Nakh-Daghestanian Languages: Digital Resources and some Examples of their Use

Nina Dobrushina (Lyon)

The Nakh-Daghestanian language family is the biggest in the Caucasus in terms of different languages and language branches. In the last decade, there have been several collaborative projects aimed at documenting and systematically studying these languages. As a result, a number of new resources on Daghestanian population, multilingualism, vocabulary, and grammar emerged. In this paper, I present the resources that are in various stages of development.

I start with a brief introduction into the area and its languages (Section 1), outline the demographic database of Daghestanian villages (Section 2), the database of Daghestanian multilingualism (Section 3), the Daghestanian loans database (Section 4), the database of Daghestanian lexemes that originate from Arabic, Persian and Turkic languages (Section 5), the database of Swadesh-100 wordlists from the languages of Daghestan (Section 6), the Typological Atlas of the Languages of Daghestan (Section 7), and the database of Rutul dialects (Section 8). I also provide a list of electronically accessible corpora and dictionaries of Nakh-Daghestanian languages.

1. Daghestan and its languages

The Republic of Daghestan is a federal subject of Russia located in the North Caucasus. It has borders with Georgia, Azerbaijan and Chechnya. The relatively small area of mountainous Daghestan (about 50 000 km^2) is inhabited by speakers of more than forty languages. They belong to the East Caucasian (alias Nakh-Daghestanian) language family, but there are also speakers of three Turkic and one Iranian language (Kumyk, Nogai, Azerbaijani and Tat, respectively).[1] Today, most Daghestanians also speak Russian.

What makes Daghestan especially attractive for all kinds of linguistic research is that, at least in the villages, there are no signs of language shift. In addition to their relative isolation by mountain ranges, rivers and other geographical obstacles, an important factor which preserves the linguistic diversity of Daghestan is the traditional ethnic endogamy within the villages which was typical of most parts of Daghestan.[2]

Language population sizes range from less than one thousand speakers (Hinuq, Mehweb, Tukita) to hundreds of thousands (Avar, Lezgian, Lak). Recent estimates of the population size of Daghestanian languages are problematic, since they are usually based on official censuses, which are not fully reliable in this respect. For example, according to the census of 2010 there were 128 people who spoke Godoberi, while Mehweb was not present in the census at all. However, as we know from field research, there are about 3,000 people speaking Godoberi, and about 800 speakers of Mehweb.

For this reason, the attempt to estimate language population sizes on the reported sizes of villages was undertaken. Due to the strict endogamy, Daghestanian villages are ethnically and

[1] For more details see Koryakov 2002.
[2] Cf. Comrie 2008; Dobrushina 2023.

linguistically very homogeneous. This means that the size of a village can serve as a reliable means of estimating the number of speakers, if all villages are linked to languages.

Since the mid-20th century, the population of some villages has undergone some changes. Some villages decreased or were even depopulated completely because of forced or voluntary relocations to the lower lands. At the same time, the population of many villages has grown as the living conditions improved. Hence it is important to have the data on the village population before intensive urbanisation has started. For this reason, several historical censuses were digitised and put online in the form of the demographic database of Daghestan.

2. The demographic database of Daghestanian villages

The demographic database of Daghestanian villages[3] is a result of digitisation of several sources, including rural registers from 1886 and 1895 and national censuses of 1926 and 2010:

- Cities and districts of the Dagestan *oblast*, a corpus of statistical data of the population of the Transcaucasian territory, taken from the household registers of 1886 and published in 1893;[4]
- Notebook of the Dagestan *oblast*, compiled by E. I. Kozubsky and published in 1895;[5]
- List of population centers of the Dagestan ASSR, from the census of 1926, published in 1927.[6]
- Census microdata from the All-Russian National Census of 2010.[7]

The registers of 1886 and 1895 and the 1926 census were digitised by students of the HSE University (School of Linguistics) and then reviewed, completed, and brought up to date by Yu. B. Koryakov. They were lastly equipped with search capabilities and infographics by programmers D. Staferova and A. Belokon.[8]

Besides the population of all villages, the sources contain information about their ethnic composition: for 1895, only the dominant ethnicity (occasionally data of other large groups is given in the commentary); for 1886, 1926, and 2010, exact numbers of every ethnic group in a given village are provided.

The first source is made up of the most important information from household registers gathered in 1886 by decree of the State Council mainly for army drafts and tax collection. This source includes data on the general size of the population of each town, as well as divisions by nationality, religion, and class.

The second source (the 1895 notebook) contains population data of villages which was taken from official sources dating to 1894, and data on ethnicities which was collected from literature and archives, as well as in part from the testimony of local experts. In addition, many local versions of place names and information about affiliation with traditional communities and landholdings are given.

[3] https://multidagestan.com/census.
[4] *Svod* 1893: VI.
[5] Kozubsky 1895.
[6] Materialy 1927.
[7] Provided by the Main Interregional Center of Rosstat. The information was retrieved by Y. Koryakov from the site https://std.gmcrosstata.ru/ , which is no longer available today (as of 21.12.2023).
[8] Koryakov et al. 2019.

