On phonology of East Caucasian languages

G. Moroz

## Introduction

There are a lot of work about phonology of indigenous languages of the Caucasus (Catford 1977; Job and Smeets 1994; Smeets 1994; Alekseev et al. 2001; Hewitt 2004; Grawunder 2017; Beguš 2021; Boris 2021a, 2021b; Koryakov and Maisak Fort) and particular dedicated East Caucasian languages (A. E. Kibrik and Kodzasov 1990). There are also a lot of work on historical-comparative phonetics (Бокарев 1960, 1981; Гудава 1964; Имнайшвили 1977; Акиев 1977; Гигинейшвили 1977; Б. Б. Талибов 1980; Nikolayev and Starostin 1994; Nichols 1994; Ardoteli 2009; Мудрак 2019, 2020) and many others. Fortunately the amount of grammatical description of particular languages is increasing, so we have a lot of more new and detailed information about phonological inventories of particular villages and do not need to extrapolate our knowledge of standard languages onto all villages, where the language is spoken. Even though we have a lot of works on different East Caucasian languages, in order to proceed with the analysis we still need a unified description of those inventories. We compiled the database of East Caucasian languages that can be downloaded from [here](https://raw.githubusercontent.com/agricolamz/new_caucasian_phonology_dataset/master/database.csv). The database has the following variables:

* id – unique id for each segment
* glottocode – language identifier after Glottolog database (Hammarström, Forkel, and Haspelmath 2017)
* language – language name
* idiom
* type
* value — segment in IPA notation
* value\_source — segment as it was presented in the source
* annotation — annotation of type of the segment (consonant vs. vowel)
* source — bibtex entry of the source of the information
* page
* comment

The database is compiled from different sources and therefore some languages can be present multiple times: in some cases it is different idioms (e. g. Darwa dialects), in other cases it is different sources of the same languages (e. g. Khinalug).

On average East Caucasian languages (and other indigenous languages of the Caucasus) have more consonants and vowels then other languages of the world. The main reasons for this are the following

* as other indigenous languages of the Caucasus East Caucasian languages have **ejective consonants** (except Udi);
* as other indigenous languages of the Caucasus East Caucasian languages have **uvular consonants**;
* Andic and Tsezic languages have **lateral obstruents**;
* **labialization**, gemination and fortis/lenis destinction is widespread across Dagestan;
* common among languages of the world triangle vowel system (*i*, *e*, *a*, *o*, *u*) is complicated with **nasalization** (mostly Andic and some Tsezic languages), **long vowels** (Nakh, Andic and Tsezic), **pharyngealisation** and **umlaut vowels** (on the South, Azerbaijanian influence?).

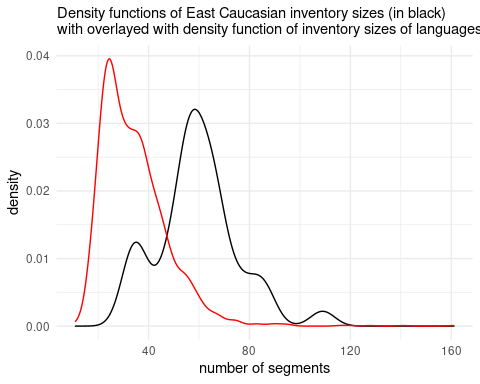
In this introductory part of the atlas we will be focusing on inventory sizes, but there are also the following separate subchapters:

* Number of contrasts in voiceless consonants
* Presence of fortis/geminates
* Presence of labialized consonants
* Presence of laterals
* Presence of long vowels
* Presence of nasal vowels
* Pharyngeal inventories
* Presence of pharyngealized segments
* Inventory of v-like consonants
* Presence of palatalized consonants
* Presence of umlaut vowels
* Presence of uvular voiced stop
* Presence of velar voiced fricative
* Presence of ejective pʼ
* Presence of ejective fricatives

## Inventory size

Here is the map of inventory sizes of East Caucasian languages:

As we can see inventory size of languages in our dataset are ranged from 33 (Georgian and Kumyk) to 109 (Northern Akvakh). We can compare the obtained numbers with the PHOIBLE database (Moran, McCloy, and Wright 2014) which contains inventories of the languages of the world[[1]](#footnote-22):

