

Phonological adaptation of Russian borrowings in Avar-Andic languages

Abstract

Aims and Research Questions: The paper investigates processes of phonological adaptation of Russian borrowings in Avar-Andic languages with an aim to determine whether degrees of phonological adaptation correlate with geographic distances and phylogenetic classifications between the languages, as well as with other factors including population size, time depth and frequency of use.

Methodology, Data and Analysis: We analyzed Russian borrowings from ten dictionaries of Avar-Andic languages. Each borrowing was annotated and matched segment-by-segment to its corresponding Russian source lexeme. We performed a qualitative analysis of the correspondences and modeled them with a mixed effect logistic regression.

Findings: Modeling the probability of loanword adaptation gives a hierarchy of languages that is partially explained by a language's history of direct contact with Russian and authorship of the dictionaries, but does not fully match with geographic distances and phylogenetic classifications, nor with population sizes. Time depth and frequency of use show the expected effect on loanword adaptations, but their predictive strength is not statistically significant.

Originality: This is the first quantitative investigation of phonological adaptations of Russian borrowings in Avar-Andic languages.

Implications: Our findings have implications for studies on language contact, and more specifically contribute to research on loanword adaptation in multilingual settings.

Keywords

Language contact, borrowing, phonological adaptation, Avar-Andic languages, Russian

Introduction

Lexical borrowing is the most common outcome of language contact. It occurs even in situations of casual contact, sometimes even when borrowers are not fluent in the source language, or when there are only few bilingual speakers in the borrowing language community (Thomason, 2001, p. 70). At such earlier stages of contact, loanwords usually undergo strong phonological adaptation, and can be “nativized” to the point that the original source word is hardly recognizable. However, when contact becomes more intense, the degree of phonological adaptation tends to decrease, and the borrowing language can ultimately adopt new sounds and sound sequences (Thomason, 2001, p. 73).

In studies on loanword adaptation, a number of factors are usually mentioned as affecting the degree of phonological integration of borrowings in the recipient language. One of these factors is the speakers' level of bilingualism: the degree of phonological adaptation is lower when speakers are more fluent in the donor language, cf., e.g., (Haugen, 1950, 1969). Another factor is time depth: loanwords tend to change their shape as time goes by, i.e., early loanwords show a higher degree of phonological adaptation as compared to loanwords introduced at a later (and more intense) stage of contact (Poplack & Sankoff, 1984, p. 101;

Poplack et al., 1988, p. 70). Partly related to time depth is frequency of use: more frequent loanwords show a higher degree of phonological adaptation (Poplack & Sankoff, 1984; Poplack et al., 1988; Lev-Ari et al., 2014).

This study deals with phonological adaptation of Russian borrowings in languages of the Andic branch of East Caucasian (aka Nakh-Daghestanian). Dictionary data from eight Andic languages are compared to data from Avar, the closest relative of Andic within the family and a major lingua franca in northern Daghestan.

The paper is structured as follows. First, we briefly outline the sociolinguistic situation of the area where Andic languages are spoken and provide information about language populations and their repertoires. Second, we describe the data used for this study and illustrate the process of data annotation. Then we show the results of data analysis and statistical investigations. Finally, in the last section, we summarize and discuss our findings.

The Andic area: languages, speakers, and repertoires

The Andic branch of East Caucasian includes eight languages (Akhvakh, Andi, Bagvalal, Botlikh, Chamalal, Godoberi, Karata-Tukita, and Tindi) spoken in 77 villages in the northern Caucasus (Republic of Daghestan, Russian Federation). These villages belong to three¹ administrative districts of Daghestan: Botlikhsky, Tsumadinsky and Akhvakhsky districts, with Botlikh, Agvali and Karata as their administrative centers. Andic villages are surrounded by Avar villages (Figure 1).

¹ One Chamalal village is located outside Daghestan, in the Chechen Republic, and three Akhvakh villages are located in the Shamilsky District of Daghestan.

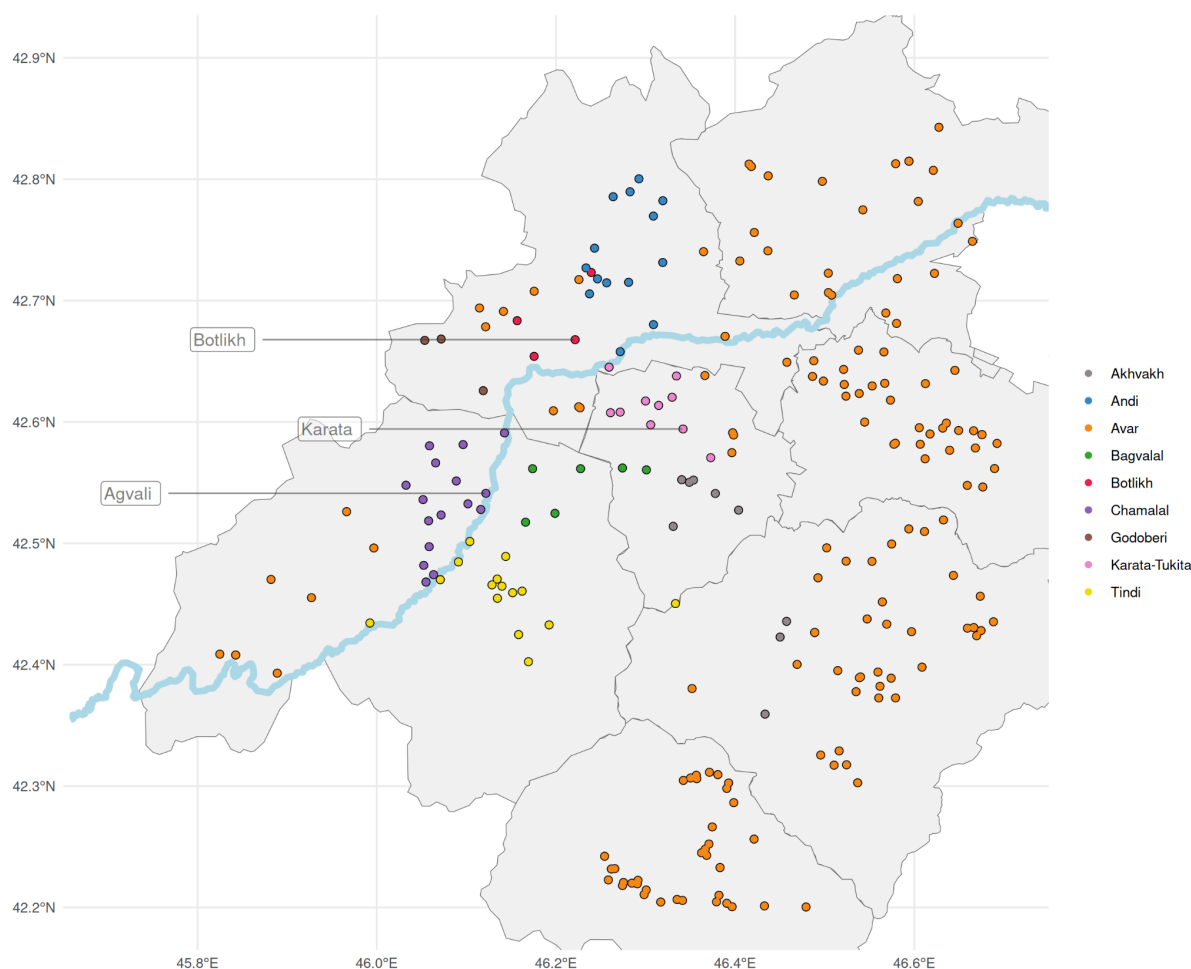


Figure 1. The Andic area (map created with the R package *lingtypology* (Moroz, 2017)).

This territory was controlled by the Avar Khanate (13th–19th centuries), and later became part of the Andiyskiy District (1861–1928), one of the administrative districts of the Daghestan Oblast, which was created in 1860 as a province of the Russian Empire. The administrative center of the Andiyskiy District was Botlikh, which is up to date the largest settlement of the region. Botlikh has been an important economic center for the whole Andic area since the Middle Ages, and people from neighboring villages used to go there to exchange goods and livestock at the local market (Alimova & Magomedov, 1993, p. 45). Furthermore, the main routes to other trade centers (in Georgia, Chechnya, and lowland Daghestan) all passed through Botlikh, which also made it a major spot for contacts among Andic peoples (Alimova & Magomedov, 1993, p. 48).