The third source is compiled on the basis of the results of the first all-union census of 1926. The official day of the census throughout the entire USSR was 17 December. In the 1926 census, the Andi-Tsez nationalities had not yet been merged with the Avar, nor had the Kaytagians and Kubachinians with the Darginians.

Finally, material from the 2010 census was taken from the database of microdata of the Main Interregional Center of Russia, which provides information on ethnic composition, native language, and first and second languages in every population center. A downside of these data is the artificial skewing of all numerical data within a small range (±5 people) for the purpose of preserving anonymity.

It is important to understand that the area of Daghestan was considerably smaller at the end of the 19th century than it is today. In the beginning of the 1920s, the Kizlyarsky and Khasav-Yurtovsky districts and the Achikulaksky region were added to it. Later, part of the Achikulaksky region was returned to the Stavropol Territory. Only those towns which were within Dagestan in 1886 and 1895 are in the database, including a few towns which today (as in 1926) are located in Chechnya (the Camalal-speaking town of Kenkhi and neighbouring villages) and Azerbaijan (the Lezgian towns on the southern slopes of the Caucasus).

During the review process, towns from all four sources were tied together by a single index. The basis for matching towns from different years was their geographic location. Thus, if a town moved partly or completely to a new place (at a distance of more than 2–3 kilometers), the editors of the database considered them to be two different towns, even if their names were the same. A few villages match with two villages in past censuses and vice versa. More often than not this occurs with ‘Upper’ / ‘Lower’ village pairs, but occasionally two closely located villages merge into one. Sometimes for two old villages, one contemporary town is indicated and vice versa (especially when they are located very near to one another geographically), but more frequently the modern merged town is matched with just one of the old villages, and merely the coordinates are given for the second.

Many villages no longer exist and their names may be indicated on maps along with ‘des.’ (deserted), with a symbol with the label ‘ruins’, or simply by the word ‘landmark [name]’ without an indication of the exact location, in those cases when the settlement did not leave even a trace behind with the exception of the topographical place name. Along with other sources, old topographical military maps of the Worker-Peasant Red Army were used.

The creation of this database made it possible to study the dynamics of the population of Daghestanian languages across 150 years with great accuracy.[9] Since Daghestanian villages are ethnically and linguistically homogeneous, the population of villages provides very good estimates for the number of language speakers starting from the end of the 19th century.

High linguistic density in a small territory was the reason why, before the advent of Russian, many inhabitants of Daghestan spoke several languages. The results of a large-scale field study of Daghestanian multilingualism are presented in the Atlas of multilingualism in Daghestan.

3. Atlas of multilingualism in Daghestan

While the information on the language population was retrieved from official sources (even if indirectly, through the population of villages), the quantitative data on the multilingualism of the Daghestanian population can be only collected by means of field research. The database

[9] See Dobrushina & Moroz 2021.

'Atlas of multilingualism in Daghestan' (MultiDag) is based on an extensive field study of the language repertoires of the residents of rural highland Daghestan.[10]

The Atlas makes a contribution to the study of the phenomenon which is referred to as 'small-scale multilingualism'.[11] Investigating situations of small-scale multilingualism is important to reconstruct social conditions that favored linguistic diversity in the precolonial world. Another emergent reason is the fact that the traditional multilingual settings are highly endangered. Indeed, the competence in small local L2 is rapidly displaced by the usage of lingua franca - Russian in case of Daghestan.

Daghestanian multilingualism is discussed in some detail in several surveys[12] but was never studied systematically, i.e. by comparing various locations, according to a similar set of parameters, and on the basis of quantitative data. The project described in this section aims at collecting quantitative data about the multilingualism of Daghestanians in a representative set of locations across the region and presenting it in the form of maps and descriptions. The project was launched in 2009 and is run by a large team of researchers collecting and processing data.[13]

A research of traditional bilingualism in Daghestan needs data on local multilingual patterns from the period before the spread of Russian. This period, which can be called 'traditional bilingualism', provides a window into the sociolinguistic past of the region. There are some documentary sources, such as works and notes by Uslar, Dirr and other Russian researchers and civil servants who worked in the Caucasus in the 19th century.[14] There are also several reports on bilingualism by Soviet anthropologists.[15] All such documents, however, provide qualitative and fragmentary information. The information is usually given in the form of general observations and assessments such as "During the historically accessible period, the Tsakhurs are in a state of a permanent Tsakhur-Azerbajani bilingualism",[16] without providing specific quantitative details, such as counts of people who had a command of specific languages as L2.

