

# Formation of linguistic demography on the national e-government platform: problems and their solutions

*Gulnara Ch. Nabibayova*

Institute of Information Technology, B. Vahabzade str., 9A, AZ1141 Baku, Azerbaijan

[gnabibayova@gmail.com](mailto:gnabibayova@gmail.com)

[orcid.org/0000-0001-8743-7579](https://orcid.org/0000-0001-8743-7579)

## ARTICLE INFO

<http://doi.org/10.25045/jpis.v14.i2.09>

### Article history:

Received 8 February 2023

Received in revised form

14 April 2023

Accepted 3 July 2023

### Keywords:

Demography

Dialect

Linguistic demography

OLAP

Electronic demographic system

Data Warehouse

Native language

Language of daily communication

Language of use

## ABSTRACT

The advancing of information and communication technologies has contributed to their introduction into almost all areas of human activity, as well as the field of demography. The advent of such a field as electronic demography has become a precondition for the improvement of an electronic demographic system on the e-government platform. Demography is a multidisciplinary field of science, and along with other areas, it also includes linguistic demography. Therefore, in the electronic demographic system, along with other segments, there is a segment of linguistic demography. Each area in the data warehouse (DW) of the electronic demographic system has its own data mart (DM). This article highlights several complications in the formation of DM of linguistic demography. These complications are linked to the fact that linguists-researchers offer several approaches to the definitions of the notions of "language" and "dialect". Moreover, when establishing a language indicator in a country, one of its three options can be approved - as a native language, as a language of everyday communication or as a language of use. In this regard, certain difficulties arise in the establishment of a segment of linguistic demography in the electronic demographic system. This article shows the complications of establishing linguistic demography and presents certain proposals for their solution.

## 1. Introduction

The implementation of a demographic policy targeted at enhancing the demographic situation in the country is one of the crucial tasks facing each state. The demographic policy includes the solution of essential tasks such as prolonging the life expectancy of the population, increasing the birth rate, decreasing morbidity and mortality, regulating migration, etc. Demographic research is of great importance in solving these problems. The outcomes of these studies constitute the basis of demographic policy.

Today, as a result of the effects of information technology on all fields of human life and activity, along with the areas as electronic science, electronic medicine, etc. a new field of demography - electronic demography appeared on

The implementation of a demographic policy targeted at enhancing the demographic situation in the e-government platform in Azerbaijan (Nabibayova, 2023, b).

Demography (Greek δῆμος - people, Greek γράφω - I write) is a field of science that explores the patterns of population growth, its dependence on socio-economic conditions, natural factors, migration, population density, and its distribution. An important feature of demography is that it is an interdisciplinary field of science (Demography).

Note that demography is closely linked to economics, political science, ethnography, medicine, geography, etc. Therefore, it is divided into a number of specialized areas, each of which studies certain demographic processes (Nabibayova, 2023, b).

Figure 1 illustrates the fields of demography, among which there is linguistic demography.

The second chapter of the article defines the concept of linguistic demography, presents the place of linguistic demography on the national platform of the electronic government, and emphasizes its interdisciplinary feature.

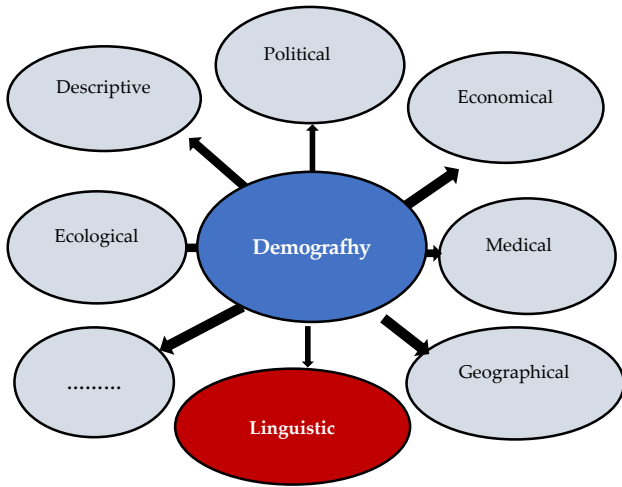


Fig. 1. Fields of demography

It presents a brief description of OLAP technology, as well as the transition from traditional OLAP to smart OLAP in the era of big data. This chapter presents a structure of DW consisting of several DMs, one of which relates to linguistic demography.