Language classification within the Andic branch is not uncontroversial. For some of the Andic languages, dialectal differentiation is rather considerable, and in some cases different village varieties are so divergent from one another that researchers have reconsidered their status as separate languages, e.g., Southern Akhvakh varieties are now classified separately from Northern Akhvakh in Glottolog (Hammarström et al., 2024).

Genealogical relationships between the languages of the branch are also disputed, and several phylogenetic hypotheses have been proposed in the literature. As discussed in (Koile et al., 2022), qualitative classifications and classifications based on lexical cognacy give rather different results. As the authors point out, the first show higher correlations with geographic distances, but they “have an implicit geographic bias or are based on selections of isoglosses that naturally tend to behave better in spatial terms than shared lexical retentions and innovations, on which quantitative phylogenies are usually based” (Koile et al., 2022, p. 4). Nonetheless, different classifications still share some correspondences. For instance, all sources agree on the similarity of Bagvalal and Tindi on the one hand, and Godoberi and Botlikh on the other. However, for the other languages of the branch, rather divergent groupings have been proposed (see Figure 2).

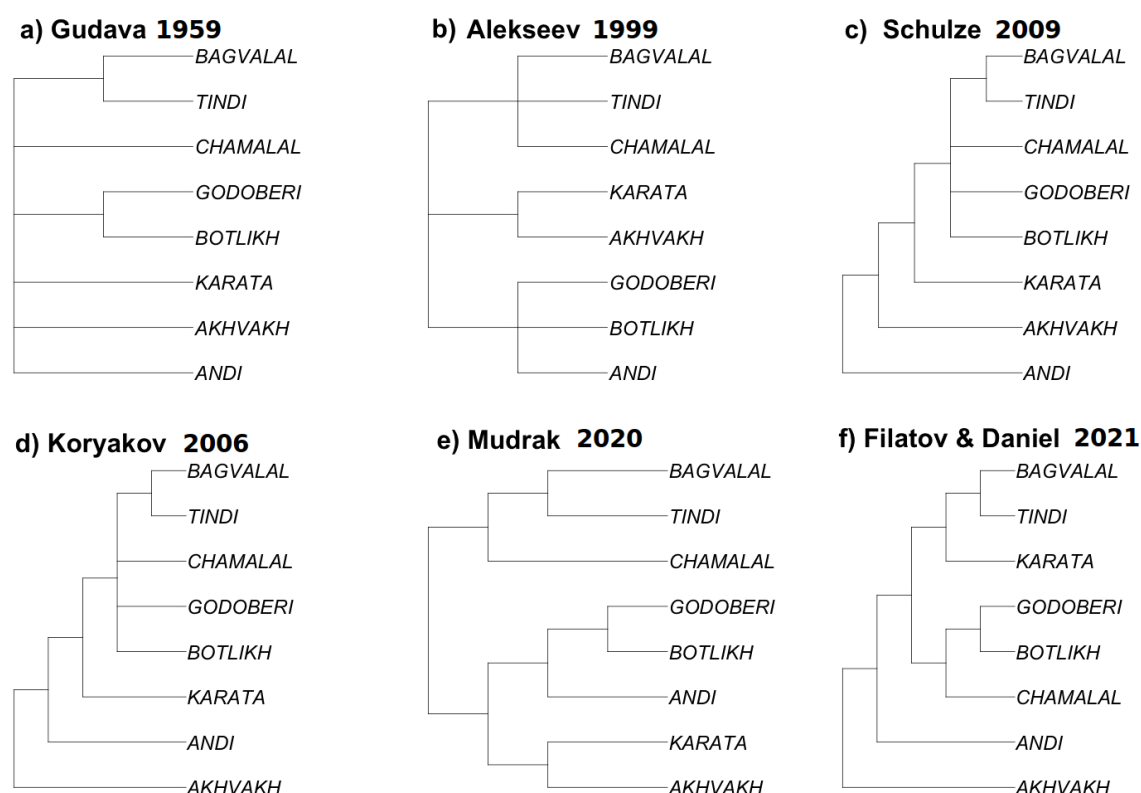


Figure 2. Qualitative approaches to the classification of Andic languages after (Koile et al., 2022, Figure 2). The plot is reproduced with the authors’ permission.

As for population sizes, it is quite problematic to give precise estimations because information from official censuses is not fully reliable. During the Soviet period, it was common practice to subsume Andic peoples under Avars (the largest ethnic group of the region), and many still self-identify as such, despite recognizing that their native language is very different from Avar. However, considering that most Daghestanian villages are ethnically homogeneous, we can use data on village sizes to get an idea of the approximate language population sizes. Note, however, that figures for the district centers are necessarily less reliable, since these villages are becoming increasingly multi-ethnic.

Table 1. Population sizes according to the censuses of 1926 and 2010. Data from the Atlas of multilingualism in Dagestan (Dobrushina et al., 2017).

Language	1926	2010
Avar	124,211	189,888
Andi (Upper)	6,484	11,075
Karata	4,763	8,873
Tindi	4,128	4,252
Chamalal	3,651	5,132
Bagvalal	3,052	4,057
Akhvakh (Northern)	2,216	5,977
Botlikh	1,694	12,501
Godoberi	1,197	3,114

Population size is an important aspect to consider in works on language contact, and specifically for Daghestan, this factor has been found to correlate with multilingualism of language communities: the smaller the language population, the higher its level of multilingualism (Dobrushina & Moroz, 2021). Since levels of multilingualism can be implied in processes of loanword adaptation, this seems an important aspect to consider.

Concerning the sociolinguistic situation, the overall picture is rather homogeneous throughout the Andic area. Most villages of the region are largely monoethnic. Mixed marriages are not common, and in the rare case when women get married outside their native village, they usually learn the local language and use it with their children (Dobrushina, 2023).

All Andic languages are officially unwritten and evaluated as “endangered” by sources like Ethnologue (Eberhard et al., 2024). Nonetheless, they are all still actively used and passed on to younger generations, at least in the original settlements in the highlands. At school, children are taught Avar, and the subject is called “mother tongue” (Rus. *rodnoj jazyk*). Before school age, children are often monolingual in their native language, although some knowledge of Russian is now quite common. Avar is the closest relative of the Andic languages on a family level and is one of the official languages of Daghestan. In the ethnic territory of Andic peoples, Avar has historically had the role of lingua franca for interethnic communication. In a region of such a high linguistic diversity, knowledge of a common language was necessary to communicate with people from the neighboring settlements. Since Avars have historically been the largest and most influential people in the area, their language also ended up having a dominant role, and almost all highlanders from this region have traditionally had active command of Avar (Volkova, 1967, pp. 29–30). A study conducted by

Dobrushina and Zakirova (2019) in the Karata area has shown that this is still the case for more than 90% of the population of that region, although it seems that knowledge of Avar among the younger generations is now becoming mostly passive (Magomedov & Magomedova, 2021).

The other important lingua franca of the area (as for all Daghestan) is Russian, which is now gaining dominance at the expense of Avar. The spread of Russian as a lingua franca in Daghestan began relatively recently, but it proceeded at an extremely high rate. At the end of the 19th century only one out of 1,000 people in Daghestan could speak Russian (Volkova, 1967, p.31), but by the end of the 20th century it had already become a second language (L2) for most of the population, and the first language (L1) for the younger generations who were born in towns (Dobrushina & Kultepina, 2021, p. 339).

Russian is the first language to have become a common lingua franca for all Daghestan. Until the first half of the 19th century, several languages were used as lingua francas in different areas: Avar in the north, Azerbaijani in the south, and Kumyk in the lowlands (Dobrushina & Kultepina, 2021, p. 341). In some regions, other patterns of multilingualism prevailed over the “lingua franca model”, and people often spoke the language of their neighbors. In such cases, asymmetric bilingualism was the most frequent situation, i.e., people from one of the villages could speak the language of their neighbors, but not the other way around. However, cases of symmetric bilingualism were also attested (Dobrushina, 2013; cf. also Chirikba, 2008, pp. 30–31).

