e-Journal of Linguistics
Available online at https://ojs.unud.ac.id/index.php/eol/index
Vol. 18, No. 1, January 2023, pages: 01--10
Print ISSN: 2541-5514 Online ISSN: 2442-7586
https://doi.org/10.24843/e-jl.2024.v18.i01.p01

Dialectometry and the Azerbaijani Language: Problems, Solutions and Perspectives

[1]Afruz Gurbanova
Institute of Information Technology, Baku, Azerbaijan, afruz1961@gmail.com

[2]Mehriban Baghirova
Institute of Information Technology, Baku, Azerbaijan, mehriban.amea@gmail.com

Article info

Received Date: 17 July 2023

Accepted Date: 26 July 2023

Published Date: 31 January 2024

Keywords:
dialectology, dialectometry, dialectological atlas, linguistic geography

Abstract

Dialectology is the study of dialects, while dialectometry is the measurement of dialect variations, i.e. linguistic differences, the distribution of which is determined primarily by geography. New sources of information and analytical software are expanding the scope of dialectometry. Dialectometric methods were analyzed in the research work. The possibilities of using dialectometric methods in Azerbaijan were investigated and recommendations were given in the direction of their application

1. Introduction

The changes occurring in society in modern times, strengthening the integration of scientific fields necessitate a new approach to dialectological research.

The relevance of studying dialects is determined by the following features:

- dialects serve to understand the process of the historical development of a language: dialects often retain archaisms necessary to recreate a broad language movement;
- dialects serve to establish a mutual relationship between the history of the language and the history of the people, as dialect facts allow us to trace how tribes and peoples migrated in ancient times;
- dialects serve to understand the variety of words, sounds and forms of the modern language, practically to consider the features of local speech.

The largest areas of study of dialects are dialect lexicography and linguistic geography (Levina, 2016). The oldest branch of dialectology, today called "dialect geography", studies the geographical distribution of language varieties.

Methodological innovation in the field of linguistic geography is related to the development of computer programs that allow direct analysis of large volumes of data and graphical visualization in a simple way, starting from the 1980s.

The systematic study of the dialects of a language by the linguistic-geographical method began in Europe in the second half of the 19th century. This method consists of collecting data from a large number of settlements through a single questionnaire. Answers to each question of the questionnaire form the basis of a dialect map reflecting their territorial distribution, and a set of dialect maps compiled for the same settlements is combined in a dialectological atlas (Arkhangelsky, 2021). The first such atlas was G. Wenker's atlas of German dialects (1876-1881) (Herrgen, 2010). The French language atlas and an atlas of other European languages were published by J. Gillieron (1902-1910) [Goebl, 2010]. All major European languages now have language atlases, including hundreds or thousands of maps. The Linguistic Atlas of Europe, published since 1975, is a project that collects data on many languages. This atlas is the largest scientific project for the study of languages using the linguistic-geographical method. The Atlas covers six language families present on the European continent: Altaic, Basque, Indo-European, Caucasian, Semitic and Uralic. These families are divided into 22 language groups comprising 90 languages and dialects (Viereck, 2006).

Today, the rapid development of mass media and new technologies has greatly impacted local languages and dialects, putting them in danger of extinction (Gurbanova, 2023). In order to preserve local languages and dialects, it is necessary to carry out dialectometric studies and to determine the dialect variations of the language, to assess the similarities and differences between them.

The necessity of a deeper study of the dialect differences of national languages by using mathematical methods, especially dialectometric research, in the current era of rapid globalization in the research work, was justified, and proposals were developed in this direction.

2. Research Method

2.1 Dialectology

A dialect is a regionally or socially distinctive variety of a language defined by a particular set of words and grammatical structures. Spoken dialects are also usually associated with a different pronunciation or accent.

A dialect is a variety of language used in a particular geographical location. So dialects are changed and influenced by a group of people who use it. Other social factors such as class, occupation, and age can also create and influence dialects.