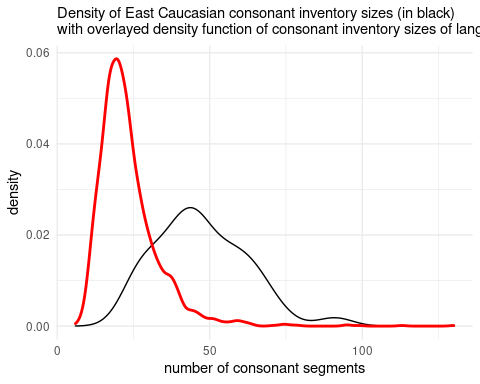


As we can see, East Caucasian languages in general have big segment inventories (with mean, median and mode near 60 segments) comparing to the rest languages of the world. A small peak around 40 can be explained by non-indigenous languages present in our dataset. From the next sections we will see that overall large inventories are mostly

### Consonant inventory size

We can do the same comparison for the consonant inventories:

As we see inventory size are differe from 25 (Budukh) to 91 (Northern Akvakh). We can compare the obtained numbers with the PHOIBLE database (Moran, McCloy, and Wright 2014):

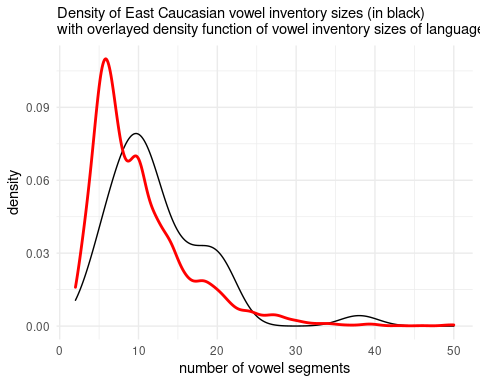


As we can see, the majority of languages from PHOIBLE has less consonants then East Caucasian languages.

### Vowel inventory size

We can do the same comparison for the consonant inventories:

As we see inventory size are differ from 5 (Avar) to 38 (Bezhta). We can compare the obtained numbers with the PHOIBLE database (Moran, McCloy, and Wright 2014):



As we can see, even vowel inventory sizes of East Caucasian languages are bigger than the average.

## Series of stops

It looks like ejectives appear in non-indigenous languages of the Caucasus like Azerbaijani (Асланов 1974; Daniel 2021), Kumyk (Дмитриев 1940: 32), Ossetic (Abaev 1964: 6).

## Pharyngeals

(Colarusso 2013)

## OLD

## Labialization

Labialization is widely attested in East Caucasian languages, however there are some languages that lost this feature (e. g. in Budukh, (Б. Б. Талибов 2007)). Unfortunately, sometimes scholars do not put the whole list of labialised consonants attested in a language. After (Catford 1972) it is worth mentioning that there are several types of labialisation:

* /w/-labialisation — this kind of labialization is more or less independent of the main place of articulation. This kind of labialisation is typical for velar and uvular places fricatives and stops;
* /ɥ/-labialisation — this kind of labialization is typical for pharyngeals. It looks like the tongue body is retracted (due to pharyngeal place of articulation) and raised, that results /ɥ/-like sound;
* /y/-labialisation — this kind of labialization is typical only for post-alveolars in Abkhaz and Abaza (North Caucasian), but also may be a case for Tabasaran (Кибрик et al. 1982: 7–8; A. E. Kibrik and Kodzasov 1990: 337) and Aghul (A. E. Kibrik and Kodzasov 1990: 338).

## Geminates/Fortis

Geminated of fortis/lenis destinction attested through all branches of East Caucasian languages except Tsesic. Unfortunately some scholars understand this differently, so sometimes under this label aspirated vs unaspirated disctinction is used.

## Lateral obstruents

Laterals is the clear feature of Andic and Tsezic languages, hover there is one Lezghian language, Archi, where they also were attested. In most cases lateral system is present by fricative ɬ and two affricates tɬ and tɬ’. Sometimes the subsystem even bigger when geminated version are present.

## Labial approximmant vs fricatives

The most common labial sound in East Caucasian languages is bilabial **w** that is present in most languages, however in the South Dagestan approximant become a frecative **v**. I also visited the Dargwa village Ayalakab, where I witnessed the bilabial fricative realisation **β**.