The data of the Atlas are meant to be both quantitative and diachronic. They are focused on specific locations, often remote from administrative centers, and target areas of contact between major languages, a major language and a minority language or between several minority languages (or, sometimes, dialects). They are collected by the method of retrospective family interviews, specifically designed to obtain the quantitative data about multilingualism in the past.[17] The method suggests that the respondents are interviewed about language inventories,

[10] See https://multidagestan.com/.
[11] Cf. Lüpke 2016; Pakendorf et al. 2021.
[12] Cf. Wixman 1980; Chirikba 2008; Magomedkhanov 2008; Nichols in preparation.
[13] Anna Aksenova, Anastasija Alekseeva, Aleksei Baklanov, Darya Barylnikova, Nikita Beklemishev, Zinaida Budilova, Ilya Chechuro, Anastasija Cheveleva, Maria Chudnovskaja, Mikhail Daniel, Faina Daniel, Yuliy Daniel, Nina Dobrushina, Anna Dyachkova, Aleksej Fedorenko, Anastasija Fedorenko, Konstantin Filatov, Dmitry Ganenkov, Anastasija Ivanova, Polina Kasyanova, Aleksandra Khadzhijskaya, Aleksandra Konovalova, Kirill Koncha, Elizaveta Kozhanova, Aleksandra Kozhukhar, Semen Kudriavtsev, Marina Kustova, Yury Lander, Yevgeniy Lapin, Aleksandr Letuchiy, Timur Maisak, Aleksandra Martynova, Maksim Melenchenko, Stepan Mikhailov, George Moroz, Valeria Morozova, Yevgenij Mozhaev, Timofey Mukhin, Polina Nasledskova, Ivan Netkachev, Elena Nikishina, Aleksandr Orlov, Ilya Sadakov, Olga Shapovalova, Semen Sheshenin, Aleksandra Sheshenina, Maria Sheyanova, Aleksandr Shiryaev, Mikhail Sonkin, Samira Verhees, Alexandra Vydrina, Egor Yatsishin, Aigul Zakirova, Pavel Zavjalov.
[14] E.g. in the journals *Sbornik svedeniy* and *Sbornik materialov*; cf. *Gorcy* 1992.
[15] Genko 2005; Lavrov 1953, 1978; Volkova 1967.
[16] "С исторически обозримого времени цахурский этнос находится в состоянии перманентного цахурско-азербайджанского двуязычия" (Magomedkhanov 2008: 47).
[17] Dobrushina 2013.

both their own and those of their elder – often deceased – relatives. Only those relatives whom they claim they remember clearly, are added to the database.

Interviews about relatives allow us to reach back into the 19th century, starting with people born around 1850, with more dense data from the 1880s on. This time span covers the situation typical of the village before the drastic social changes of the 20th century.

The interviews were all held in Russian. In very few cases – when the respondent had a really poor command of Russian – the communication via an interpreter, most often a younger relative, took place. The answers were put together in a table (spreadsheet) and aggregated.

The data were collected by Nina Dobrushina and Michael Daniel in 2009–2012, by Dmitry Ganenkov in 2013, and by Nina Dobrushina leading a field team of students in 2013–2022.[18] At the time when this paper is being written, the data from 69 villages have been put online.[19]

The villages form geographic clusters of two to four adjacent villages with two to four native languages per cluster. The cluster did not necessarily include villages whose socioeconomic relations were especially tight. A cluster is a unit of analysis more than a real social unit. The idea behind studying clusters rather than individual villages is that bilingualism is (at least) a binary relation between two (or more) ethnic groups. Counts of bilingualism within one of the groups only cannot provide for a robust sociolinguistic interpretation of the patterns of interethnic communication unless complemented by similar counts from the other group(s). All adjacent villages in Daghestan are closely related in terms of socioeconomic interaction. Each pair of villages in our clusters neighbour one another. The villages are usually within 20 to 90 minutes of walking distance.

The site of MultiDagcontains the database on Daghestanian multilingualism with a search interface. The users can use different parameters to build their own graphs and diagrams, such as particular villages, neighbourhoods, years of birth, genders, native languages, and second languages. The online database is constantly updated.

The results of this study were used in several papers where the patterns of Daghestanian multilingualism were analysed[20] and compared with the results of the study of lexical borrowings,[21] presented in the next section.

5. The database of Daghestanian loanwords

It is known that multilingualism is often the cause of language change – convergence of languages in the domains of vocabulary, phonetics and grammar. The ‘DagLoans’ project emerged as an attempt to test the correlation between the number of borrowings from different languages and the intensity of multilingualism among the speakers of the same languages.[22] A database of borrowed words collected in a number of villages in Daghestan is available online.

The authors look into lexical transfer between different Daghestanian languages at a microlevel, i.e. at the level of granularity that is sensitive to differences between village varieties.[23] For this purpose, a fixed shortlist of some 160 concepts was compiled, and a protocol for quick data collection in the field was developed. Using a fixed list of concepts for comparison allowed the

[18] Cf. footnote 13 above.
[19] Dobrushina et al. 2017.
[20] Dobrushina 2013; Dobrushina et al. 2019; Dobrushina & Kultepina 2021.
[21] Daniel et al. 2021.
[22] http://lingconlab.ru/dagloans/.
[23] Chechuro et al. 2019.

authors to quantify lexical transfer and to look for correlations with qualitative differences between areas, such as the spread of a certain lingua franca, the presence and degree of contact with particular languages, as well as migratory processes.