The aims and methods of dialect geography are as follows:

- Pure form - Dialect geography studies the relationship between language and geography, identifying the local dialect, especially in villages, where the dialect is in its purest form before being polluted, weakened and completely lost. He looks for the most "original" and the most "typical" form of speech used in a certain field.
- Non-educated Old Rural Males - NORM - The purest form of a dialect is mostly taken from old rural males.
- Rural area - In order to obtain the purest form of the dialect, teaching should be concentrated in villages where the language is less "contaminated" by foreign elements.
- Raw data - Data collected for research is presented in raw form.
- Linguistic Mapping - After the interviews are completed, data is collected, responses are tabulated, and linguistic maps are constructed to show dialect variations.

This method of studying dialects is also known as traditional dialectology (Chapter 2: Literature Review).

Dialects of a particular language keep the history, culture, ethnography and folklore of the people alive in the language of their great-grandparents, while preserving the intricacies of the language and conveying it to future generations. Therefore, the study of dialects and dialects of the language is one of the most significant issues of linguistics. The collection and study of the materials of each dialect or dialect plays an important role in the study of the language and history of the people.

Dialect differences are the main research object of dialectology. Dialectology is the branch of linguistics that studies regional dialects, dialect differences, and dialect language in its present state and history.

In traditional dialectology, sources of information are dialects, dictionaries, dialect atlases, and other materials.

As in the case of other languages, further strengthening of the state care for the Azerbaijani language has opened wide opportunities for the development of various fields of Azerbaijani linguistics. The inclusion of the issue of "Ensuring the study of various dialects and dialects of the Azerbaijani language in accordance with the requirements of the modern era" into the State Program "On the use of the Azerbaijani language in accordance with the requirements of the time and the development of linguistics in the country" put the study of the dialects of the Azerbaijani language as an important task (Decree of the President of the Republic of Azerbaijan, 2013). In the direction of the implementation of the program, "Nakhchivan Dialectological Atlas of the Azerbaijani Language", "Karabakh and Eastern Zangezur Dialectological Atlas of the Azerbaijani Language", "Nakhchivan Dialectological Dictionary of the Azerbaijani Language", "Karabakh Dialectological Atlas of the Azerbaijani Language", "Karabakh Dialects of the Azerbaijani Language" monograph and "Karabakh Dialect of the Azerbaijani Language" dialectological dictionary" was published.

"Nakhchivan dialectological atlas of the Azerbaijani language" contains 278 maps covering about 1000 settlements of the Nakhchivan Autonomous Republic, and a CD with the voices of the informants during the expedition. In the atlas, the characteristics of dialects that create isogloss are given with special signs, and their distribution area is determined.

In the "Karabakh and Eastern Zangezur dialectological atlas of the Azerbaijani language", 692 settlements were coded and reflected in 278 maps. In each map of the atlas, the specific characteristics of the dialect words of the region, changes, usage forms and variants, replacement cases, and synonyms of the words used in the settlements are also given.

The Azerbaijani language has a rich dialect system. Dialects and dialects of our language have been studied from different aspects and grouped according to their geographical location and level characteristics. Dialects and dialects of the Azerbaijani language are divided into 4 groups according to the principle of geographical area:

- Eastern group dialects and dialects of the Azerbaijani language - Guba, Baku, Shamakhi dialects and Mughan, Lankaran dialects;
- Western dialects and dialects of the Azerbaijani language - Kazakh, Karabakh, Ganja dialects and Ayrym dialects;
- Dialects and dialects of the northern group of the Azerbaijani language - Sheki dialect and Zagatala-Gakh dialects;
- Southern dialects and dialects of the Azerbaijani language - Nakhchivan, Ordubad, Tabriz dialects and Iravan dialect. (Shiraliyev, 2008).

An in-depth study of dialects is necessary for proper assessment of variability at all levels of language structure, development of optimal grammatical orthoepic and other norms. The main

differences between the dialects are investigated based on the available speech data of those dialects.

From sociolinguistic methods in the study of dialects: long-term panel observation, various types of speech writing, diaries, questionnaires, interviews, surveys, secret tests, etc., is used.

Dialects are the unwritten form of language, dialectologists use the questionnaire method and direct observation method to study them.