The third chapter defines the notions of language and dialect, and highlights the existence of various approaches to the problem of assessing the number of speakers of a specific language. Several approaches to the classification of the dialects of the Azerbaijani language are presented. Different approaches to the description of the concept of “native language” are presented. Recommendations are given for solving problems arising when entering data on linguistic demography into an electronic demographic system.

Finally, the fourth chapter concludes the article by offering a summary of research findings in this interesting area.

## 2. Formation of linguistic demography within the electronic demography system

Linguistic demography studies the numerical or statistical data of languages among all population groups, as well as their geographical distribution. The key sources of information include the population census, on which large-scale studies are performed. Conducting a census with a frequency of 5-10 years is optimal. The linguistic data collected in this manner helps linguists engaged in the linguistic composition of a country, state or region and the dynamics of its changes (Pauwels A., 2016).

Figure 2 illustrates the position of linguistic demography within the framework of national e-government. As the figure shows, e-demography, which is based on the electronic demographic system, is an element of the e-government along with e-medicine, e-education, etc. Linguistic demography, in turn, is one of the fields of e-demography. Note that the creation of an electronic demographic system is an urgent task. Through an electronic demographic system, demographic processes can be understood better, deeper research may be performed at different levels and in different sections, that is, social demography, medical demography, linguistic demography, etc. Furthermore, the data of the electronic demographic system enables conducting demographic statistics in various areas more effectively

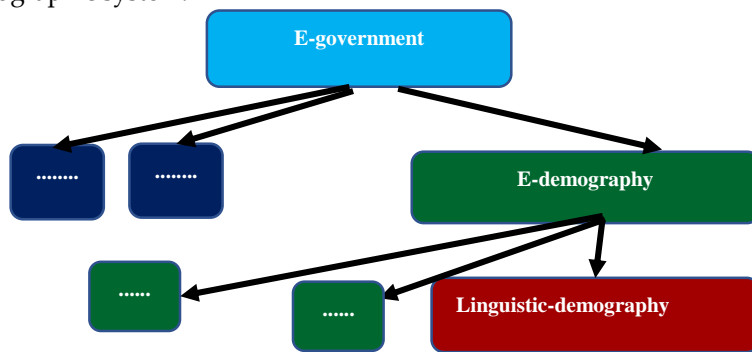


Fig. 2. The position of linguistic demography on the e-government platform in Azerbaijan

One of the chief units of an electronic demographic system is “on-line analytical processing”, i.e. OLAP technology, which is a key component of the DW (Alguliyev et al., 2019). OLAP technology has been widely used in both public and private organizations in recent years. OLAP technology is based on the representation of data in the form of OLAP cubes. They contain data that ensure a full image of the organization’s activities and are used for analysis and management decision-making.

Traditional OLAP typically includes the following operations: a) consolidation (roll-up) - an operation that determines the upward transition from a detailed data representation to an aggregated one; b) drill-down - a reverse operation of consolidation, defines a downward transition from an aggregated representation of data to a detailed one; c) slice; d) dice; e) pivot (Codd, 1993).

However, the upsurge in the amount of data and the enormous variety of data sources generated today have posed serious problems for traditional OLAP, as research has become limited. In this regard, a need to shift from traditional OLAP to smart OLAP will greatly expand the opportunities of OLAP. The benefits of smart OLAP lie in the use of data mining methods. For instance, the use of blockchain technology may provide data security, as mentioned in (Nabibayova, 2023, a). In addition, the approach of combining OLAP and neural networks for big data is very effective. Obviously, by learning complex relationships in data, neural networks can solve classification problems, as well as predict non-existent indicators. Apparently, the use of these methods and technologies for linguistic demography within the electronic demographic system containing big data is very beneficial.

