iii* (Figure. 1).

The list of the selected languages is given in Table 1. This set was used to build two phylogenetic trees: on the ASJP data (basing on the lexical and phonetic data) and on the preprocessed data of WALS Program, which include only structural information.

In order to build a tree on the WALS data, for all the pairs of languages the Hamming distance (Hamming, 1950), (Wong & Kim, 2014), i.e. the percentage of unmatching feature values, was calculated and a distance matrix was built.

Finally, using MEGA 7 (Kumar et al., 2016) software, we built phylogenetic trees for the selected set of the languages.

In order to build a phylogenetic tree on the material of the ASJP using a neighbor-joining algorithm we used 40-item lists of basic

Methods

Let us take a closer look at the main method used for the present research: linguistic phylogenetics (Nichols and Warnow 2008). Phylogenetics has been widely used, first in genetics (Edwards and Cavalli-Sforza 1964; Felsenstein 2003), and it was then adopted by linguists as a method of defining the relationship between languages. Some of the first works in comparative linguistics relying on phylogenetics are from Gray and Atkinson (2003) and Ringe et al. (2002).

Linguistics phylogeny is a universal and widely acknowledged method of comparison and classification of languages. Phylogenetic trees can reflect the quality of data only if a set of well-studied and well-described languages is used. The initial data can be lexical and phonetic data, as in the ASJP, grammatical data, such as from the WALS or from the database “Languages of the World” IL RAS (Polyakov et al. 2009), or a combination of lexical and phonetic and grammatical data (Ringe 2002). The developed trees can further be used in future studies.

For this reason, linguistic phylogeny was chosen as the method for testing the new set of languages from the WALS in comparison with the ASJP. The applicability of glottochronology for marking the dates of major linguistic events was disputed by Bergsland and Vogt (1962). After years of attempts to improve the methods of glottochronology, most linguists began using the term and the methods of lexicostatistics (and, in a similar way – grammastatistics) to mark the sequence of languages’ divergence with a more vague reference to dates.

The list of the selected languages in Table 1 was used to build two phylogenetic trees: one with the ASJP data (based on the lexical and phonetic data) and one with the preprocessed data of WALS Program, which include only structural information.

In order to build a tree using the WALS data, for all the pairs of languages, the Hamming distance (Hamming 1950; Wong and Kim 2014), that is, the percentage of unmatching feature values, was calculated, and a distance matrix was built.

Finally, using MEGA 7 (Kumar et al. 2016) software, we built phylogenetic trees for the selected set of the languages.

In order to build a phylogenetic tree on the material of the ASJP using a neighbor-joining algorithm, we used 40-item lists of basic

vocabulary written in the form of the ASJP-code and MEGA7 program (Kumar et al., 2016). Below are the trees and their analysis.

Results

Thus, the phylogenetic tree built on the ASJP data is closer to the traditional views of comparative linguistics. The phylogenetic tree built on the WALS data taken with limitations (Polyakov et al., 2016) marks problems that require explanation.

For example, why did the English borrow some elements of Swedish grammar and not of Latin grammar, though it is widely known that Britain was first conquered by Ancient Rome (I-V cc. AD), and then became part of Doggerland (modern Holland and Denmark) (VIII-X cc. AD) (BAUGH & CABLE, 2002).

It can probably be accustomed for by different sociolinguistic situations accompanying these events. Anyway, this question will require a separate research.

Another example is Greece and Bulgaria. The ASJP classifies Modern Greek as a separate branch of the Indo-European languages (Gray & Atkinson, 2003). As opposed to the ASJP, the WALS-tree classifies Modern Greek with the Balto-Slavic languages and defines Bulgarian as its closest relative.

Bulgaria is the closest geographic neighbor of Greece; consequently, it is not surprising that Bulgarian borrowed its grammar structure from Greek. However, the history of Bulgaria has been thoroughly studies. Bulgaria existed under Roman protectorate (I-V cc. AD), then under the protectorate of the Byzantine Empire (in fact – Greece), at the same time Bulgaria was experiencing strong Gothic influence (IV c. AD), in VII-VIII c. there was a Turkish invasion, later, in IX-XIV cc. AD there existed The Second Bulgarian Empire, in XIV-XIX cc. there was another Turkish conquest. Finally, in 1878 Bulgaria was freed, and the new Bulgarian state appeared (Maslov, 2005).