The collection of information about the linguistic features of the dialect by the questionnaire method is carried out by receiving written answers to the questions of a specially designed questionnaire from linguistically competent persons (teachers, rural intellectuals, etc.).

One of the methods of studying modern dialects is direct observation, when the researcher identifies linguistic features while listening to live dialect speech. With the method of direct observation, dialectologists record the live speech of dialect speakers on the basis of a pre-designed questionnaire program. In order to determine which generations and speakers have more dialects, it is necessary to measure dialectics. Using special computer programs, it is possible to obtain more information than tape recordings, which allow you to repeat and analyze speech segments continually. A fund of such writings will preserve modern dialect speech for future researchers (Methods for studying dialects, 2015).

2.2 Dialectometry

Obtaining an appropriate measure of the distance between two pronunciations is important not only for dialectologists interested in establishing relationships between different dialects, but also for sociolinguists studying the effects of boundaries on spoken language.

The presence of a measure of distance between word pronunciations necessitates quantitative analysis to investigate geographic and sociolinguistic factors (Wieling et al., 2014).

In the second half of the 20th century, a large number of collected dialectological materials and the development of information technologies led to the emergence of a new approach to the study of dialects - dialectometry. Dialectometry studies language variation using statistical methods. Dialectometric methods allow dialectologists to quantify dialect differences and measure language change based on this (Nerbonne & Kretzschmar, 2006; Pickl & Rumpf, 2012).

Dialectometry is a branch of geolinguistics that deals with the measurement, visualization, and analysis of common dialect similarities or distances depending on the characteristics of a geographic location. Dialectometric studies form computational approaches to identify "general, apparently hidden structures from large numbers of features" and focus on quantitative, cartographic visualization, and exploratory data analysis to extract patterns. Dialectometry provides methods for estimating the linguistic distance between two arbitrary dialects in dialectological atlas projects. The main goal is to determine the degree of closeness of dialects, confirm existing dialect classifications, and solve problems related to dialect division (Vozenilek et al., 2022).

In addition to dialectological studies, dialectometry has made theoretical contributions to comparative dialectology and the study of dialect distribution.

3. Discussion

In (Mehrabani & Hansen, 2015), the main differences between dialects or closely related languages are explored based on the available speech data of those dialects/languages. A method is proposed to measure spectral acoustic differences between dialects based on the analysis of volumetric space within a 3D model using log probability distributions derived from traditional cepstral Mel Frequency Coefficient and Gaussian Mixture Models. The proposed dialect affinity

measures are evaluated and compared on a corpus of Arabic dialects as well as a corpus of closely related South Indian languages.

In (Goebl, 2010), documented the taxometric and cartographic achievements of the Salzburg school of dialectometry. Problems related to Romance dialectology and Romance language geography, historical linguistics, numerical classification, statistics and statistical cartography were investigated. Issues such as measuring linguistic atlas data, creating a data matrix, choosing a similarity index, creating appropriate similarity and distance matrices, similarity maps, parameter maps and cartographic email of dendrograms were analyzed and their visualization was carried out using "Visual DialectoMetry" software.

In (Donoso & Sánchez, 2017), an information-theoretical approach to geographic language variation is proposed using a corpus based on Twitter. Dialectometric measurements (cosine similarity and Jensen-Shannon difference) are used to quantify the linguistic distance between the hollows created in a single grid (set) on the map. The authors believe that social networks can be used qualitatively for dialectometric analysis.

In (Asadpour, 2011), the method of measuring the cumulative degree of lexical, phonological, morphological and syntactic differences between the dialect variations of the Azerbaijani language is proposed. Using hierarchical cluster analysis, dialect distances are analyzed, the benefits of applying the methods developed and tested in Turkic languages to the Azerbaijani language are shown.