The rise of Russian as a lingua franca is explained by Dobrushina and Kultepina (2021) as mainly due to the introduction of compulsory school education in the 1930s. By analyzing information on more than 3,500 people collected with the method of retrospective family interviews in 27 villages of Daghestan, the authors demonstrate that a steep rise in the command of Russian is observed in the generation born in the 1920s, which is correlated with the opening of Soviet schools in most Daghestanian villages in the 1930s. Starting from the generation born in the 1950s, the percentage of people with a (spoken and written) command of Russian is almost 100% (Dobrushina & Kultepina, 2021, pp. 342–343). The authors report that, after an initial resistance concerning particularly female education, the adoption of Russian (and the ideology connected with it) was largely welcomed by the population of Daghestan because this was associated with the idea of future and progress, and hence, social lift (Dobrushina & Kultepina, 2021, pp. 354–355). Importantly, the same development did not take place in other post-Soviet areas with similar educational systems like Azerbaijan, where Azerbaijani remained the dominant language used for interethnic communication in multilingual regions. As the authors suggest, “school education can add an L2 to the language repertoire in a large territory only under the conditions of a specific linguistic ecology” (Dobrushina & Kultepina, 2021, p. 355). Daghestan, with its extreme linguistic diversity and the absence of a language common to the whole region, was a fertile ground for the spread of Russian as lingua franca.

Given the role of Russian in Daghestan, it is not surprising to encounter Russian borrowings in the local languages. Lingua francas are known to play a major role as lexical donors in multilingual contexts, and this is often explained as due to the high level of bilingualism that arises as a consequence of a lingua franca being used as the main means of interethnic communication (Brown, 1996, 2011). However, as Daniel et al. (2021) point out,

high levels of bilingualism are not a sufficient explanation in the case of Daghestan: although knowledge of languages other than the lingua franca was high in some regions, this did not lead to substantial borrowing from these languages. Borrowing from a lingua franca is more intense because “a lingua franca might not be as strongly associated with ethnic identity as languages that are used only in communicating with L1 speakers of that language, and lexical borrowing from it does not threaten anyone’s identity” (Daniel et al., 2021, p. 520).

Although direct contact with Russian is a relatively recent phenomenon, Russian borrowings are pervasive in different semantic fields. Considering the Loanword Typology meaning list (Haspelmath & Tadmor, 2009), Comrie and Khalilov (2009, p. 421) and Chumakina (2009, p. 438) found, for Bezhta (Tsezic) and Archi (Lezgian) respectively, that in the semantic field “Modern world”, which is the one with the highest proportion of loanwords in these languages, more than 60% of the borrowings are from Russian. This is not surprising, since modernization in the 20th century came from Russia, so Russian has since then been the main source of knowledge of the modern world and its artifacts.

As far as we know, comparable quantitative investigations for Andic languages have not been conducted so far, but impressionistic observations on the types of Russian borrowings into Andic languages and their approximate dating are sometimes reported in the literature we have access to. In general, the observed trends are similar for all Andic languages: especially starting from the beginning of the 20th century, borrowings in the most various semantic fields largely come from Russian, and this trend is in constant growth, cf. (Alisultanova, 2009, pp. 138–149) on Andi, (Magomedova & Abdulaeva, 2007, p. 722; Creissels, Forth.) on Akhvakh, (Magomedova, 2004, p. 592) on Bagvalal, (Saidova & Abusov, 2012, p. 565; Alexeyev & Verhees, Forth.) on Botlikh, (Magomedova, 1999, p. 435) on Chamalal, (Sirazudinov, 2015) on Godoberi, (Magomedova & Xalidova, 2001, p. 471; Pasquereau, Forth.) on Karata, (Magomedova, 2003, pp. 615–616) on Tindi. Some of these sources also mention that earlier borrowings underwent strong phonological adaptation, whereas more recent borrowings are adapted with few or no modifications at all.

By analyzing processes of phonological adaptation of Russian borrowings in Avar-Andic languages, we will address the following research questions:

- a) Do degrees of similarity between Avar-Andic languages in terms of loanword adaptation correlate with geographic distances and phylogenetic classifications?
- b) Do we observe lower degrees of phonological adaptation in languages with a longer history of direct contact with Russian (i.e., Avar in the first place, but also Botlikh, which has long been an important administrative center of the area)?
- c) Is there a correlation between population sizes and degrees of phonological adaptations (i.e., the larger the language population, the lower the degree of phonological adaptation)?
- d) Do time depth and frequency of use correlate with degrees of phonological adaptations? Specifically, do earlier and more frequent loanwords show lower degrees of phonological adaptation?

Data and methods

Data selection for this study was based on borrowing annotations in the Avar-Andic database (Moroz et al., 2025). The database contains data from ten dictionaries of the following idioms:

- Andi, Gagatli dialect (Salimov, 2010);
- Avar (Gimbatov, 2006);
- Bagvalal (Magomedova, 2004);
- Botlikh (Saidova & Abusov, 2012; Alekseev & Azaev, 2019);
- Chamalal (Magomedova, 1999);
- Godoberi (Saidova, 2006);
- Karata-Tukita (Magomedova & Khalidova, 2001);
- Northern Akhvakh (Magomedova & Abdulaeva, 2007);
- Tindi (Magomedova, 2003).

For this study we performed a manual phonological and morphological annotation of the dataset. Each lexeme was provided with an IPA transcription and divided into segments, which was done for both the Russian source lexeme and the target language loanword. For the sake of comparability, in cases where epenthesis or elision was detected, we inserted 0-segments to ensure the identical length of the material. Metathesis was marked in a separate column so as not to be confused with a series of insertions and deletions. In case borrowings attached Andic morphology, we took into account only the stem.

The IPA transcriptions for the target languages were already present in the Andic database, while Russian transcriptions were added for this study. We should make some clarifications on several deviations from standard IPA correspondences for Russian in our dataset. Firstly, we annotated vowel reduction for unstressed syllables differently. Russian is usually described as a language with a three-term system where the largest number of allophones is distinguished in the stressed syllable. Moderate reduction is found in the syllable immediately preceding the stressed one or in onsetless syllables, while radical reduction is attested in all other cases (Iosad, 2012, p. 524). Russian orthography, on the other hand, is mostly based on morphological principles, and it usually reflects the variant attested in the strong position, which causes a mismatch between pronunciation and spelling (Kerek & Niemi, 2009). For our study we did not consider the distinction between moderate and radical reduction. Contrary to the IPA standards, where the stress sign is positioned before the stressed syllable, we positioned the stress sign directly before the stressed vowel. Stressed vowels in our system are written in lowercase, while any other vowels are capitalized. The choice of the vowel in all pre- and post-stressed syllables is determined by the prototypical allophone of the phoneme that appears in this syllable in strong positions. In (1) we provide an example of how the word *korova* ‘cow’ would be transcribed in the traditional system for Russian (1a) and in our dataset (1b).

- (1) *korova* ‘cow’
- a. [kɐˈrɔvə]
 - b. [kɒrˈɒvʌ]

Another deviation from classical descriptions of Russian regards assimilatory changes of the voice feature, as well as final devoicing. Obstruents in Russian undergo devoicing in final position, and they undergo voice assimilation when they are followed by another obstruent in consonant clusters (Hayes, 1984). In any such case in our annotation system, the segment undergoing (de)voicing is capitalized and left with its original voice feature. In (2) we provide an example of how the traditional way of transcribing (2a) differs from ours (2b).

- (2) *podklad* ‘lining (in sewing)’
 a. [pɐt̚'klat̚]
 b. [pADkl'aD]

All these changes were needed because we worked with dictionary data but had no access to audio recordings and pronunciation of borrowed lexemes. As Andic languages have no sustainable orthography, and given that the authors of the dictionaries knew Russian, in some cases it was impossible to determine whether the spelling reflects real pronunciation, or it simply follows the orthographic rules of Russian. For instance, the word *velosiped* ‘bicycle’ is reported as *w-i-l-a-s-i-p-'e-d* in the Bagvalal dictionary (Magomedova, 2004), which accounts for vowel reduction but not for final devoicing. In the Botlikh dictionary (Alekseev & Azaev, 2019) we find *w-e-l-a-s-a-p-'e-t*, which accounts for final devoicing but only partly for vowel reduction. The way of reflecting these processes is not uniform within the dictionaries. Therefore, we annotated Russian transcriptions in a way that can automatically switch between pronunciation and spelling so as not to confuse “dictionary effects” with real changes. By “changes” here and in the rest of the paper we refer to phonological adaptations that emerged during the adoption of a borrowed lexeme in the target language.