As noted above, demography is an interdisciplinary science, due to which the DW of the electronic demographic system is broken down into several DMs. Each DM corresponds to a

separate area of demographics. One of them relates to linguistic demography (Figure 3).

Data on the distribution and use of languages is included in the DM of Linguistic Demography. Specifically, these data contain information on both the territorial distribution of languages and the number of native speakers by region.

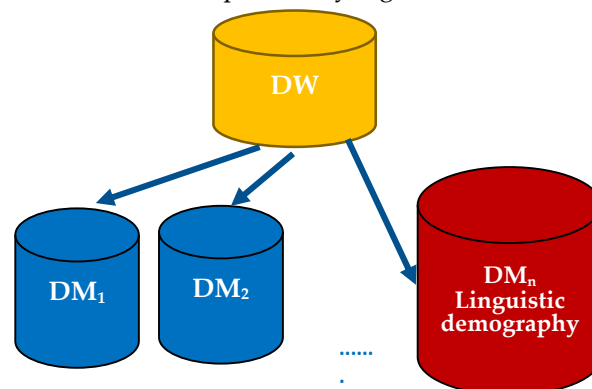


Fig. 3. DW structure

In this case, the regions include villages, cities, territorial administrative regions of the country.

### 3. The concepts of “language” and “dialect” can be defined differently”

A language, by definition, is a system of phonetic, lexical and grammatical means, which is a tool for delivering thoughts, feelings, wills and serving as the most important communication means among people.

There are currently more than 7 000 languages throughout the world. Some languages have become the utmost extensively used languages in the world. Figure 4 displays the positions of the 10 most spoken languages in the world (Languages of the world).

Currently, linguistic competition and the extinction of languages can be seen all over the world.

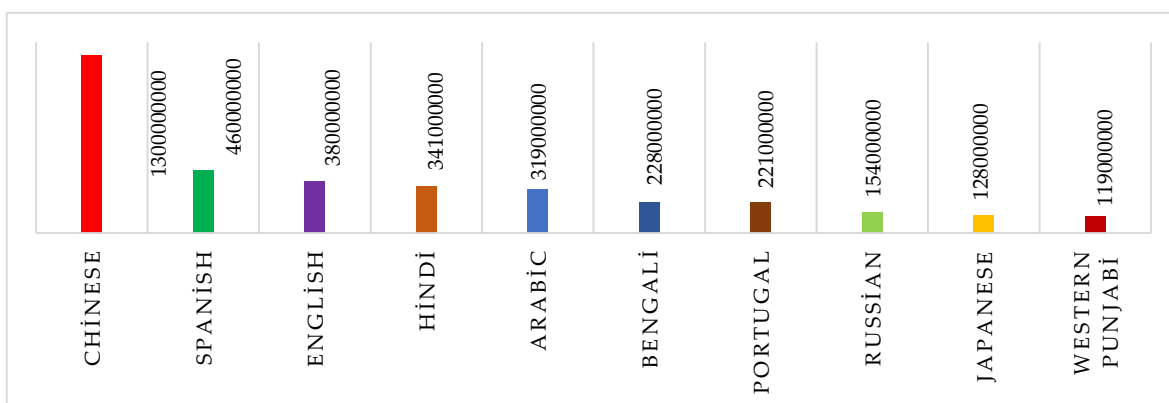


Fig.4. Top 10 most spoken languages in the world

Linguists think that at least half of some languages will extinct in the 21st century due to the extensive use of some languages, such as English, Chinese, Spanish, Indian etc.

Many different socio-economic, political and cultural factors that reduce linguistic diversity are available. These factors may include globalization, urbanization of the population, and migration. In this regard, groups speaking different languages have a need for a common language of communication. Some languages began to perform this function and therefore, their importance in official and unofficial communication has greatly increased (Kandler, 2009).

A dialect is a language characteristic of a certain area, community, or group, often with comparatively slight alterations in vocabulary, style, spelling, and pronunciation (Dialect). The need to study geographically determined features of the use of languages and pronunciation, that is, dialects, resulted in the rise of a special section of linguistics - dialectology.