4. Novelty

Dialectometric methods are constantly improving, opening new possibilities for explaining linguistic variations:

- focus on identifying the most important (diagnostic) individual language elements that form the basis of the general dialect variation;
- understanding that lexical and social factors can determine geographical variation;
- new methods for evaluating linguistic change in dialects;
- dialectological theory;
- more attention to dialect grammar and morphosyntax;
- to use new data sources in addition to the traditional dialect atlas data;
- to create new (online) software that allows dialect researchers to use dialectometric tools more easily (Wieling & Nerbonne, 2015).

The changes occurring within the language are not only related to the geographical distance of the population speaking the dialect variants. Factors such as education, access to audiovisual media, self-affirmation, and cultural expansion can influence language use.

The main concepts of dialectometry are:

- measures of difference or similarity between linguistic varieties for one or more linguistic functions;
- aggregation and clustering algorithms for organizing large data sets by proximity/difference;
- tools to present changes measured across time, space and social groups.

Dialectometric methods began to develop actively in the 1970s and 1980s after Jean Séguy (France) and Hans Goebl (Austria) applied statistical approaches to the study of Romani dialects using previously collected atlas data (Arkhangelsky, 2021).

At the beginning of the 21st century, John Nerbonne (American) and Wilbert Heeringa, following the research started by Goebl, developed and tested new analytical methods based on various statistical procedures, in addition, they included a quantitative measure of articulation distance. These studies contributed to mastering the theoretical foundations of dialectometry, expanding its application areas and perspectives.

Dialectometry began to develop further with the application of computer tools for the analysis of geolinguistic data. The most common and used statistical analysis programs can be applied to the study of language variations. Widely used by statisticians and data collectors, the R open-source statistical package is commonly used to obtain figures involving distance measurement, similarities, cluster analysis, and many other complex analyses. R packages such as rMaps make it easy to create and share interactive maps from R.

In recent years, research in the field of dialectometry has used various methods of data visualization that represent the main methods of GIScience (geographical information science). GIS software systems offer a number of software products that can be used directly to represent geolinguistic data, the results obtained from the statistical analysis of this data in a simple way (Dubert & Sousa, 2016).

Dialectometric methods are based on the concept of "distance" between settlements. By distance here we mean the mathematical function d(X, Y) that calculates the value based on the atlas data for any two settlements (*X* and *Y*), showing how the answers to the survey questions at one point differ from the answers to the questions at another point. In practice, different functions can be applied, but in any case they must meet at least three requirements (Arkhangelsky, 2021):

• If the answers to all questions of the questionnaire coincide at two points, the distance between them is equal to 0 d(X, Y)=0;

• The function is symmetric, i.e. d(X, Y) = d(Y, X);

• The more the questionnaires X and Y differ on a large number of questions, the greater the distance between them.

Dialectometrics is a new direction emerging from classical dialectology, where the differences between different dialects in a region are statistically calculated and presented using dialect maps and atlases.

A linguistic map is a thematic map showing the geographical distribution of speakers of a language or isoglosses of a language family. Linguistic atlases serve as empirical databases documenting in detail the dialect profile of a large number of locations. A variety of well-known numerical classification methodologies are used to summarize and visualize the underlying pattern from the vast amount of data contained in linguistic atlases.

The matrix of double points calculated on the basis of the data of the dialectological atlas can be used in several ways. First, it provides to find important groups of isoglosses, i.e. lines, where the boundary of variants is crossed at the same time in many dialect maps. Secondly, the classification of dialects and dialects can be obtained with the help of the automatic clustering method. In this case, clusters (sub-dialects) will connect points that are slightly different from each other, but significantly different from neighboring clusters (Batagelj et al., 1992).

Since it took many years to create dialect atlases, researchers have also tried to use other sources of information to study dialect variation. Szmrecsanyi called these research approaches "Corpus dialectometry". When studying pronunciation variation, they compare using the data collected by the compilers of linguistic atlases - the pronunciations of the same word in different places.

Corpus dialectometry is a younger field, where statistical methods are applied to Corpus data rather than dialect maps (Szmrecsanyi, 2011).

Assessing the closeness between several dialects of the language is an interesting but complex research topic. This type of assessment shows how often dialects are mixed or differentiated in the given language space. Thus, the data obtained determine dialect differences, isogloss, clusters, etc., and it opens up ample opportunities for visualization.

