

Issues of linguistic demography formation on the national e-government platform

Gulnara Nabibayova

Institute of Information Technology, Baku, Azerbaijan

gulnarara@mail.ru

Abstract— The emergence of the field of electronic demography has become a prerequisite for developing 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. In this regard, in the electronic demographic system, along with other segments, there is a segment of linguistic demography. Each area in the Data Warehouse (DW) has its own Data Mart (DM). This research emphasizes that when forming a DM of linguistic demography, it is necessary to solve some problems. These problems are related to the presence of different approaches to the definitions of the concepts of “language” and “dialect”, as well as to the establishment of a language indicator in the country, for which three options are possible - as a native language, as the language of everyday communication or as the language of use.

Keywords — *demography; Linguistic Demography; OLAP; electronic demographic system; Data Warehouse; Data Mart; native language; language of daily communication; language of use.*

I. INTRODUCTION

The implementation of a demographic policy aimed at improving the demographic

situation in the country is one of the important tasks facing each state. Demographic policy includes the solution of important tasks such as prolonging the life expectancy of the population, increasing the birth rate, reducing morbidity and mortality, regulating migration, etc. Demographic research plays an important role in solving these problems. The results of these studies constitute the basis of demographic policy.

Today, as a result of the impact of information technology on all spheres 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 e-government platform in Azerbaijan.

Demography (Greek δῆμος - people, Greek γράφω - I write) is a field of science that studies the patterns of population growth, its dependence on socio-economic conditions, natural factors, migration, population density and its distribution. Obviously, an important feature of demography is that it is an interdisciplinary science [1]. Demography is closely related to economics, political science, ethnography, medicine, geography, etc. In this regard, it is divided into a number of

specialized areas, each of which studies certain demographic processes.

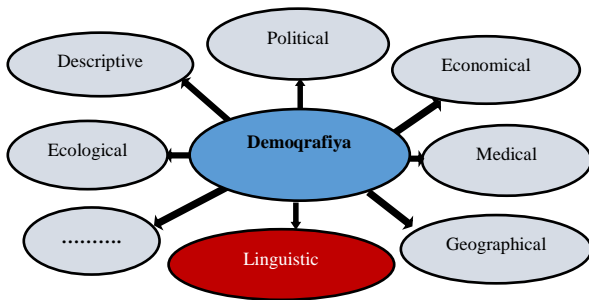


Fig. 1. Fields of demography

Fig. 1 shows the fields of demography, among which there is linguistic demography.

The emergence of electronic demography became a prerequisite for the need to develop an electronic demographic system [2]. One of the main unit in this system is “on-line analytical processing”, i.e. OLAP technology, which is a key element of the DW. OLAP technology is based on the representation of data in the form of OLAP cubes. They contain data used for analysis and decision making.

As noted above, demography is an interdisciplinary science, as a result of which the DW of the electronic demographic system is divided into several DM. Each DM corresponds to a distinct area of demographics. One of them deals with linguistic demography (Fig. 3).

II. FORMATION OF LINGUISTIC DEMOGRAPHY WITHIN THE ELECTRONIC DEMOGRAPHY SYSTEM

Linguistic demography is the statistical study of languages among all groups of populations. Fig. 2 illustrates the position of linguistic demography within the framework of national e-government.

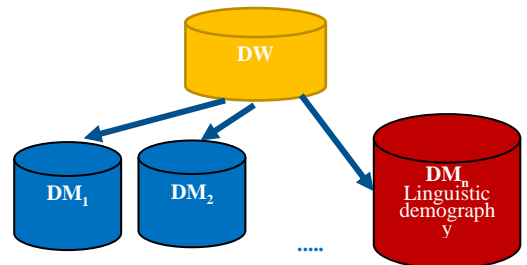


Fig. 3. DW structure

The data on the distribution and use of languages are included into the Linguistic Demographics DM. Specifically, these data include both the geography of the distribution of languages and the number of speakers of these languages.