Note that the dialects of a language may preserve the most ancient characteristics of the language, owing to which they become a real linguistic treasury. This also refers to the dialects of the Azerbaijani language, which are prevalent in the territory of Azerbaijan. In (Shiraliyev, 1967), the author classifies the dialects of the Azerbaijani language, based on the historical-geographical principle. According to this principle, dialects in Azerbaijan are classified into 4 groups: eastern (Kuba, Baku, Shamakhi dialects, Mughan and Lankaran accent), western (Kazakh, Karabakh, Ganja dialects and Ayrim accent), northern (Shaki dialect and Zagatala-Gakh accent) and southern (Nakhchivan, Ordubad, Tabriz dialects and Irevan accent).

But in the establishment of the main dialects, physical and geographical barriers, administrative divisions, centres uniting the surrounding population, substrative influence, population migration, and historical isolation play a certain role. Taking this into consideration, in (Azizov E., 1999), the author classifies the dialects of the Azerbaijani language into 3 groups: northeastern (covers 3 groups of dialects - Shamakhi-Baku, Kuba, Derbent), western or middle (covers 3 groups of dialects - Karabakh, Kazakh-Borchali, Ayrim accent) and southern (covers several groups of dialects - Tabriz, Ardabil, Lankaran, Urmia, Nakhchivan, etc.). In this research, the author, along with the main dialects, also shows their transitional types. Since the transitional forms of dialects are intermediate, i.e., they are on the border of two dialects, the properties of different

dialects are observed in these dialects.

In Azerbaijan, the study of the dialects of the Azerbaijani language is being addressed at the state level. By the decree of the President of the Republic of Azerbaijan Ilham Aliyev dated April 9, 2013, 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" was approved (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). According to the Action Plan for the realization of this program, the Article "Ensuring the study of various dialects and the dialects of the Azerbaijani language in accordance with the requirements of the modern era", the study of the dialects of the Azerbaijani language in accordance with the requirements of the time and the achievements of modern linguistics has become a significant mission.

Thus, there are certain problems associated with approximating the number of speakers of a certain language, linked with different approaches to this topic. Therefore, estimates can vary significantly.

The problems complicating the studies in this field primarily include:

- availability of different approaches to the definition of the notions of "language" and "dialect" –the definition of the status of a certain variety of a language as a separate language or as a dialect of this language is challenging; this problem refers to one of the multifaceted difficulties of linguistics;
- availability of multilingualism among some individuals, which obscures the definition of the concept of "native language" – there are individuals who fluently speak several languages, and therefore, it is difficult for them to determine their native language. Sometimes multilingualism becomes a consequence of the fact that several languages can be official languages at once in a country, for instance, Switzerland, Belgium, Luxembourg, etc.
- lack of sufficient population census data in many countries - in general, the study of the results of the population census contributes to better planning in areas such as transport, health, social services, economics, education, etc. The question of the status of a native speaker in the demographic data of the population

census is mandatory. Therefore, a regular census can determine the number of speakers of a specific language, including the dynamics of changes in this field.

The availability of different approaches to the explanations of the notions of “language” and “dialect” is directly associated to the availability of diverse points of view on these notions (Wichmann, 2019; Koryakov, 2017). When the notion of “language” is considered from a political point of view, then the language is the one that the people regularly speak as a nation. For instance, from 1850 to 1992 there was a Serbo-Croatian language, which included several dialects, as well as Serbian, Croatian and Bosnian. Nonetheless, following the breakup of Yugoslavia into several independent states in the mid-1990s, these dialects were recognized as separate languages. Spanish is another example. It is spoken in Spain and significantly differs from its Latin American version. Nonetheless, despite the fact that it may well be considered a distinct language, it is known as a dialect (Nabibayova G., 2023).

(Koryakov, 2017) describes the criteria for explaining the concepts of language and dialect. Main criteria are: a) general supra-dialect norm, b) ethnic orientation, c) opinion of speakers, d) official status.