The main issue of dialectometric analysis is the acquisition of a modern “map of similarity”of linguistic idioms with each other. In this case, the similarity can arise not only by the relationship of idioms, but also by the recent migration or the general influence of another language. Evaluation can be carried out in the following sections (Mehrabani & Hansen, 2015):

- differences in the physical sounding of speech;
- differences in the linguistic expression of speech;
- signs of perception assessment;
- differences of classifiers of automatic speech sounding systems.

There are a number of ways to compare languages, dialects, or other types of speech. Various string distances such as Levenstein, Euclidean (Jeszenszky et al., 2017) and Manhattan (Heeringa et al., 2009) distance are used to account for pronunciation differences between dialects.

Levenstein distance is a measure of the difference between two sequences or strings. Such sequences can be words of the language being studied. When measuring the Levenshtein distance between two pronunciation variants of a word, the minimum number of operations (insertion, deletion and replacement of characters) that one variant must go through in order to transform it into another is calculated (Heeringa, 2004).

The Wagner-Fischer algorithm is a dynamic programming algorithm that measures the Levenstein distance between two character strings.

A dialect continuum is a group of language dialects that vary within an area. On the dialect continuum, the further apart two dialects are, the more different they are, the more difficult to understand each other, or not at all. People in close proximity on the dialect continuum can understand each other when they speak.

Multivariate scaling (MS) is a statistical method used to study dialect continua. MS converts complex distance data into interpretable low-dimensional images (Klis & Tellings, 2020).

Clustering algorithms are applied to classify the studied set of objects by identifying the closest clusters in that set. The source material for the analysis is the matrix of distances between the studied objects, and the result of the algorithm can be presented as a hierarchical structure that shows the sequence of clusters (Galdino & Maciel, 2019).

The UPGMA (Unweighted Arithmetic Mean Pair Grouping Method) method is one of the simplest and most widely used hierarchical clustering algorithms for creating a dendrogram from a distance matrix. Here, the local topological relationships are obtained in descending order of similarity and the dendrogram is constructed stepwise. That is, the two closest data points are identified first and grouped in the dendrogram. After the first grouping, the two closest data points are treated as one data point (composite) and new distances are calculated using the mean of the distances between the simple data point and the components of the composite data point. Then the next closest data points are added to the dendrogram until all data points are included.

The WPGMA (Weighted Pair Group Method using Arithmetic Averages) algorithm is similar to its unmeasured variant, the UPGMA algorithm. In the WPGMA algorithm, the distance between clusters is calculated as a simple average. WPGMA gives a simple average weighted result, while in UPGMA it gives a proportional average weightless result (Garcia-Vallvé & Puigbo, 2009).

Hans Gebl and Edgar Heimerl developed special software called “VisualDialetometry (VDM)”.

VDM was developed as a dialectometry project at the University of Salzburg between 1998 and 2000 and implements algorithms that support dialectometric analyses of Dialectological Atlas data. It offers functionality for managing pre-classified Atlas data, various dialectometric

approaches to data analysis, and various methods for visualizing the results of such analyzes (dendrograms, diagrams, or maps) is one of the most used tools for dialectometric analysis of various languages of VDM (Galdino & Maciel, 2019; Goebl, 2006).

Gabmap is a web application for dialectometry and cartography. It allows for comparison and statistical analysis of dialect data. Gabmap is a graphical user interface that performs not only comparison of vocabulary or other categories of information, but also comparison of pronunciation using editing distance. Gabmap allows researchers in dialectology to perform computer-assisted exploration and calculations (Website "Dialektometrie Projekt" – Salzburg; Nerbonne et al., 2011).

Quantitative analysis methods make it possible to reveal the relationship existing between the two distributions of data. The method used in this case is called correlation dialectometry. The method allows you to visualize and compare geolinguistic relationships between the distribution of phonetic data and other morphological data, as well as analyze the relationship between linguistic and geographical distances (Montemagni, 2008).