Changes were counted automatically, taking into account some decisions made about the transcriptions. For the sake of reproducibility, the dataset and the R script for all visualizations and analyses are available in the supplementary materials (see <https://osf.io/unsfd/>).

Firstly, we decided to avoid counting correspondences of Russian phonemes to Avar-Andic phonemes as changes if the mapping of the former to the latter was regular. For instance, we did not count the correspondence of the Russian voiced labiodental fricative [v] to the Avar-Andic labiovelar approximant [w] as a change. The reason for this is the absence of the phoneme /w/ in Russian and of the phoneme /v/ in all Avar-Andic. Speakers do not distinguish between these realizations in their L1, so the acoustically closest variant is chosen. The same decision was taken for the following pairs: Russian [tɕ] and Avar-Andic [tʃ], Russian [ɕ:] and Avar-Andic [ʃ:], Russian [z] and Avar-Andic [ʒ].

The second issue concerns nasal vowels, which are attested in all Andic languages but not in Avar. They as a single segment (e.g., *ã*) correspond to sequences of vowel + nasal (e.g., *a-n*) in Russian, which results in a mismatch in the length of IPA transcriptions between Russian and the target language. To overcome this mismatch, we introduced a 0 symbol in the target language transcription but did not count Russian nasal - 0 correspondence as a change if it was preceded by a nasal vowel, cf. (3) for an example of the automatic segment-by-segment mapping.

(3) Marking of nasals

A-v-A-n-ti-'u-r-A (Russian) → a-w-ã-0-t-'u-r-a (Bagvalal) ‘adventure’ — 1 change

Labialized consonants matched to sequences of the consonant + [v] were treated similarly.

Another difference between Russian and the majority of Avar-Andic languages is the presence of palatalization as part of the phonological system in the former and its absence in the latter, except for velars in Godoberi and Tindi. We decided to consider the difference in the palatalization feature as a distinct change only for velar consonants in Godoberi and Tindi, as they are expected to preserve palatalization from Russian where it is present.

Afterwards, we annotated the approximate time of borrowing based on the first occurrence of the lexeme in the Russian National Corpora (RNC) (Savchuk et al., 2024). All lexemes were divided into two categories: pre-Soviet lexemes and Soviet lexemes. The latter includes lexemes coined after 1917 or used no more than 10 times before 1917. Other lexemes were classified as pre-Soviet. Although the first occurrence in the RNC does not reflect the time of borrowing to Avar-Andic languages directly, this classification can help to distinguish between potentially early borrowings and more recent ones, providing us with an opportunity to assess the change in number and inventory of adaptations across time.

Each change in segments was labeled with a brief description. We decided to label only systematic adaptations observed for several classes of segments (Table 2). All other adaptations, such as change of vowel quality, were labeled as *other*.

Table 2. Types of changes annotated in our dataset.

Type of changes	Example
initial vowel epenthesis	<u>0</u> -s-t-'o-l (Russian) → <u>u</u> -s-t-'o-l (Andi) ‘table’
initial consonant epenthesis	<u>0</u> -U-f-'a-n-k-A (Russian) → <u>b</u> -u-f-'a-n-k-a (Bagvalal) ‘ushanka-hat’
middle vowel epenthesis	v- <u>0</u> -r-'a-tɛ (Russian) → w- <u>a</u> -r-'a-tʃ (Botlikh (Alekseev & Azaev, 2019)) ‘doctor’
middle consonant epenthesis	d-U- <u>0</u> -'e-lʲ (Russian) → d-u- <u>ʒ</u> -'e-l (Botlikh (Alekseev & Azaev, 2019)) ‘duel’
final vowel epenthesis	mʲ-'e-t-r- <u>0</u> (Russian) → m-'e-t-r- <u>a</u> (Akhvakh) ‘meter’
metathesis	m- <u>r</u> -' <u>a</u> -m-O-r (Russian) → m- <u>a</u> - <u>r</u> -m-'a-r (Godoberi) ‘marble’
middle vowel elision	j-'a- ɐ:- <u>ɪ</u> -k-0 (Russian) → j-a-f- <u>0</u> -k-a (Andi) ‘box’

middle consonant elision	k-O-n- <u>fi</u> -e-t-A (Russian) → k-e-m- <u>0</u> -e-t-i (Akhvakh) ‘candy’
final vowel elision	O-di-E-j-'a-l- <u>O</u> (Russian) → a-d-i-j-'a-l- <u>0</u> (Bagvalal) ‘blanket’
final consonant elision	si-'i-ti-E- <u>ts</u> (Russian) → tʃ:-'i-t-i- <u>0</u> (Akhvakh) ‘chintz’
ejectivization	n-A-l-'o- <u>G</u> (Russian) → n-a-l-'o- <u>k</u> ' (Bagvalal) ‘tax’
nasalization	<u>I-n</u> -k-U-b-'a-t-O-r (Russian) → <u>ĩ</u> -0-k-u-b-'a-t-o-r (Bagvalal) ‘incubator’
palatalization	b-'a-n- <u>k</u> -A (Russian) → b-'a-n- <u>ki</u> -a (Godoberi) ‘burk’
voicing	A- <u>p</u> -ti-'e-k-A (Russian) → a- <u>b</u> -t-'e-k-a (Godoberi) ‘apotheker’
devoicing	<u>b</u> -'o-tɛ-k-A (Russian) → <u>p</u> -u-tʃ-k-'a (Botlikh (Alekseev & Azaev, 2019)) ‘barrel’

As a result, we obtained a dataset consisting of 21,749 observations with the following columns:

- language: language;
- reference: source dictionary;
- dictionary_translation: dictionary translation unified across all dictionaries;
- lemma_frequency_ipm: frequency of the dictionary translation in the RNC;
- russian_ipa: modified IPA transcription of the Russian word or part of the word;
- target_ipa: IPA transcription of the target language word or part of the word;
- change: binary coding for the change for each segment correspondence;
- type_of_change: coding of the type of change (e.g., apocope, epenthesis, etc.);
- total: total number of units in the word;
- changes_per_word: number of segments marked as changes per word;
- time_of_borrowing: approximate time of borrowing based on data from the RNC.

The content of our database is visualized in Figure 3.