## Pharyngealization

Pharyngealization — is a complex feature that phonetically reallised as a constriction in the pharynx or epiglottis zone. Different scholars used different strategy in describing locus of the pharyngealisation: sometimes it is analysed as a vowel feature (Forker 2013), sometimes as a consonant feature (Khalilova 2009) and sometimes as a suprasegmental feature (Aleksandr E. Kibrik 1994; Moroz 2019). It is common that pharyngealisation spreads through the word, hovewer most researchers distinguish core pharyngealised syllable (Aleksandr E. Kibrik 1994; Б. Б. Талибов 2007; Moroz 2019; Беляев, n.d.). Everything become even more complicated since some East Caucasian languages developed two different types of pharyngealisation (e. g. Беляев, n.d.). As we can see from the map, pharyngealization is common for central and south Daghestan for languages of Dargic, Lak, Tsezic and Lezgian branches.

## Nasalized vowels

Nasalised vowels is a comon feature of Andic and Tsezic languages, hovewer some researchers report sporadic final **n**-delition that leads to development of nasalised vowels in other branches (e. g. Bezhta, (Б. Б. Талибов 2007)).

## Long vowels

Long vowels are grouped in the North Dagestan, Chechnya and Ingushetia. It is not present in Avar and and South Dagestan. Hovewer there are some cases when they were reported as a result of some intervocal consonant delitions.

## Umlaut vowels

Umlaut vowels are present in the South Dagestan and in Nakh languages. I can speculatively propose that they historically developed from pharyngealisation, but this hypothesis needs to be accurately checked with shared lexicon.

## More features?

* syllable structure?
* vowel harmony?
* acoustic study of stress?

Abaev, V. I. 1964. *A Grammatical Sketch of Ossetic*. Bloomington: Indiana Univesrity Press.

Alekseev, M. E., S. A. Starostin, G. A. Klimov, and Ya. G. Testelets. 2001. *Yazyki Mira. Kavkazskiye Yaziki*. Academia.

Ardoteli, Nodar. 2009. *Xunjuri-Andiuri-Didouri Enat’a Istoriul-šedarebit’i p’onetika*. Gamomc’emloba Universali.

Beguš, Gašper. 2021. “Segmental Phonetics and Phonology.” In *The Oxford Handbook of Languages of the Caucasus*, edited by Maria Polinsky, 1–40. Language Science Press.

Boris, Lena. 2021a. “Tone and Intonation in Languages of the Caucasus.” In *The Oxford Handbook of Languages of the Caucasus*, edited by Maria Polinsky, 1–30. Oxford University Press.

———. 2021b. “Word Stress in the Languages of the Caucasus.” In *The Oxford Handbook of Languages of the Caucasus*, edited by Maria Polinsky, 1–30. Language Science Press.

Catford, J. C. 1972. “Labialisation in Causasian Languages, with Special Reference to Abkhaz.” In *Proceedings of the Seventh International Congress of Phonetic Sciences/Actes Du Septième Congrès International Des Sciences Phonétiques*, 679–82. De Gruyter Mouton.

———. 1977. “Mountain of Tongues: The Languages of the Caucasus.” *Annual Review of Anthropology* 6 (1): 283–314.

Colarusso, John. 2013. “The Typology of Gutturals.” In *Base Articulatoire Arrière: Backing and Backness. Munich: LINCOM Europa*, edited by Jean Léo Léonard and Samia Naïm, 93–109.

Daniel, Michael. 2021. “Ejectives in the Azerbaijani of Ilisu.”

Forker, Diana. 2013. *A Grammar of Hinuq*. De Gruyter Mouton.

Grawunder, Sven. 2017. “The Caucasus.” In *The Oxford Handbook of Languages of the Caucasus*, edited by R. Hickey, 356–95. Cambridge University Press.

Hammarström, Harald, Robert Forkel, and Martin Haspelmath. 2017. “Glottolog 4.4.” *Max Planck Institute for the Science of Human History*.

Hewitt, George. 2004. *Introduction to the Study of the Languages of the Caucasus*. Lincom.

Job, D. M., and R. Smeets, eds. 1994. *The Indigenous Languages of the Caucasus*. Vol. 3. Caravan Books.