Collecting data in adjacent villages allows researchers to show variation between villages on the map. This reveals the contours of various zones of influence for specific L2s. For example, lexical influence from local Turkic languages (Azerbaijani, Kumyk and Nogai) is found throughout Daghestan. In the south, however, where Azerbaijani served as *lingua franca* for a very long time, this influence is much stronger. In the north of Daghestan, bilingualism in Turkic languages was not so common, and almost all Turkic borrowings in minor local languages in the list seem to be shared with Avar, a major native language.[24]

The 'Dagloans' database was aimed at detecting loans from local languages spoken in adjacent communities. In contrast, the DAG < APT database (see next Section) studies the loans from the languages which were important in the domains of religion, literacy and culture.

6. The DAG < APT database

The DAG < APT database[25] contains Daghestanian lexemes that originate from Arabic, Persian and Turkic languages, as well as the original donor lexemes.

Arabic, Persian and Turkic languages are prolific borrowing sources for Daghestanian languages due to their cultural importance. Arabic is the language of religion, and Persian and Turkic were important languages for trade and exchange of knowledge. Of course, these languages have also influenced each other, so a borrowing of an Arabic lexeme into a Daghestanian language may have been mediated through Persian and/or a local Turkic language like Azerbaijani.

DAG < APT consists of three parts. The main database is a collection of target lexemes and their origins extracted from available literature. The second part is an overview of attested donor lexemes and their translations. Finally, there is a list of sources on borrowings in Daghestanian languages.

The authors' aim is, first, to combine information from the rich literature on lexical borrowing from Arabic, Persian and Turkic into various Daghestanian languages into a single, searchable database that can be used for reference. The second aim is to create a base of target lexemes that can be compared in terms of adaptation patterns and geographical distribution.[26] This can help to uncover different historical and regional layers of borrowing processes, and perhaps identify cases of mediated borrowing.

While both the DagLoans and DAG < APT projects address loanwords and aim at identifying patterns of language contact and their result, the database of Swadesh lists looks at stable vocabulary.

7. The database of Swadesh-100 word lists from the languages of Daghestan

'DagSwadesh (Swadesh lists with village granularity)' is a database of Swadesh-100 wordlists from the languages of Daghestan.[27] The main aim of the project is to substantiate a

[24] See Daniel et al. 2021.
[25] https://lingconlab.github.io/DAG-APT/.
[26] Dedov & Verhees 2022.
[27] http://lingconlab.ru/dagswadesh/.

lexicostatistical classification of the East Caucasian idioms by controlling the provenance of the wordlist at the village level including data from reliable sources[28] and collecting new data ourselves.[29] The recordings fill gaps in the existing datasets, and account for eventual differences between villages speaking what is conventionally seen as one and the same language. As of 2020, the database presents vocabulary from 21 village varieties, belonging to the Avar-Andic branch of the East Caucasian family.[30]

Users can browse DagSwadesh in search mode (in case they are interested in investigating a list from a particular idiom or in inspecting a cognate set) or in comparative mode (in case they want to compare concepts across different idioms). All the data are also available for downloading in TSV CLDF-compatible format.

The DagSwadesh project allows linguists to clarify the genealogical classification of not only languages, but also dialects, and to raise questions related to the correlation of linguistic and geographical distance. In contrast to this, the Typological Atlas of the Languages of Daghestan aims at a typological comparison of the languages of Daghestan, primarily on the domains of grammar and phonology.

8. Typological atlas of the languages of Daghestan

The systematic comparison of phonetic, lexical, and grammatical features of the languages of Daghestan is reflected in the form of 'DagAtlas', a typological atlas of the languages of Daghestan.[31] In this database, the languages spoken on the territory of Daghestan (including Turkic languages, Armenian and Georgian) are compared on the basis of a number of features which are displayed on maps in a uniform way. The authors aim at developing a tool for the visualisation of information about linguistic structures that are characteristic of Daghestan.[32] The atlas is based mainly on data from published grammars, and can therefore be used for bibliographical research and as a source of references on parameters of interest. A key task of the project is the creation of maps and visualisations that allow the researcher to combine metadata and genealogical parameters with information on a particular feature.

The version of the TALD released in July 2022 contained 28 chapters on such topics as the phonological inventory, evidentiality, standards of comparison, ordinal numerals, optatives, prohibitives, caritives, adpositions, etc. Each chapter consists of the description of the feature, the discussion of its variation in the languages of Daghestan, and its geographical and genealogical distribution.

Being one of the most linguistically diverse parts of Russia, Daghestan is emerging as an extremely important area for investigating the role of language contact in the rise, and the development and distribution of certain linguistic structures. Data from the Atlas can be used to formulate hypotheses about the area and scenarios for the distribution of certain phenomena. For example, the chapter on optatives shows that Nakh-Daghestanian languages typically have dedicated inflectional optatives. It also shows that optatives which are based on imperatives are typical for Avar-Ando-Tsezic languages, and do not occur in the south of Daghestan.