Thus, the question that remains unanswered is why Bulgarian borrowed elements of the grammar structure from Greek and not from Latin, Gothic or Turkish. The fact that the Byzantine Empire was historically the first conqueror cannot serve as a plausible explanation for the phenomenon, as in case of Britain the first conquerors were Romans, but the grammar

vocabulary written in the form of the ASJP-code and MEGA7 program (Kumar et al. 2016). The trees and their analysis are given below.

Results

Thus, the phylogenetic tree built from the ASJP data is closer to the traditional views of comparative linguistics. The phylogenetic tree built from the WALS data taken with limitations (Polyakov et al. 2016) brings up problems that require explanation.

For example, why did English borrow some elements of Swedish grammar and not of Latin or Celtic grammar, though it is widely known that during the Roman invasion in I–V cc. AD, the native languages were Celtic (Brittonic), and not Germanic (Schrijver 2013)? This may be explained by the fact that in the IX century, part of England became a territory where laws of the Danes held sway (Hornung 2017). As is known, Danish and Swedish are closely related languages. The lexical borrowings from Latin came to English mainly from Medieval French (Norman conquest, XII c.) (Lutz 2017).

Another example is Greece and Bulgaria. The ASJP classifies Modern Greek as a separate branch of the Indo-European languages (Gray and Atkinson 2003). As opposed to the ASJP, the WALS-tree classifies Modern Greek with the Balto-Slavic languages and defines Bulgarian as its closest relative.

The history of Bulgaria has been thoroughly studied. First, this territory was inhabited by Thracians. The first Bulgarian Empire was established by proto-Bulgarians, Slavs, and Thracians in the VII c. AD (Angelov 1971). In VII–VIII c. there was a Turkish invasion, and later, in IX–XIV cc. AD, the Second Bulgarian Empire existed, and in XIV–XIX cc., there was another Turkish conquest. Finally, in 1878, Bulgaria was freed, and the new Bulgarian state appeared (Maslov 2005).

Thus, the question that remains unanswered is why Bulgarian borrowed elements of the grammar structure from Greek and not from Turkic languages (Bulgar languages or languages of Turkic conquerors). Other languages that could have influenced the Bulgarian grammar are those of proto-Slavic and Thracian tribes. However, as is widely known, Old Slavic was used as a liturgical language and was strongly influenced by Byzantine Greek; that could explain the

structure was borrowed from Danish and Swedish Vikings (or vice versa). This question will require further study.

Let us look at the third situation – the order in which the languages separated from the Indo-European tree. According to (Gray & Atkinson, 2003), the Indo-European languages first divided into two groups: the Indo-Aryan languages and the languages of Europe (6900 years ago), then the Iranian languages separated from the Indian languages (4600 years ago). The Indo-European languages divided into Eastern and Western (6500 years ago). The WALS-tree reflects that scenario better than the ASJP-tree.

Let us make a preliminary conclusion. The ASJP allows building phylogenetic trees basing on lexical and phonetic data to the depth up to 3000–6000 years with the quality comparable with the quality of the trees built manually. At the same time the WALS, which is based on grammar data, allows rendering the relationship of languages at further distance (over 6000 years) and raises questions on borrowing of grammar structure of languages that do not have any traces of lexical and phonetic similarity.

similarities between Modern Greek and Bulgarian.

Let us look at the third situation – the order in which the languages separated from the Indo-European tree. According to Gray and Atkinson (2003), the Indo-European languages first divided into two groups: the Indo-Aryan languages and the languages of Europe (6900 years ago), then the Iranian languages separated from the Indian languages (4600 years ago). The European part of Indo-European languages then divided into Eastern and Western (6500 years ago). The WALS-tree reflects that scenario better than the ASJP-tree.

Let us make a preliminary conclusion. The ASJP allows for the building of phylogenetic trees based on lexical and phonetic data to the depth of up to 3000–6000 years with a quality that is comparable to the quality of manually built trees. At the same time, the WALS, which is based on grammar data, allows for the rendering of the relationships between languages at a further distance (over 6000 years) and raises questions about borrowing of grammar structures between languages that do not have any traces of lexical and phonetic similarities.