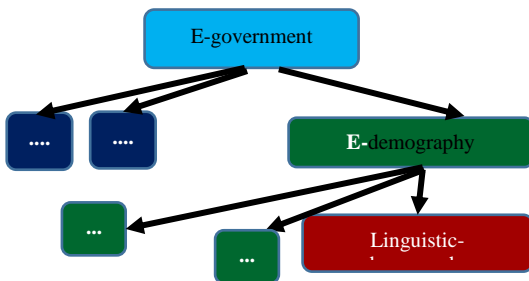


Fig. 2. The position of linguistic demography on the e-government platform in Azerbaijan

III. DIFFERENT APPROACHES TO THE DEFINITION OF THE CONCEPTS OF “LANGUAGE” AND “DIALECT”

There are currently over 7,000 languages in the world. Some languages have become the most widely used languages in the world. Fig. 4 displays the ranking of the 10 most spoken languages in the world [3]

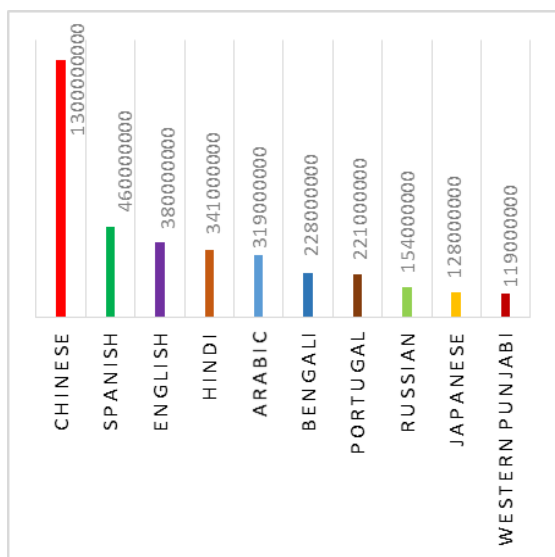


Fig.4. Top 10 most spoken languages in the world

It should be noted that the Azerbaijani language ranks 30th in the world in terms of the number of speakers. The area of distribution of the Azerbaijani language is Azerbaijan, Iran, Russia (Dagestan), Turkey, Georgia, and Iraq.

However, there are certain problems related to the estimation of the number of speakers of a particular language, and different estimates may vary significantly. The problems complicating the conduct of

research in this field primarily include: a) availability of different approaches to the definition of the concepts of “language” and “dialect”; b) availability of multilingualism among some individuals, which obscures the definition of the concept of “native language”; c) and the lack of sufficient population census data in many countries.

The availability of different approaches to the definitions of the concepts of “language” and “dialect” is directly related to the presence of different points of view on these concepts [4, 5]. If the concept of “language” is considered from a political point of view, then the language is the one that the people usually speak as a nation. For example, from 1850 to 1992 there was a Serbo-Croatian language, which included several dialects, including Serbian, Croatian and Bosnian. Nevertheless, following the breakup of Yugoslavia into several independent states in the mid-1990s, these dialects were recognized as separate languages. Another example is Spanish. The Spanish spoken in Spain significantly differs from its Latin American version. Nonetheless, despite the fact that it may well be considered a separate language, it is recognized as a dialect.

In order to carry out statistical studies in the field of linguistic demography, it is necessary to determine the current 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. Note that 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”.

It should be noted that not only multilingualism, but also the presence of different approaches to the concept of “native language” complicate the definition of this concept.

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 opinion 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.

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.

CONCLUSION

Due to the existing problems listed above, there are some difficulties when entering data on linguistic demography. For example: a) Should the dialect of the language be accepted as a new language? b) What does the term “native language” mean? and so forth.

All of the above problems facing world demographers are also important for Azerbaijani demographers.

In this regard, certain problems arise when entering data on linguistic demography into an electronic demographic system.

These problems may include:

- Whether the dialect of the language is accepted as another language or not

- Identification of native language of multilingual people
- Definition of the concept of “native language”.
- The need to conduct a census at least once every 5 years to determine the dynamics of the distribution of languages;
- What should be used as a language indicator - native language, everyday or spoken language.

These issues need to be addressed as soon as possible.

REFERENCES

- [1] <https://www.cambridge.org/core/books/essential-public-health/demography/CB1BF80C417CE4E5F980DF4D0A0D8D04>
- [2] Alguliyev R. M., Aliguliyev R. M., Yusifov F. F., Alakbarova I. Y. Formation of electronic demography as an effective tool for social research and monitoring of population data / Issues of state and municipal management. 2019. No. 4. pp. 61–86
- [3] <https://www.ethnologue.com/statistics/>
- [4] Soren Wichmann. How to Distinguish Languages and Dialects. Computational Linguistics, Volume 45, Issue 4 December 2019 <https://aclanthology.org/J19-4007/>
- [5] Koryakov Yu. The problem of “language or dialect” and an attempt of a lexical and statistical approach // Problems of Linguistics. 2017. No 6, pp. 79-101