[28] Such as Kibrik & Kodzasov 1988, 1991.
[29] Filatov & Daniel 2022.
[30] The results of this study were used in Koile et al. 2022.
[31] http://lingconlab.ru/dagatlas/.
[32] Daniel et al. 2022.

The Typological atlas of the languages of Daghestan is intended as a dynamic resource, whose contents are constantly updated by adding new topics. The project of a documentation of the dialectal variation of the Rutul language (see next Section) is conceived in a somewhat similar way, though with a much higher level of granularity and on a much smaller geographical scale.

9. Dialectological atlas of Rutul

The database of Rutul dialects is based on a field survey of twelve villages, and is in the process of being put online in the form of an atlas where each feature is reflected on a map.[33] Twelve Rutul villages are located in the Rutulsky and Akhtynsky districts of the Republic of Daghestan, Russian Federation. These settlements are Amsar, Dzhilikhur, Ikhrek, Kala, Khnov, Kiche, Kina, Kufa, Luchek, Myukhrek, Rutul, and Shinaz.

The Rutul language[34] belongs to the Lezgic branch of the Nakh-Daghestanian (East Caucasian) language family. The majority of speakers (approximately 33,000, according to the latest census of 2020) live in Daghestan.

The dialectological survey was carried out in July 2022 by a group of linguists from HSE University.[35] Speakers of various Rutul dialects were interviewed using the same questionnaire compiled by the research team. The questionnaire was partly based on the existing descriptions of the varieties of Rutul. In addition to documentation and quantitative assessment of the established parameters of dialectal variation, the authors aimed at identifying new, previously unreported regional variants.

For each phenomenon selected for the comparison, several Russian stimuli were compiled. Translations of these sentences by two, three, or four speakers in each of the villages in the survey were recorded and transcribed. Dialectal differences at all levels of language structure, including lexicon, phonetics, morphology, and syntax were investigated.

The aim was a systematic mapping of dialectal variation of Rutul in a set of pre-selected parameters. As expected, in many cases, variation in the realisation of a phenomenon was attested. The variation could be observed not only between villages, but also between villagers (different speakers from the same village) and also at the intra-speaker level (in the data recorded from an individual speaker). Such cases are reflected on the map as circles containing sectors of different colors. The relative frequency of variations is not reflected on the maps. The project will make it possible to carry out a quantitative comparison of divergence between Rutul dialects.

10. Summary

Although most languages of Daghestan are still spoken and transmitted to children, the sociolinguistic situation in the region changes very quickly. The author of this paper saw children who prefer speaking Russian rather than their native languages even within their village, which was unimaginable until the very recent past. Therefore, the documentation of local languages becomes an increasingly urgent task.

The present survey of digital resources of the languages of Daghestan reflects only a small part of the work currently being done by linguists. Numerous efforts lie in the domain of creating

[33] https://lingconlab.github.io/rutul_dialectology/.
[34] ISO 639–3 rut; glottocode rutu1240.
[35] Alekseeva et al. 2023.

digital dictionaries and corpora. Currently available online are the corpora of Standard Avar, Aghul, Archi, Bagvalal, Dargwa, Sanzhi Dargwa, Godoberi, Lak, Tabasaran, Tsakhur, Tsez, dialectal Lezgian, most of them rather small (see the list of corpora in Appendix 1). There are also numerous collections of texts,[36] as well as dictionaries, available only in printed form, but also electronic dictionaries of the Kina dialect of Rutul, Mehweb, Tukita and several varieties of Dargwa (see Appendix 2).

Appendix 1. Electronically available corpora of Nakh-Daghestanian languages

Avar Text Corpus https://baltoslav.eu/avar/

Aghul corpus http://web-corpora.net/AghulCorpus/search/

Archi corpus http://web-corpora.net/ArchiCorpus/search/

Archi corpus http://www.philol.msu.ru/~languedoc/eng/archi/corpus.php

Avar corpus http://web-corpora.net/AvarCorpus/search/

Bagvalal corpus http://web-corpora.net/BagvalalCorpus/search/

Dargwa http://web-corpora.net/VanDenBergDargwaCorpus/search/?interface_language=en

Kadar Dargwa http://lingconlab.ru/kadar_dargwa/

Muira Dargwa http://lingconlab.ru/muira_dargwa/

Sanzhi Dargwa corpus http://web-corpora.net/SanzhiDargwaCorpus/search/

Sanzhi Dargwa https://multicast.aspra.uni-bamberg.de/ - Haig, Geoffrey & Schnell, Stefan (eds.). 2022. *Multi-CAST: Multilingual corpus of annotated spoken texts.* Version 2207. Bamberg: University of Bamberg. (multicast.aspra.uni-bamberg.de/)