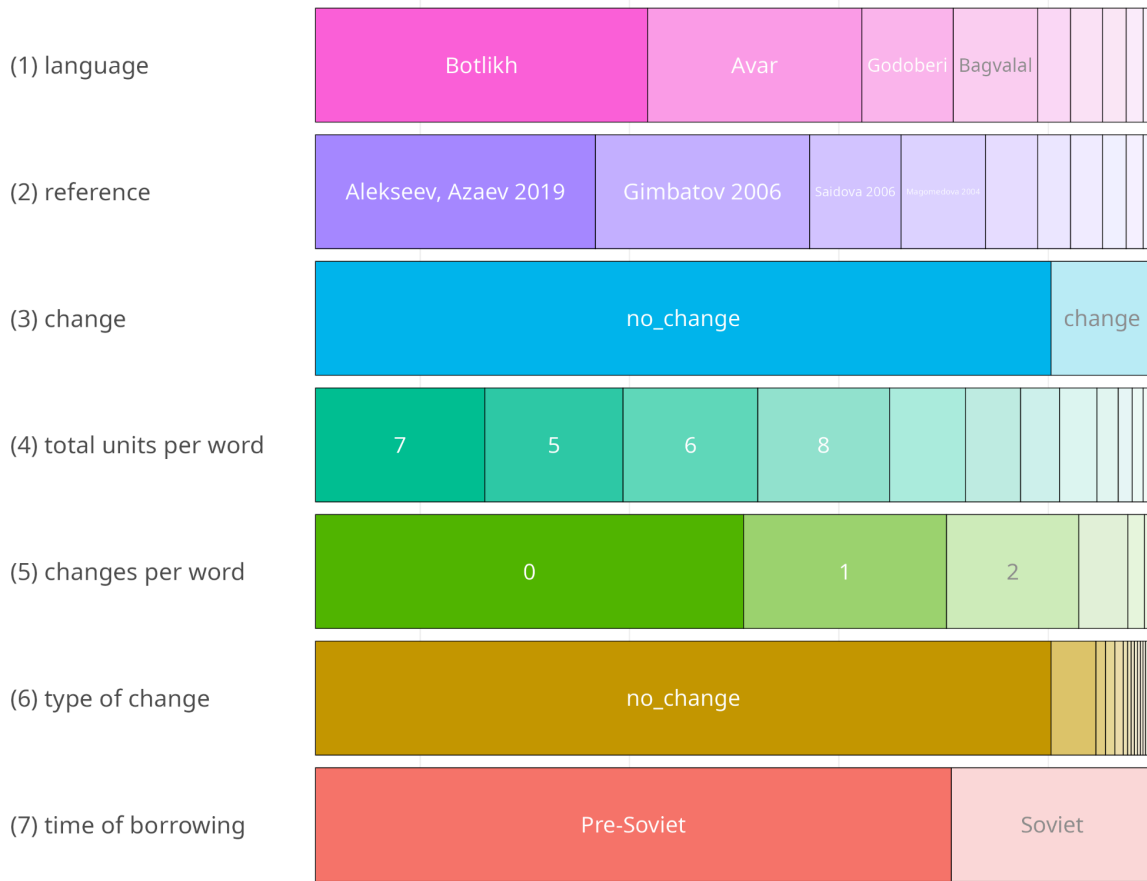


Figure 3. The database and parameters of data annotation (plot created with the R package `inspectdf` (Rushworth, 2022)).

Results

We created a model that predicts the average number of changes based on the dictionary, approximate time of borrowing and dictionary translation frequency. To do so, we applied a mixed effect logistic regression. The models were generated with the R (R Core Team, 2024) package `lme4` (Bates et al., 2015) with the following formula:

```
(4) change ~ dictionary*approximate time of borrowing +
dictionary translation frequency + (1|dictionary lemma
translation)
```

The random effect of the model is the unified dictionary lemma translation. We included the interaction of two variables: dictionary and approximate time of borrowing. Since this model compares values with some baseline, Soviet borrowings from the Avar dictionary (Gimbatov, 2006) were used as a baseline. Differences between all dictionaries turned out to be statistically significant. Approximate time of borrowing (p-value = 0.07570) and dictionary translation frequency (p-value = 0.05672) turned out not to be statistically significant. Just a few interactions of the dictionary variable with the time of borrowing variable turned out to

be statistically significant, i.e., for Andi and Tindi. Adding census data from Table 1 affected the model convergence, so we decided to exclude this variable.

Figure 4 shows the estimated probability of change by dictionary and approximate time of borrowing. While we could have expected clear-cut clusters of languages, what we see is rather a smooth continuum in which languages intersect with one another. Dictionaries are sorted according to the mean value of predicted probability and form a hierarchy of languages: Avar < Botlikh < Bagvalal < Andi < Godoberi < Karata-Tukita < Chamalal < Akhvakh < Tindi. As shown in Figure 4, borrowings from the pre-Soviet period on average present more changes than the Soviet ones. The only exceptions are observed in Tindi and Chamalal. These exceptions can be attributed to the smaller number of loanwords available for these languages in our dataset (47 lemmata for Chamalal and 74 lemmata for Tindi), among which Soviet loanwords are particularly rare (8 lemmata for Chamalal and 3 lemmata for Tindi), which makes it impossible for the model to make accurate predictions for these languages.

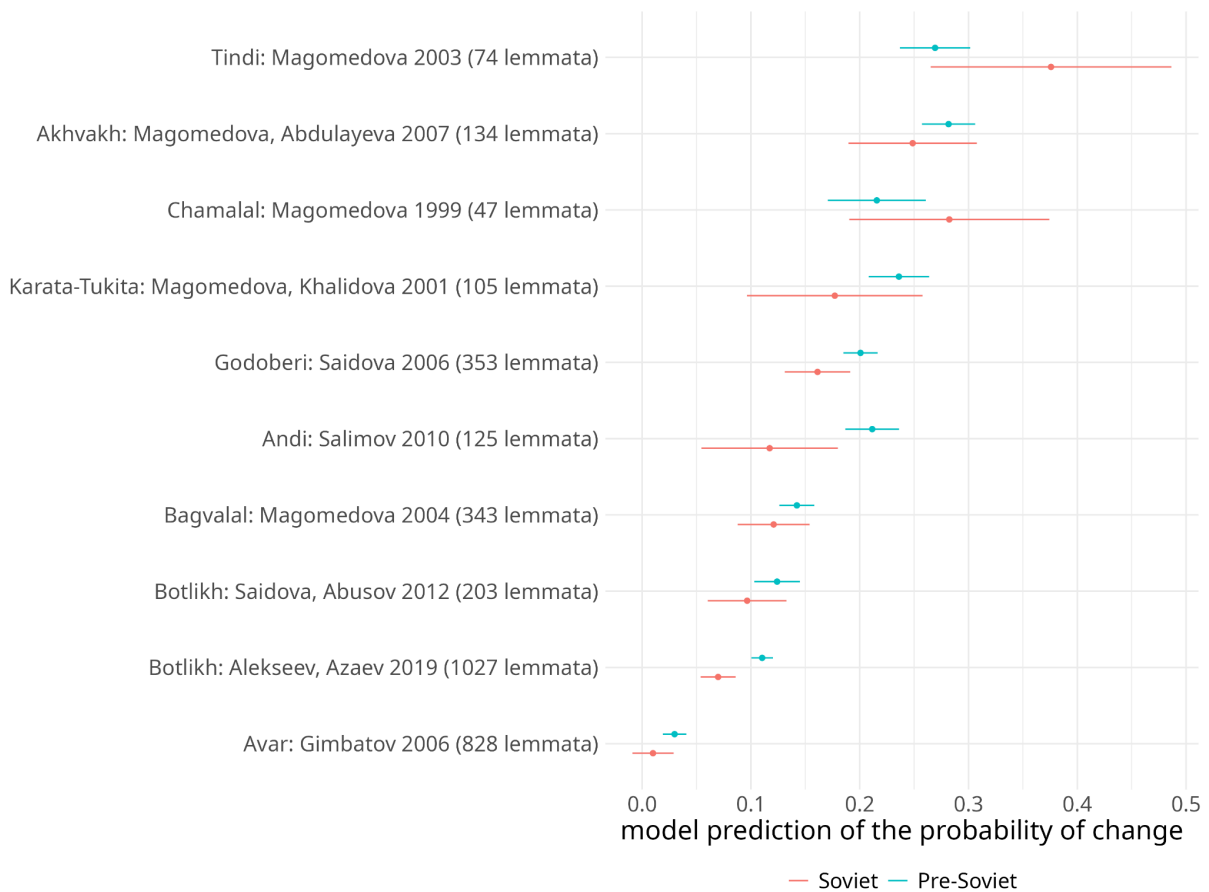


Figure 4. Probabilities of change by dictionary and approximate time of borrowing with 95% confidence intervals. Dictionaries are sorted according to the mean value of model prediction for both Soviet and pre-Soviet borrowings.

If we consider the phylogenies in Figure 2, we could expect Bagvalal and Tindi to cluster together. However, in the obtained hierarchy they are quite far from one another. The same is true for Godoberi and Botlikh, which ended up being separated by Bagvalal and Andi

in the hierarchy. Note, however, that the distance between them is smaller as compared to Bagvalal and Tindi.