5. Conclusion

Dialectometric methods were analyzed in the research work. It was determined that dialectometric methods are constantly being improved, significant progress has been made in this field. So that:

- Various methods have been developed by specialists in the field of dialectometry to simultaneously analyze the linguistic and social factors behind geographical differences and to assess their relative strength;
- Dialectometry has been greatly improved to assess linguistic changes in dialects;
- Dialectometry uses data sources other than traditional dialect atlases, especially dialect corpora built from online sources, to study dialect variation;
- With the creation of new (online) applications, many dialectologists use dialectometric tools.

The importance of in-depth study and application of dialectometric methods in Azerbaijan was determined. The possibility of using dialectometric methods for calculating the dialect differences of the Azerbaijani language is investigated and the following is recommended:

- It is necessary to create a web portal to enter the data of the "Dialectological Atlas of the Azerbaijani language" into the database and ensure the availability of this data for everyone. The portal will allow the creation of a single dialectal environment of the Azerbaijani language and the operative search of dialects.
- It is possible to use modern tools, including mathematical methods, especially algorithms widely used in dialectometric research - cluster analysis and multidimensional scaling - to analyze the data entered into the database.
- As a part of the ecosystem of the Azerbaijani language on the e-state platform (Alguliyev et al., 2021), the creation of a single dialect system of the Azerbaijani language will create a technological basis for deeper study, classification and evolution of these dialects.

6. Acknowledgements

We express our gratitude to the Institute of Information Technologies for the support provided in conducting the research.

7. References

Alguliyev R., Yusifov F., Gurbanova A. (2021). Protection of Azerbaijani Language in e-government platform, *E-Journal of Linguistics*, Vol. 15(2), pp. 155-161. https://ojs.unud.ac.id/index.php/eol/issue/view/3948

Arkhangelsky, T. (2021). Application of the dialectometric method to the classification of Udmurt dialects. *Ural-Altai studies*, V 2, 7-20.

Asadpour, H. (2011). A Survey of Language Varieties in Azerbaijan-e Qærbi through Dialectometric Analysis. *Journal of Persian Academy of Language, Dialectology*, V.1, 173-202.

Batagelj, V. et al. (1992). Automatic Clustering of Languages. *Computational Linguistics*, V. 18(2), 339-352.

Chapter 2: Literature Review http://studentsrepo.um.edu.my/5575/3/3.Chapter_2.pdf

Decree of the President of the Republic of Azerbaijan on the approval of the "State Program on the use of the Azerbaijani language in accordance with the requirements of the time and the development of linguistics in the country", (2013). Baku, https://e-qanun.az/framework/25537

Donoso, G. & Sánchez, D. (2017). Dialectometric analysis of language variation in Twitter. http://aclanthology.lst.uni-saarland.de/W17-1202.pdf

Dubert, F. & Sousa, X. (2016). On quantitative geolinguistics: an illustration from Galician dialectology. *Dialectologia: revista electronica*, 191-221.

Galdino, S. & Maciel, P. (2019). Weight Pair Group Average Mean Clustering for Interval-valued Data. *IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 1-7.

Garcia-Vallvé, S. & Puigbo, P. (2009). DendroUPGMA: a dendrogram construction utility. *Universitat Rovira i Virgili*, 1-14.

Goebl, H. (2006). Recent advances in Salzburg dialectometry. Literary and linguistic computing, V. 21(4), 411-435.

Goebl, H. (2010). Dialectometry: theoretical pre-requisites, practical problems, and concrete applications (mainly with examples draw from the" Atlas linguistique de la France", 1902-1910). *Dialectologia: revista electrònica*, 63-77.

Gurbanova A.M. (2023). Problems and Prospects for Minority Languages in the Age of Industry 4.0. *Lecture Notes on Data Engineering and Communications Technologies (LNDECT) – Springer Publisher*, V. 158, pp. 722-734. https://link.springer.com/chapter/10.1007/978-3-031-24475-9_59, DOI: 10.1007/978-3-031-24475-9_59

Heeringa, W. (2004). Measuring dialect pronunciation differences using Levenshtein distance. 323 p. http://www.wjheeringa.nl/thesis/thesis.pdf.