Godoberi http://web-corpora.net/GodoberiCorpus/search

The Parsed Corpus of Modern Lak http://erwinkomen.ruhosting.nl/lbe/crp/

Tsakhur corpus http://web-corpora.net/TsakhurCorpus/search/

Tabasaran https://multicast.aspra.uni-bamberg.de/ - Haig, Geoffrey & Schnell, Stefan (eds.). 2022. *Multi-CAST: Multilingual corpus of annotated spoken texts.* Version 2207. Bamberg: University of Bamberg. (multicast.aspra.uni-bamberg.de/)

Tsnal Lezgi http://lingconlab.ru/tsnal_lezgi/

The Tsez Annotated Corpus Project https://tsezacp.clld.org/ - A.K. Abdulaev, I.K. Abdullaev, André Müller, Evgeniya Zhivotova, & Bernard Comrie. (2022). The Tsez Annotated Corpus Project (v1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7096350

Appendix 2. Electronically available dictionaries of Nakh-Daghestanian languages

Avar http://avar.me/

Agul https://www.webonary.org/aghul/en/overview/copyright/ - Roman Kim (ed.). 2016. Aghul-Russian Dictionary. Pathways. Research and Development.

Dargwa http://lingconlab.ru/dargwa_dict/

Dargwa https://www.webonary.org/dargan-mez/ Sychev, Sergey N. 2019. "Dargan-Russian-English Dictionary." *Webonary.org*. SIL International.

Mehweb http://lingconlab.ru/MehwebDict/ - A. Musaev, V. Morozova, M. Daniel. Mehweb Jena wordlist. 2020. Moscow: Linguistic Convergence Laboratory, HSE University.

Rutul https://linghub.ru/rutuldict/

Sanzhi-Dargwa https://dictionaria.clld.org/contributions/sanzhi

Tabasaran. https://www.webonary.org/tabasaran - Shikhalieva, Sabrina. 2019. "Tabasaran - Russian Dictionary." *Webonary.org*. SIL International.

[36] Cf., e.g., https://proclac.cnrs.fr/projets/projets-turcaucase/immocal/.

Tsakhur https://www.webonary.org/tsakhur/ - Sackett, Kathleen; Shamkhalov, Magommedsharif; Davudov, Axmed; Ismayilov, Nusrat; Shamkhalov, Vugar; and Agalarov, Magommed (eds.). 2022. "Tsakhur - Azerbaijani - Russian - English Dictionary." *Webonary.org*. SIL International.

Tukita http://lingconlab.ru/TukitaDict/ - Magomedgazhieva, P. and M. Daniel (2023). *Dictionary of Tukita (v2.0.0)*. Moscow. DOI: 10.5281/zenodo.7803955.

Shiri. https://www.webonary.org/shiri/ - Oleg Belyaev (ed.). 2020. *Shiri – English Dictionary*. Otto-Friedrich Universität Bamberg.

References

Alekseeva et al. (2023): Asya A., Nikita Beklemishev, Michael Daniel, Nina Dobrushina, Konstantin Filatov, Anastasiya Ivanova, Timur Maisak, Maksim Melenchenko, George Moroz, Ivan Netkachev, and Ilya Sadakov, *Atlas of Rutul dialects*. Moscow: Linguistic Convergence Laboratory, HSE University. https://lingconlab.github.io/rutul_dialectology/index.html.

Chechuro et al. (2019): Ilya Ch., Michael Daniel and Samira Verhees, "Daghestanian loans database". Linguistic Convergence Laboratory, HSE. http://lingconlab.ru/dagloans/.

Chirikba (2008): Viacheslav A. Ch., "The problem of the Caucasian Sprachbund". In: Pieter Muysken (ed.), *From linguistic areas to areal linguistics*. Amsterdam / Philadelphia: John Benjamins, 25–94. https://www.academia.edu/570408.

Comrie (2008): Bernard C., "Linguistic Diversity in the Caucasus". *Annual Review of Anthropology* 37, 131–143. https://www.annualreviews.org/doi/pdf/10.1146/annurev.anthro.35.081705.123248.

Daniel et al. (2021): Michael D., Ilya Chechuro, Samira Verhees and Nina Dobrushina, "Lingua Francas as Lexical Donors: Evidence from Daghestan". *Language* 97/3, 520–560. https://muse.jhu.edu/article/806347/pdf; https://doi.org/10.1353/lan.2021.0046.

— (2022): Michael D., Konstantin Filatov, George Moroz, Timofey Mukhin, Chiara Naccarato, and Samira Verhees, *Typological Atlas of the Languages of Daghestan (TALD)*, v. 1.0.0. Moscow: Linguistic Convergence Laboratory, NRU HSE. http://lingconlab.ru/dagatlas/; DOI: 10.5281/zenodo.6807070.

Dedov & Verhees (2022): Timofei D. and Samira V, "*DAG < APT, An online database of borrowed vocabulary in the languages of Daghestan*, v. 1.0.0. Moscow: Linguistic Convergence Laboratory, NRU HSE. https://lingconlab.github.io/DAG-APT/.