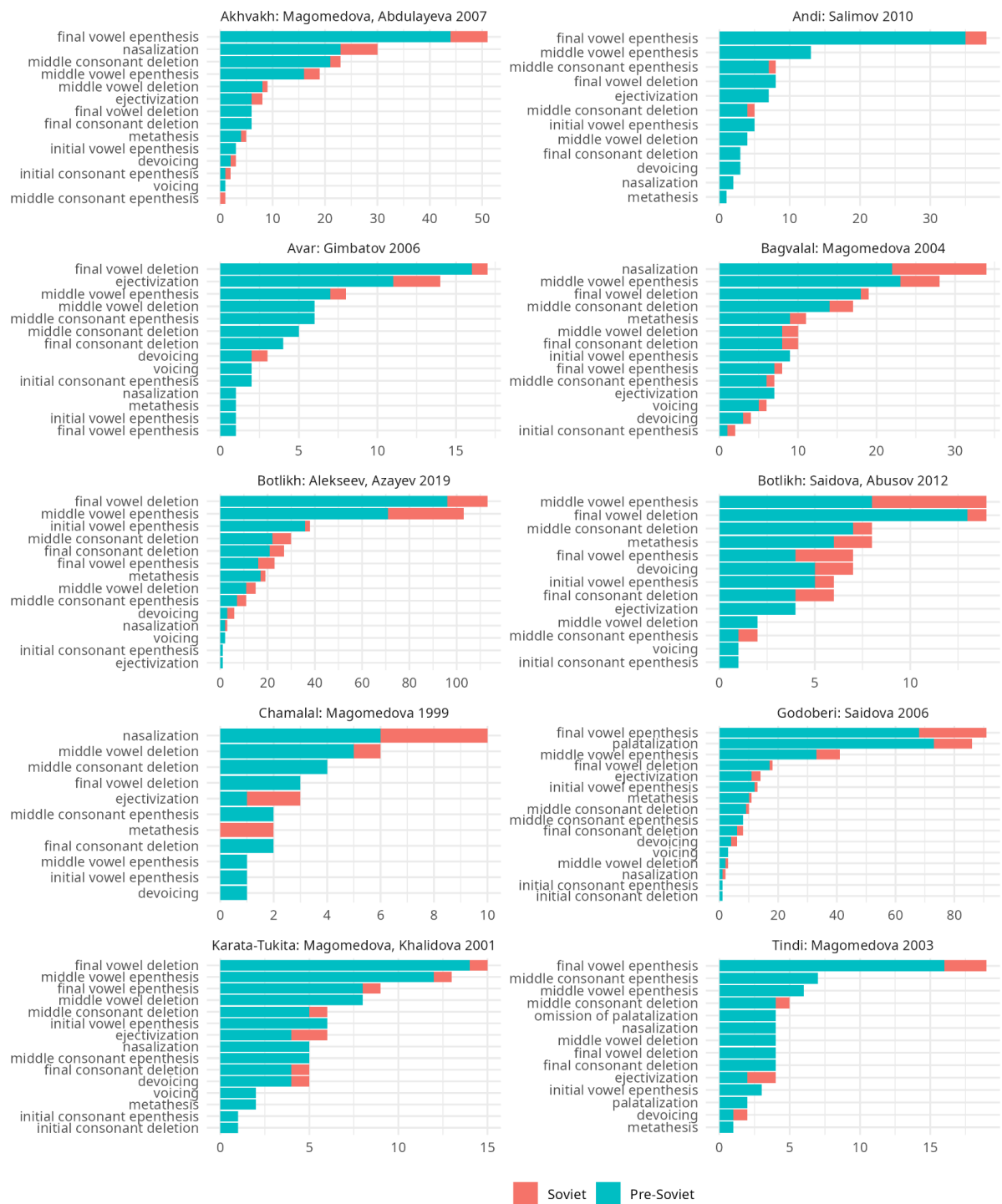


Figure 5. Distribution of changes across dictionaries. Changes marked as “other” are omitted. The distribution of all changes by dictionary and time of borrowing can be found in Table 1 in the supplementary materials.

Figure 5 shows the distribution of changes across time and the inventory of changes per language. The stacked bars display the overall number of instances for each type of change, while approximate time of borrowing is marked by color.

Concerning the hypothesis that more recent borrowings show lower degrees of phonological adaptation, we should note that most changes are preserved and apply to both pre-Soviet and Soviet borrowings. For instance, vowel prothesis is present in all Soviet borrowings (except for those in Avar) starting with [+sib][+plos -voice] to avoid initial clusters with such quality (5).

(5) Vowel prothesis before [+sib][+plos -voice] clusters in Botlikh (Alekseev & Azaev, 2019)

- a. Pre-Soviet
0-f-pi-'i-li-k-A → i-f-p-'i-l-k-a 'pin'
- b. Soviet
0-f-t-'a-m-p → i-f-t-'a-m-p 'stamp'

The difference between pre-Soviet and Soviet adaptations mostly concerns the reduced number of changes per word rather than a change in the adaptation inventories. It seems from our data that the process of erosion of borrowing adaptations does not cause some changes to disappear, but mostly makes existing mechanisms of adaptation apply less regularly. For instance, in pre-Soviet loanwords in Godoberi, final vowel epenthesis is highly regular after complex codas. As for the Soviet loanwords, this change applies non-systematically (6).

(6) Final vowel epenthesis after final consonant clusters in Godoberi

- a. Pre-Soviet loanword with change
A-r-ti-'i-s-t-0 → a-r-t-'i-s-t-i 'performer'
- b. Soviet loanword with change
z-'a-G-s-0 → z-'a-k-s-a 'registry organization'
- c. Soviet loanword without change
t-0-r-A-k-t-O-ri-'i-s-t → t-a-r-a-χ-t-a-r-'i-s-t 'tractor driver'

Figure 5 also illustrates the profile of each language by inventory of changes, which can serve as a basis for further elaborations on the hierarchy from Figure 4. Avar, which is taken as a baseline for comparison in the model for Figure 4, differs from all Andic languages in a number of features. First, it is the only language with non-native phonemes used in loanwords. We can observe Russian [f] retention in all borrowings into Avar, while in all Andic languages it is mapped to [p]. Moreover, other changes such as initial and final vowel epenthesis are rarely applied to borrowings in Avar. Even though it has the largest amount of data in our dataset, final vowel deletion as the most popular type of change counts only 17 instances, which is extremely few. This observation confirms our initial hypothesis that Avar shows lower degrees of phonological adaptation as compared to the other languages, which is likely due to its longer history of direct contact with Russian. This might have caused Avar to stop adapting loanwords sooner as a result of the increasing levels of bilingualism. The

second place in the hierarchy can be explained in a similar way, since Botlikh has long been an important administrative and trade center of the region, which also led to a more intense contact with Russian as compared to the other Andic settlements.

Another important observation concerns nasalization, which is a frequent adaptation strategy only in Akhvakh, Bagvalal and Chamalal. On the one hand, in these languages nasal vowels play a more important role in phonology. On the other hand, we cannot be sure that this difference can be attributed to linguistic factors alone, as all three dictionaries were co-authored by P. T. Magomedova. Nasalizations are frequent in Karata-Tukita and Tindi too, whose dictionaries were also co-authored by P. T. Magomedova. In the current state of research, it is impossible to determine whether different authors are inclined to different spelling of some acoustically close environments, like vowel + nasal and nasal vowel, or whether Avar-Andic languages actually differ in terms of frequency of nasal vowels in borrowings. Another example of a possible “author effect” can be observed by comparing the two Botlikh dictionaries available. Among other things, we noticed that, in the same lexemes, Saidova and Abusov (2012) mark ejectives more often than Alexeev and Azaev (2019) do.

Another essential point is that some changes demonstrate opposite tendencies in the alternation of syllabic structures. Examples of such opposite trends are final vowel epenthesis (7) and final vowel deletion (8).

(7) Final vowel epenthesis

- a. present
p-0-l-'a-ɛ:-0 (Russian) → p-a-l-'a-f:-i (Godoberi) ‘coat’
- b. absent
p-0-l-'a-ɛ: (Russian) → p-a-l-'a-f: (Bagvalal) ‘coat’

(8) Final vowel deletion

- a. present
k-A-l-'o-f-A (Russian) → k-a-l-'o-f-0 (Avar) ‘type of shoes’
- b. absent
k-A-l-'o-f-A (Russian) → kɪ-a-l-'u-f-a (Godoberi) ‘type of shoes’

If we consider such processes as a choice of each language to overcome constraints on some patterns in syllabic structure, we can divide the languages into two groups (Table 3).

Table 3. Grouping of languages by preference of final vowel epenthesis or elision.