Heeringa, W. et al. (2009). Measuring Norwegian dialect distances using acoustic features. *Speech Communication*, V. 51(2), 167-183.

Herrgen, J. (2010). The digital wenker atlas (www. diwa. info): an online research tool for modern dialectology. *Dialectologia: revista electrònica*, 89-95.

Jeszenszky, P. et al. (2017). Exploring global and local patterns in the correlation of geographic distances and morphosyntactic variation in Swiss German. *Journal of Linguistic Geography*, V. 5(2), 86-108.

Levina, M. (2016). Linguistic geography as a basis for areal studies of the Mordovian languages. *Finno-Ugric world*, V. 3 (28), 50-59.

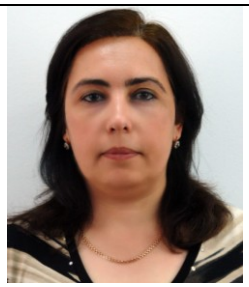
Martijn van der Klis & Tellings, J. (2020). Multidimensional scaling and linguistic theory. https://www.researchgate.net/publication/346857889_Multidimensional_scaling_and_linguistic_theory.

Mehrabani, M. & Hansen, J. (2015). Automatic analysis of dialect/language sets. *International Journal of Speech Technology*, V. 18, 277-286.
Methods for studying dialects, https://studfile.net/preview/2378303/page:4/
Montemagni, S. (2008). The space of Tuscan dialectal variation: A correlation study. *International Journal of Humanities and Arts Computing*, V. 2(1-2), 135-152.
Nerbonne, J. & Kretzschmar, W. (2006). Progress in dialectometry: toward explanation. Literary and Linguistic Computing, V. 21(4), 87-397.
Nerbonne, J. et al. (2011). Gabmap-a web application for dialectology. *Dialectologia: revista electronica*, 65-89. https://www.raco.cat/index.php/Dialectologia/article/view/245345.
Pickl, S. & Rumpf, J. (2012). Dialectometric concepts of space: Towards a variant-based dialectometry. *Dialectological and folk dialectological concepts of space: Current methods and perspectives in sociolinguistic research on dialect change*, V. 17, 199-214.
Shiraliyev, M. (2008). Basics of Azerbaijani dialectology. *Baku, "East-West"*, 416 p.
Szmrecsanyi, B. (2011). Corpus-based dialectometry: a methodological sketch. Corpora, V. 6(1), 45-76.
Viereck, W. (2006). The linguistic and cultural significance of the Atlas Linguarum Europae. *Gengojohogaku Kenkyuhokoku (Memoir for Linguistic Informatics)*, V. 9, 58-80.
Vozenilek, V. et al. (2022). Mapping, synthesis and visualization of Czech dialects. *International Journal of Cartography*, V. 8(1), 148-163.
Website “Dialektometrie Projekt” – Salzburg. URL: http://www.dialectometry.com/.
Wieling M. et al. (2014). A cognitively grounded measure of pronunciation distance. *PloS one*. 9(1), e75734
Wieling M. & Nerbonne, J. (2015). Advances in dialectometry. *Annual Review of Linguistics*, V. 1, 243-264.

Biography of Authors

[1] 

Afruz Gurbanova graduated from the Applied Mathematics faculty of the Baku State University in 1984 and was appointed to the Institute of Information Technology of the Azerbaijan Academy of Sciences. He worked at the Institute of Information Technology as a programmer, researcher, and sector manager. Her area of interest includes web technologies, computational linguistics, computational terminology, linguistics demography. She currently works as a senior researcher. She is the author of more than 40 papers.
Email: afruz1961@gmail.com

[2] 

Mehriban Baghirova graduated from the Faculty of Automated Control Systems of the Azerbaijan Technical University in 1986 and was appointed to the Institute of Information Technologies of the Azerbaijan Academy of Sciences. She worked as a programmer, senior researcher at the Institute of Information Technologies. Her interests include web technologies, computational linguistics, and linguistic demography. She is the author of several articles.
Email: mehriban.amea@gmail.com