Dirr, Adolf. 1906. Kratkij grammatičeskij očerk andijskogo jazyka s tekstami, sbornikom andijskix slov i russkim k nemu ukazatelem. In: *SMOMPK*, Issue 36. Part IV. Tiflis: Upravlenie Kavkazskogo uchebnogo okruga.

Dobrushina (2013): Nina D., "How to study multilingualism of the past. Investigating traditional contact situations in Daghestan". *Journal of Sociolinguistics* 17/3, 376–393. https://onlinelibrary.wiley.com/doi/full/10.1111/josl.12041.

— (2023): Nina D., "Language ideology in an endogamous society: the case of Daghestan". *Journal of Sociolinguistics* 27/2, 159–176. https://www.researchgate.net/publication/ 364447448.

Dobrushina & Kultepina (2021): Nina D. and Olga K., "The rise of a lingua franca: the case of Russian in Dagestan". *International Journal of Bilingualism* 25/1, 338–358. https://journals.sagepub.com/doi/10.1177/1367006920959717; https://doi.org/10.1177/1367006920959717.

Dobrushina & Moroz (2021): Nina D. and George M., "The speakers of minority languages are more multilingual". *International Journal of Bilingualism* 25/4, 921–938. https://journals.sagepub.com/doi/10.1177/13670069211023150; https://doi.org/10.1177/13670069211023150.

Dobrushina et al. (2017): Nina D., Daria Staferova and Alexander Belokon (eds.), *Atlas of Multilingualism in Dagestan Online*. Linguistic Convergence Laboratory, HSE. https://multidagestan.com.

— (2019): Nina D., Aleksandra Kozhukhar and George Moroz, "Gendered multilingualism in highland Daghestan: story of a loss". *Journal of Multilingual and Multicultural Development* 40, 115–132. https://www.researchgate.net/publication/326303743; http://dx.doi.org/10.1080/01434632.2018.1493113.

Filatov & Daniek (2022): Konstantin F. & Michael D. (eds.), *DagSwadesh: 100 Swadesh lists from Daghestan. An online database of basic vocabulary divergence across neighboring villages*. Moscow: Linguistic Convergence Laboratory, HSE University. http://lingconlab.ru/dagswadesh/.

Genko (2005): Анатолий Н. Г., *Табасаранско-русский словарь.* Москва: Академия. https://disk.yandex.ru/i/EO93AE49hL7MFA.

Gorcy (1992): *Кавказские горцы. Сборник сведений*. 1992. Москва: Адир. http://web2.anl.az:81/read/page.php?bibid=vtls000190090. [Reprint of *Сборникъ свѣденій о кавказскихъ горцахъ* I, Тифлисъ 1868; https://rusneb.ru/catalog/001199_000087_220/.]

Kibrik & Kodzasov (1988): А. Е. Кибрик & С. В. Кодзасов, *Сопоставительное изучение дагестанских языков: глагол*. Москва: Издательство Московского Университета. https://libarch.nmu.org.ua/bitstream/handle/GenofondUA/31341/0ebd1d3bb7e2a17103af4a0bde436732.pdf?sequence=1&isAllowed=y; http://ir.nmu.org.ua/handle/GenofondUA/31341.

— (1990): А. Е. Кибрик & С. В. Кодзасов, *Сопоставительное изучение дагестанских языков: имя, фонетика.* Москва: Издательство Московского Университета. https://libarch.nmu.org.ua/bitstream/handle/GenofondUA/56547/875e87bb12cca7efdb62d1c85884b9ba.pdf; http://ir.nmu.org.ua/handle/GenofondUA/56547.

Koile et al. (2022): Ezequiel K., Ilya Chechuro, George Moroz and Michael Daniel, "Geography and language divergence: The case of Andic languages". *PlosOne* 17/5. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0265460; https://doi.org/10.1371/journal.pone.0265460.

Koryakov (2002): Yuri B. K., *Atlas of Caucasian languages with Language Guide*. Moscow: Institute of Linguistics, Russian Academy of Sciences. http://lingvarium.org/raznoe/publications/caucas/ACL-all.pdf.

Koryakov et al. (2019): Yury B. Koryakov, Daria A. Staferova, Alexander A. Belokon, Nina R. Dobrushina *Population of Dagestan from data of different census years*. Linguistic Convergence Laboratory, HSE University. https://multidagestan.com/census.

Lavrov, Leonid I. (1953): Леонид И. Л., "Некоторые итоги Дагестанской экспедиции 1950–1952 гг." *Краткие сообщения Института этнографии* 19, 3–7.

— (1978): Леонид И. Л., "О причинах многоязычия в Дагестане". In: Леонид И. Л., *Историкоэтнографические очерки Кавказа*. Ленинград: Наука, 29–32. http://apsnyteka.org/file/Lavrov_Istoriko-etnograficheskie_ocherki_Kavkaza.pdf.