Preferred change	Languages
final vowel epenthesis	Andi, Akhvakh, Godoberi, Tindi
final vowel deletion (apocope)	Avar, Bagvalal, Botlikh
inconsistent	Chamalal, Karata-Tukita

The presence of both final vowel epenthesis and final vowel deletion for the languages preferring epenthesis can seem misleading, but the reason for this is that the same *final vowel deletion* label is also used for the annotation of *jV* final syllable deletion after vowels, which, unlike elision of a single final vowel, is present in all Avar-Andic. As a result, this deletion does not produce a final closed syllable, which distinguishes it from regular final vowel deletion (9).

- (9) Final *jV* deletion
 k-O-n-s-ti-I-t-'u-ts-I-**j**-**A** (Russian) → ki-a-n-s-t-i-t-'u-ts-i-**0**-**0** (Godoberi) ‘constitution’

The presence of final vowel epenthesis in languages with preferable final vowel elision can be attributed to constraints on consonant clusters at the end of the word, but not to constraints on closed syllables in general (10).

- (10) Final vowel epenthesis after consonant clusters
 m-A-r-k-si-'i-z-m-**0** (Russian) → m-a-r-k-s-'i-z-m-**a** ‘Marxism’ (Botlikh)

The inconsistency of data for Chamalal according to the parameter chosen is attributable to the small number of examples available, so it is not possible to make any reliable conclusions. As for Karata-Tukita, the choice seems to be lexically determined.

If we exclude data from inconsistent languages, we could conclude that languages lower on the hierarchy are prone to final vowel elision rather than final vowel epenthesis, while languages higher on hierarchy behave in the opposite way. These changes could serve as an indirect indicator of linguistic mechanisms underlying the hierarchy of languages in Figure 4. The difference in the way languages adapt the end of the word shows that the syllabic structure of the target languages is one of the linguistic factors at play.

Discussion and concluding remarks

The paper has discussed strategies of phonological adaptation of Russian borrowings in Avar-Andic languages. A thorough phonological and morphological annotation of loanwords from the Avar-Andic database and their corresponding Russian lexemes allowed us to conduct a quantitative investigation of the number and types of changes that Russian borrowings have undergone in the languages of the sample. Information on time depth and frequency of use was retrieved from the RNC and included in the analysis.

The results of a mixed effect logistic regression showed a statistically significant difference between all dictionaries, which can be sorted in a hierarchy based on the predicted probabilities of change: Avar < Botlikh < Bagvalal < Andi < Godoberi < Karata-Tukita < Chamalal < Akhvakh < Tindi. This hierarchy does not show clear correlations with geographic distances and phylogenetic classifications. Importantly, however, the hierarchy supports our hypothesis that languages with a longer history of direct contact with Russian (Avar and Botlikh) present lower degrees of phonological adaptation. As we have discussed, being the languages spoken in important administrative centers of northern Dagestan, Avar

(to a greater extent) and Botlikh (to a lower extent) have historically undergone more intense contacts with Russian as compared to the other languages spoken in the region, which likely caused them to stop adapting loanwords sooner as a result of the increasing levels of bilingualism. It is also worth mentioning that the high prestige of Avar in the nineteenth century has led to its role as a mediator of Russian borrowings into Andic languages. Some traces of this are still visible in some of the borrowings in our sample, but in most cases it is hard to discriminate whether they actually underwent some modifications through Avar or not. A comparison with other East Caucasian languages could probably give us a clue for at least some of the lexemes, but this falls beyond the scope of this study.

The small sample size makes it hardly possible to draw any conclusions on the effect of population sizes on degrees of loanword adaptation. We found a small negative correlation between the hierarchy in Figure 4 and census data in Table 1 (Spearman correlation values: -0.13 for the 1926 census and -0.5 for the 2010 census). However, we cannot derive any theoretical implications based on this data.

Some of the variables included in the model, i.e., approximate time of borrowing and dictionary translation frequency, do not significantly affect the predictions of the number of changes. However, we still see an effect of these predictors in that earlier and more frequent borrowings on average present more changes. This tendency is not observed in Tindi and Chamalal, but this might be due to the smaller amount of data available for these languages, and to the underrepresentation of Soviet loanwords in these two dictionaries.

Another interesting observation that emerged from the analysis is that some sort of “author effect” could play a role in the clusterization of dictionaries. As we have seen, nasalization of vowels is represented regularly only in dictionaries by P. T. Magomedova, which might affect their closeness on the hierarchy.

In sum, the analysis has demonstrated that variation in degrees of phonological adaptation of borrowings depends on multiple (linguistic and sociolinguistic) factors. If, on the one hand, languages as they are represented in the available sources consistently and significantly differ among one another, their affinity in terms of adaptation strategies does not straightforwardly correspond to expectations based on other types of classifications. However, on the other hand, it is probably not that unexpected to detect the absence of certain local patterns in situations in which a lingua franca spreads across such a large area. Russian as an L2 in Daghestan is progressively expanding, and new borrowings will necessarily undergo fewer and fewer changes. In this sense our study contributes to capturing the evolution of certain linguistic processes that are doomed to disappear.

References

- Alekseev, M. E. (1988). *Sravnitel'no-istoričeskaja morfologija avaro-andijskix jazikov* [Historical-comparative morphology of Avar-Andic Languages]. Nauka.
- Alekseev, M. E., & Azaev, X. G. (2019). *Botliksko-russkij slovar'* [Botlikh-Russian dictionary]. Academia.
- Alexeyev, M., & Verhees, S. (Forth.). Botlikh. In Y. B. Koryakov, T. A. Maisak, & S. Kaye (Eds.), *The Caucasian Languages. An International Handbook. Volume 1. Nakh-Daghestanian languages: Nakh, Avar-Andic and Tsezic*. Mouton de Gruyter.

- Alimova, B. M., & Magomedov, D. M. (1993). *Botlixcy XIX – nač. XX v.* [Botlikhs of the 19th and beginning of the 20th centuries]. Daghestan Scientific Centre, Russian Academy of Sciences.
- Alisultanova, M. A. (2009). *Leksika andijskogo jazyka* [Lexicon of the Andi language]. *PhD thesis*. Institute of Language, Literature and Arts G. Cadasa, Daghestan Scientific Centre, Russian Academy of Sciences.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Brown, C. H. (1996). Lexical acculturation, areal diffusion, lingua francas, and bilingualism. *Language in Society*, 25(2), 261–282.
- Brown, C. H. (2011). The role of Nahuatl in the formation of Mesoamerica as a linguistic area. *Language Dynamics and Change*, 1(2), 171–204.
- Chirikba, V. A. (2008). The problem of the Caucasian Sprachbund. In P. Muysken (Ed.), *From Linguistic Areas to Areal Linguistics* (pp. 25–93). John Benjamins.
- Chumakina, M. (2009). Loanwords in Archi, a Nakh-Daghestanian of the North Caucasus. In M. Haspelmath, & U. Tadmor (Eds.), *Loanwords in the world's languages: a comparative handbook* (pp. 430–446). De Gruyter Mouton.
- Comrie, B., & Khalilov, M. (2009). Loanwords in Bezhta, a Nakh-Daghestanian of the North Caucasus. In M. Haspelmath, & U. Tadmor (Eds.), *Loanwords in the world's languages: a comparative handbook* (pp. 414–429). De Gruyter Mouton.
- Creissels, D. (Forth.). Northern Akhvakh. In Y. B. Koryakov, T. A. Maisak, & S. Kaye (Eds.), *The Caucasian Languages. An International Handbook. Volume 1. Nakh-Daghestanian languages: Nakh, Avar-Andic and Tsezic*. Mouton de Gruyter.
- Daniel, M., Chechuro, I., Verhees, S., & Dobrushina, N. (2021). Lingua francas as lexical donors: Evidence from Daghestan. *Language*, 97(3), 520–560.
- Dobrushina, N. (2013). How to study multilingualism of the past: Investigating traditional contact situations in Daghestan. *Journal of Sociolinguistics*, 17(3), 376–393.
- Dobrushina, N. (2023). Language ideology in an endogamous society: The case of Daghestan. *Journal of Sociolinguistics*, 27, 159–176.
- Dobrushina, N., & Kultepina, O. (2021). The rise of a lingua franca: The case of Russian in Dagestan. *International Journal of Bilingualism*, 25(1), 338–358.
- Dobrushina, N., & Moroz, G. (2021). The speakers of minority languages are more multilingual. *International Journal of Bilingualism*, 25(4), 921–938.
- Dobrushina, N., Staferova, D., & Belokon, A. (Eds.). (2017). *Atlas of multilingualism in Dagestan online*. Linguistic Convergence Laboratory, HSE. <https://multidagestan.com>
- Dobrushina, N. R., & Zakirova, A. N. (2019). Avarskij kak lingva franka: issledovanie v karatinskoj zone [Avar as a lingua franca: a study in and around Karata]. *Tomsk Journal of Linguistics and Anthropology*, 1(23), 44–55.
- Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2024). *Ethnologue: Languages of the World. Twenty-seventh edition*. SIL International. Online version: <http://www.ethnologue.com>
- Filatov, K., & Daniel, M. (2021). *DagSwadesh: 100 Swadesh lists from Daghestan. An online database of basic vocabulary divergence across neighbour villages*. Linguistic Convergence Laboratory, HSE University. <https://github.com/LingConLab/dagswadesh>