In accordance with the first criterion, two idioms (an idiom combines the notions of language, dialect and accent) are estimated as the dialects of the same language if they have the same script differing from other languages; or, for unwritten languages (without scripts), a common language resulting from the mixing of several dialects and languages.

The second criterion focuses on the ethnicity of the speakers: if the speakers of two idioms refer to the identical ethnic group, then this is one language, if they vary, they belong to diverse ethnic groups. However, linguistic and ethnic boundaries too often do not coincide. For example, Switzerland has four traditional languages (German, French, Italian and Romansh), although ethnically their speakers consider themselves primarily Swiss.

In accordance with the third criterion, the opinion of native speakers is taken into account: if they consider their native speech to be a separate language, different from the languages of all neighbours, then the language they speak is a separate independent language.

The fourth criterion reviews a language as an idiom, the status of which is verified in some official document, for example, in the constitution, the language law, or other official acts of the

government.

Several other criteria on the concepts of language-dialect are available. These may include: a sufficiently large number of carriers, a large distribution area; high prestige; educational use; widespread use in media, on the Internet, etc.

To perform statistical studies in the field of linguistic demography, it is important to identify the available language indicator in the country, which can be set differently, namely as a native language, as a language of daily communication or as a language of use. Moreover, in order to make the results of these studies more reliable, demographers are recommended to ignore the language indicator of tourists, students from other countries and workers on a temporary contract.

At the session of the International Statistical Congress of 1872, St. Petersburg, it was recommended to use “native language” as a language indicator. However, this recommendation was not followed in four censuses in multilingual Austria, where the “language of use” was adopted as a language indicator. In the Czech Republic, the recognition of “language of use” as a language indicator has been challenged. This was due to the fact that many citizens of the Czech Republic who were born in this country and speak Czech since childhood, but live in German-speaking cities in Austria, may consider German as their “native language” (Nabibayova, 2023, b).

Note that not only multilingualism, but also the availability of several approaches to the notion of “native language” obscure the explanation of this notion.

The main approaches are as follows:

- *Native language is a language acquired in childhood without teaching.* A language that a person learns from childhood without special training, in a natural language environment.
- *Native language is functionally the first language.* The language in which they think automatically. In this language, a person easily expresses own thoughts in oral and written form, and this is the most convenient form of expression for him/her.
- *Native language is the language of ethnic identification.* The language of a nation or ethnic group linking a person to previous generations.

Note that the native language may not coincide with the nationality of a person. Such cases may often be found in the context of global migration processes of the 20th and 21st centuries (Nabibayova, 2023, b).

The consequences of migrations are contradictory and often unpredictable. A massive influx of migrants can cause various unpredictable consequences, which are often negative. This is due to the fact that being affected by migration, not only the social structure, ethnic composition and localization of the population change, but also the linguistic structure of society changes, which can lead to negative results as the loss of national identity, separatism, xenophobia, etc. (Yakovleva, 2017). In this regard, the linguistic aspects of migration processes require a serious attitude and scientific approach.

It is also noteworthy that recent studies in the field of linguistic demography have attracted more and more attention of demographers from the standpoint of sociolinguistics, which studies the relationship between the area of distribution of a language and the social conditions of its existence.

It is also noteworthy that recent studies in the area of linguistic demography have attracted more and more attention of demographers from the perspective of sociolinguistics, which explores the relationship between the area of distribution of a language and the social circumstances of its existence.

All of the above problems facing world demographers are also important for Azerbaijani demographers (Nabibayova, 2023, b). This is due to the fact that Azerbaijan is a multinational country, and multilingualism emerged here. Moreover, many dialects are practiced in Azerbaijan, as noted.

Thus, when entering data on linguistic demography into an electronic demographic system, a whole range of problems arises. To solve all these problems, the following recommendations are presented:

- to make a decision whether the dialect of the language will be accepted as another language or not;
- to determine which language is the mother tongue of multilingual people;
- to present a clear description of the concept of "mother tongue";
- to determine which language has to be used in the country as a language indicator, that is, native language, daily language or colloquial language;
- to hold a population census at least once every 5 years to identify the dynamics of the spread of languages and the number of speakers of different languages.