Lüpke (2016): Friederike L., "Uncovering Small-Scale Multilingualism". *Critical Multilingualism Studies* 4/2, 35–74. https://cms.arizona.edu/index.php/multilingual/article/view/100

Magomedkhanov (2008): Магомедхан Магомедович М., Дагестанцы. Этноязыковые и социокультурные аспекты самосознания. Moscow: ООО «ДИНЭМ». https://instituteofhistory.ru/media/library/publication/files/Магомедханов_2008._Дагестанцы._Этноязыковые....pdf.

Materialy (1927): *Материалы Всесоюзной переписи населения 1926 г. по Дагестанской АССР*. Вып. 1: *Список населенных мест Дагестанской АССР*. Махач-Кала: Дагестанское Статистическое Управление. http://www.lingvarium.org/docs/dagestan-1926.zip.

Nichols (in preparation): Johanna N., *The Languages of the Great Caucasian Range*.

Pakendorf et al. (2021): Brigitte P., Nina Dobrushina and Olesya Khanina, "A typology of small-scale multilingualism". *International Journal of Bilingualism* 25/4, 835–859. https://journals.sagepub.com/doi/full/10.1177/13670069211023137; https://doi.org/10.1177/13670069211023137.

Sbornik materialov (1881–1926): *Сборникъ матеріаловъ для описанія мѣстностей и племен Кавказа*. 1–45. Тифлисъ / Махач-Кала. https://books.google.de/books?q=editions:HARVARD32044099656910&id=GToEAAAAYAAJ; http://kubangenealogy. ucoz.ru/index/sb_kavkaz/0-25.

Sbornik svedeniy (1868–1881): Сборникъ свѣдѣній о кавказскихъ горцахъ. 1–10. Тифлисъ. http://elib.shpl.ru/ru/nodes/17485-sbornik-svedeniy-o-kavkaze-t-1-7-9-tiflis-1871-1885.

Svod (1893): *Сводъ статистическихъ данныхъ о населеніи Закавказскаго края, извлеченныхъ изъ посемейныхъ списковъ 1886 г.*, изданъ по распоряженію Главноначальствующаго гражданскаго частію

на Кавказѣ, Закавказскимъ Статистическимъ Комитетомъ. Тифлисъ: Типографія И. Мартиросянца. https://rusneb.ru/catalog/000199_000009_005403186/.

Volkova (1967): Наталья Г. В., “Вопросы двуязычия на Северном Кавказа”, *Советская этнография* 1, 27–40. https://www.booksite.ru/etnogr/1967/1967_1.pdf.

Wixman (1980): Ronald W., *Language Aspects of Ethnic Patterns and Processes in the North Caucasus*. Chicago: University of Chicago. https://abkhazworld.com/aw/Pdf/Language_Aspects_of_Ethnic_Patterns_and_Processes_in_the_North_Caucasus_by_Ronald_Wixman.pdf.

ნახურ-დაღესტნური ენები: ციფრული რესურსები და მათი გამოყენების რამდენიმე მაგალითი

ნინა დობრუშინა (ლიონი)

ნახურ-დაღესტნურ ენათა ოჯახი კავკასიაში ყველაზე დიდი ოჯახია, თუ გავითვალისწინებთ მასში შემავალი ენებისა და ენობრივი ვარიანტების სიმრავლეს. ბოლო ათწლეულში არაერთი პროექტი განხორციელდა საერთაშორისო თანამშრომლობის შედეგად, რომლებიც მიზნად ისახავდა ამ ენათა სისტემურ შესწავლასა და დოკუმენტირებას, რის შედეგადაც შეიქმნა ახალი რესურსები დაღესტნის მოსახლეობის, მულტილინგვიზმის, ლექსიკისა და გრამატიკის შესახებ. მოცემულ ნაშრომში განხილულია ის რესურსები, რომლებიც ამჟამად განვითარების სხვადასხვა ეტაპზეა.

ნაშრომის დასაწყისში წამოდგენილია მოკლე ინფორმაცია მოცემული რეგიონისა და მისი ენების შესახებ (ნაწილი 1), განხილულია დაღესტნის სოფლების დემოგრაფიულ მონაცემთა ბაზა (ნაწილი 2), დაღესტანში არსებული მრავალენოვნების ამსახველი მონაცემთა ბაზა (ნაწილი 3), დაღესტნურ ენებში ნასესხობების მონაცემთა ბაზა (ნაწილი 4), არაბული, სპარსული და თურქული ენებიდან დაღესტნურ ენებში შემოსული ლექსიკის მონაცემთა ბაზა (ნაწილი 5), სვადეშის მონაცემთა ბაზა – 100 სიტყვის სია დაღესტნური ენებიდან (ნაწილი 6), დაღესტნურ ენათა ტიპოლოგიური ატლასი (ნაწილი 7) და რუთულურის დიალექტოლოგიური ატლასი (ნაწილი 8). გარდა ამისა, სტატიაში წარმოდგენილია ნახურ-დაღესტნური ენების ელექტრონული კორპუსებისა და ლექსიკონების სია.