- Gimbatov, M. M. (2006). *Avarsko-russkij slovar' (rukopis')* [Avar-Russian dictionary (manuscript)]. Daghestan Scientific Centre, Russian Academy of Sciences. (Retrieved from <http://avar.me/>).
- Gudava, T. E. (1959). *Sravnitel'nyj analiz glagol'nyx osnov v avarskom i andijskom jazikax* [A comparative analysis of verb stems in Avar and Andic languages]. Institute of Language, Literature and Arts G. Cadasa, Daghestan Scientific Centre, Russian Academy of Sciences.
- Hayes, B. (1984). *The phonetics and phonology of Russian voicing assimilation*. In M. Aronoff, & R. T. Oehrle (Eds.), *Language Sound Structure* (pp. 318–328). MIT Press.
- Hammarström, H., Forkel, R., Haspelmath, M., & Bank, S. (2024). *Glottolog 5.0*. Max Planck Institute for Evolutionary Anthropology. <http://glottolog.org>
- Haspelmath, M., & Tadmor, U. (2009). The Loanword Typology project and the World Loanword Database. In M. Haspelmath, & U. Tadmor (Eds.), *Loanwords in the world's languages: a comparative handbook* (pp. 1–34). De Gruyter Mouton.
- Haugen, E. (1950). The analysis of linguistic borrowing. *Language*, 26, 210–231.
- Haugen, E. (1969). *The Norwegian Language in America: A Study in Bilingual Behavior*. University of Indiana Press.
- Iosad, P. (2012). Vowel reduction in Russian: No phonetics in phonology. *Journal of Linguistics*, 48(3), 521–571.
- Kerek, E., & Niemi, P. (2009). Russian orthography and learning to read. *Reading in a Foreign Language*, 21(1), 1–21.
- Koile, E., Chechuro, I., Moroz, G., & Daniel, M. (2022). Geography and language divergence: The case of Andic languages. *PLoS ONE*, 17(5): e0265460. <https://doi.org/10.1371/journal.pone.0265460>
- Koryakov, Y. B. (2006). *Atlas kavkazskix jazykov* [Atlas of the Caucasian Languages]. Institute of Linguistics, Russian Academy of Sciences.
- Lev-Ari, S., San Giacomo, M., & Peperkamp, S. (2014). The effect of domain prestige and interlocutors' bilingualism on loanword adaptations. *Journal of Sociolinguistics*, 18(5), 658–684.
- Magomedov, M. A., & Magomedova, M. M. (2021). Sociolingvističeskaja situacija v iskonnom areale rasprostraneniya andijskix jazykov [Sociolinguistic situation in the native area of diffusion of Andic languages]. *Sociolinguistics*, 3(7), 59–68.
- Magomedova, P. T. (1999). *Čamalinsko-russkij slovar'* [Chamalal-Russian dictionary]. Institute of Language, Literature and Arts G. Cadasa, Daghestan Scientific Centre, Russian Academy of Sciences.
- Magomedova, P. T. (2003). *Tindinsko-russkij slovar'* [Tindi-Russian dictionary]. Institute of Language, Literature and Arts G. Cadasa, Daghestan Scientific Centre, Russian Academy of Sciences.
- Magomedova, P. T. (2004). *Bagvalinsko-russkij slovar'* [Bagvalal-Russian dictionary]. Institute of Language, Literature and Arts G. Cadasa, Daghestan Scientific Centre, Russian Academy of Sciences.
- Magomedova, P. T., & Abdulaeva, I. A. (2007). *Axvaxsko-russkij slovar'* [Akhvakh-Russian dictionary]. Institute of Language, Literature and Arts G. Cadasa, Daghestan Scientific Centre, Russian Academy of Sciences.

- Magomedova, P. T., & Xalidova, R. Š. (2001). *Karatinsko-russkij slovar'* [Karata-Russian dictionary]. Institute of Language, Literature and Arts G. Cadasa, Daghestan Scientific Centre, Russian Academy of Sciences.
- Moroz, G. (2017). *lingtypology: easy mapping for Linguistic Typology*.
<https://CRAN.R-project.org/package=lingtypology>
- Moroz, G., Antsupova, A., Averin, A., Chistyakova, D., Chuprinko, K., Davidenko, A., Dolgodvorova, M., Fedorenko, N., Gnedina, T., Kuznetsov, G., Kuznetsova, S., Naccarato, C., Sadakov, I., Shkutko, Z., Tsyzoza, A., Verhees, S., & Zubkova, V. (2025). *Comparative Andic dictionary database*, v. 0.11.
https://github.com/phon-dicts-project/comparative_andic_dictionary_database
- Mudrak, O. A. (2020). *Andijskie osnovy. Etimologičeskij slovar'* [Andic stems. An etymological dictionary]. Jazyki narodov mira.
- Pasquereau, J. (Forth.). Karata. In Y. B. Koryakov, T. A. Maisak, & S. Kaye (Eds.), *The Caucasian Languages. An International Handbook. Volume 1. Nakh-Daghestanian languages: Nakh, Avar-Andic and Tsezic*. Mouton de Gruyter.
- Poplack, S., & Sankoff, D. (1984). Borrowing: The synchrony of integration. *Linguistics*, 22, 99–135.
- Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, 26, 47–104.
- R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rushworth, A. (2022). *inspectdf: Inspection, comparison and visualisation of data frames* (R Package Version 0.0.12). <https://alastairrushworth.github.io/inspectdf/>
- Saidova, P. A., & Abusov, M. G. (2012). *Botliksko-russkij slovar'* [Botlikh-Russian dictionary]. Institute of Language, Literature and Arts G. Cadasa, Daghestan Scientific Centre, Russian Academy of Sciences.
- Salimov, X. S. (2010). *Gagatlinskij govor andijskogo jazyka* [The Gagatli dialect of Andi]. Institute of Language, Literature and Arts G. Cadasa, Daghestan Scientific Centre, Russian Academy of Sciences.
- Savčuk, S. O., Arxangelskij, T. A., Bonč-Osmolovskaja, A. A., Donina, O. V., Kuznecova, Y. N., Ljaševskaja, O. N., Orexov, B. V., & Podrjadčikova, M. V. (2024). Nacional'nyj korpus russkogo jazyka 2.0: novye vozmožnosti i perspektivy razvitija. *Voprosy jazykoznanija*, 2, 7–34.
- Schulze, W. (2009). *Languages in the Caucasus (manuscript)*.
<http://wschulze.userweb.mwn.de/lgxcauc.pdf>
- Siražudinov, R. M. (2015). Osobennosti osvoenija russkix zaimstvovanij v godoberinskom jazyke [Peculiarities of the adaptation of Russian borrowings in Godoberi]. *Filologičeskie nauki. Voprosy teorii i praktiki*, 8(50), 178–181.
- Thomason, S. G. (2001). *Language Contact. An Introduction*. Edinburgh University Press.
- Volkova, N. G. (1967). Voprosy dvujazyčija na Severnom Kavkaze [Issues of bilingualism in the Northern Caucasus]. *Sovetskaja etnografija*, 1, 27–40.