Khalilova, Zaira. 2009. “A Grammar of Khwarshi.” PhD thesis, Leiden University.

Kibrik, A. E., and S. V. Kodzasov. 1990. *Sopostavitelnoye Izucheniye Dagestanskix Yazykov: Imya. Fonetika*. Vol. 2. Moskovskij Gosudarstvennyj Universitet.

Kibrik, Aleksandr E. 1994. “Archi.” In *The Indigenous Languages of the Caucasus. Vol. 4: The Northeast Caucasian Languages. Part 2*, edited by Rieks Smeets, 297–365. Delmar, NY: Caravan Press.

Koryakov, Yu., and T. Maisak, eds. Fort. *Handbook of Caucasian Languages*. Vol. 2. Berlin/New York: De Gruyter Mouton.

Moran, Steven, Daniel McCloy, and Richard Wright, eds. 2014. *PHOIBLE Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://phoible.org/>.

Moroz, George. 2019. “Phonology of Mehweb.” In *The Mehweb Language: Essays on Phonology, Morphology and Syntax*, edited by Michael Daniel, Nina Dobrushina, and Dmitry Ganenkov, 17–39. Language Science Press.

Nichols, Johanna. 1994. “The Nakh-Daghestanian Consonant Correspondences.” In *Current Trends in Caucasian, East European, and Inner Asian Linguistics : Papers in Honor of Howard i. Aronson*, edited by Dee Ann Holisky and Kevin Tuite, 207–65. John Benjamins Publication Company.

Nikolayev, S. L., and S. A. Starostin. 1994. *A North Caucasian Etymological Dictionary*. Asterisk Press.

Smeets, R., ed. 1994. *The Indigenous Languages of the Caucasus*. Vol. 4. Caravan Books.

Акиев, А. Ш. 1977. “Историко-сравнительная фонетика даргинского и лакского языков: Система консонантизма.” *Дагучпедгиз*.

Асланов, А. М. 1974. “Иберийско-кавказские абруптивы в закатальско-кахских говорах азербайджанского языка.” *Ежегодник иберийско-кавказского языкознания* 1: 224–28.

Беляев, О. И. n.d. “Два типа фарингализации в ширинском даргинском: Предварительные наблюдения.” In *ДУРХЪАСИ ХАЗНА*, edited by Т. А. Майсак, Н. Р. Сумбатова, and Я. Г. Тестелец, 18–29.

Бокарев, Евгений Алексеевич. 1960. *Основные вопросы исторической фонетики дагестанских языков*. Издательство восточной литературы.

———. 1981. *Сравнительно-историческая фонетика восточнокавказских языков*. Наука.

Гигинейшвили, Б. К. 1977. “Сравнительная фонетика дагестанских языков.” *Тбилиси: Издательство Тбилисского университета*.

Гудава, Того Евстафьевич. 1964. *Консонантизм андийских языков: Историко-сравнительный анализ*. Изд-во Академии наук Грузинской ССР.

Дмитриев, Н. К. 1940. *Грамматика кумыкского языка*. АН СССР.

Имнайшвили, Д. С. 1977. *Историко-сравнительный анализ фонетики нахских языков*. Мецниереба.

Кибрик, Бергельсон, А. Е., КК Богатырев, ОЮ Богуславская, Анна А Зализняк, СВ and Кодзасов, ИА Муравьева, ВА Плунгян, and ЕВ Рахилина. 1982. “Табасаранские этюды.”

Мудрак, Олег Алексеевич. 2019. “Аварские основы.”

———. 2020. “Аварские основы.”

Талибов, Б. Б. 1980. *Сравнительная фонетика лезгинских языков*. Наука.

Талибов, Б Б. 2007. *Будухский язык*. Academia.

1. Here and further performing such a comparison of our dataset and one, provided by PHOIBLE, arise a question: whether we need to exclude East Caucasian and non-indigenous languages of the Caucasus from the PHOIBLE subsample that we use? In this text we decided to exclude them for the sake of the comparability. However, groups that we get are far from to be comparable in the proper way: different language families are overrepresented in both groups; the dataset sizes are different (PHOIBLE’s 2169 languages vs our dataset of 50 idioms) etc. For those who are against of the exclusion we assure that this operation changes shapes of the density plots that we obtained, but the change is extremely small. [↑](#footnote-ref-22)