Obviously, the first four statements are addressed to linguists, and the last one to public authorities. The implementation of these proposals put forward will play an important role in the analysis and decision-making in the field of linguistic demography.

#### 4. Conclusion

Presently, information and communication technologies have been introduced into all spheres of human activity. Accordingly, in an area as demography, this manifested itself in the need to develop an electronic demographic system. Demography is a multidisciplinary science, which includes an area as linguistic demography. For analysis and decision making in this field, it is beneficial to apply OLAP technology, i.e., a key component of DW. This article emphasised the benefits of using smart OLAP for big data.

The article emphasized certain problems existing in the field of linguistics, which led to significant difficulties when entering data on linguistic demography into an electronic demographic system, and they are associated with different approaches of linguists to such concepts as language and dialect.

Other very significant problems were also revealed to arise when establishing a language indicator in the country. Additionally, the inevitability to give a clear explanation of the notion of "mother tongue" and to define the status of "mother tongue" was highlighted.

In regard to these difficulties, various questions arise, for example:

- a) Should the dialect of the language be recognized as a new language?
- b) What does the notion "native language" mean?
- c) What language should be recognized as a native language for a person speaking several languages or living in a country where several languages are official? etc.

To overcome these difficulties, this article presented recommendations. The execution of these recommendations will assist in entering linguistic demographic data into the electronic demographic system.

All this will enable analysts to make the right decisions based on the DW analysis.

## References

- Alguliyev R. M., Aliguliyev R. M., Yusifov F. F., Alakbarova I. Y. (2019). Formation of electronic demography as an effective tool for social research and monitoring of population data / Issues of state and municipal management. 4, 61–86. (in Russian)
- Azizov E. (1999). New classification of dialects of the Azerbaijani language / E. Azizov // Turkology. 352 p. (in Azerbaijani).
- Codd, E.F., Codd, S.B., Salley, C.T. (1993). Providing OLAP (on-line Analytical Processing) to User-analysts: An IT Mandate, Codd & Date, Inc., 32, 1-24.
- 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. <https://e-qanun.az/framework/25537> (in Azerbaijani).
- Demography. <https://www.cambridge.org/core/books/abs/essential-public-health/demography/CB1BF80C417CE4E5F980DF4D0A0D8D04>
- Dialect. <https://old.bigenc.ru/linguistics/text/4046120>
- Kandler A. (2009). Demography and Language Competition. Human Biology: 81(3), 181-210. <http://digitalcommons.wayne.edu/humbiol/vol81/iss2/5>
- Koryakov Yu. (2017). The problem of “language or dialect” and an attempt of a lexical and statistical approach // Problems of Linguistics. 6, 79-101.
- Languages of the world. <https://www.ethnologue.com/>
- Nabibayova G. (2023, a). Ensuring information security of OLAP-containing distributed electronic systems using blockchain technology Information Technologies. 29(5), 250-256. (in Russian)
- Nabibayova G. (2023, b). Issues of linguistic demography formation on the national e-government platform. [https://ict.az/uploads/konfrans/ana\\_dili/12\\_Issues\\_of\\_linguistic\\_demography\\_formation\\_on\\_the\\_national\\_e-government\\_platform.pdf](https://ict.az/uploads/konfrans/ana_dili/12_Issues_of_linguistic_demography_formation_on_the_national_e-government_platform.pdf)
- Pauwels A. (2016). Language Maintenance and Shift. Publisher: Cambridge University Press. p. 220
- Shiraliyev M. Fundamentals of Azerbaijani dialectology / M. Shiraliyev. Baku, 1967. 420 p. (in Azerbaijani).
- Wichmann S. (2019). How to Distinguish Languages and Dialects. Computational Linguistics, 45(4), 823–831. <https://aclanthology.org/J19-4007/>
- Yakovleva E. (2017). Migrations: a hypothetical view of the linguistic future of modern Europe. <https://cyberleninka.ru/article/n/migratsii-gipoteticheskiy-vzglyad-na-yazykovoe-budushee-sovremennoy-evropy> (in Russian)