Supplementary materials to the paper 'Phonological adaptation of Russian borrowings in Avar-Andic languages'

25.04.2025

Table of contents

1 Data

Read the dataset. The whole document is based on manipulations of different parts of the same dataset.

```
library(tidyverse)

read_csv("data.csv", show_col_types = FALSE) |>
  mutate(language_ref = str_c(language, ": ", reference)) ->
  df
```

The dataset consists of 21749 observations with the following columns:

- language: language;
- reference: source of data;
- 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.

Here are the most frequent values in the subsample of variables (variables `dictionary_translation`, `lemma frequency_ipm`, `russian_ipa` and `target_ipa` are omitted due to the huge amount of values).

```
library(inspectdf)

df ▷
  mutate_all(as.factor) ▷
  select(language, reference, change, total_units_per_word, changes_per_word, time_of_borrowing, type_of_change) ▷
  rename(`(1) language` = language,
         `(2) reference` = reference,
         `(3) change` = change,
         `(4) total units per word` = total_units_per_word,
         `(5) changes per word` = changes_per_word,
         `(6) type of change` = type_of_change,
         `(7) time of borrowing` = time_of_borrowing) ▷
  inspect_cat()    ▷
  show_plot()+
  labs(title = NULL, subtitle = NULL, text = element_text(size = 30))+
  theme(axis.text.y = element_text(hjust = 0))
```
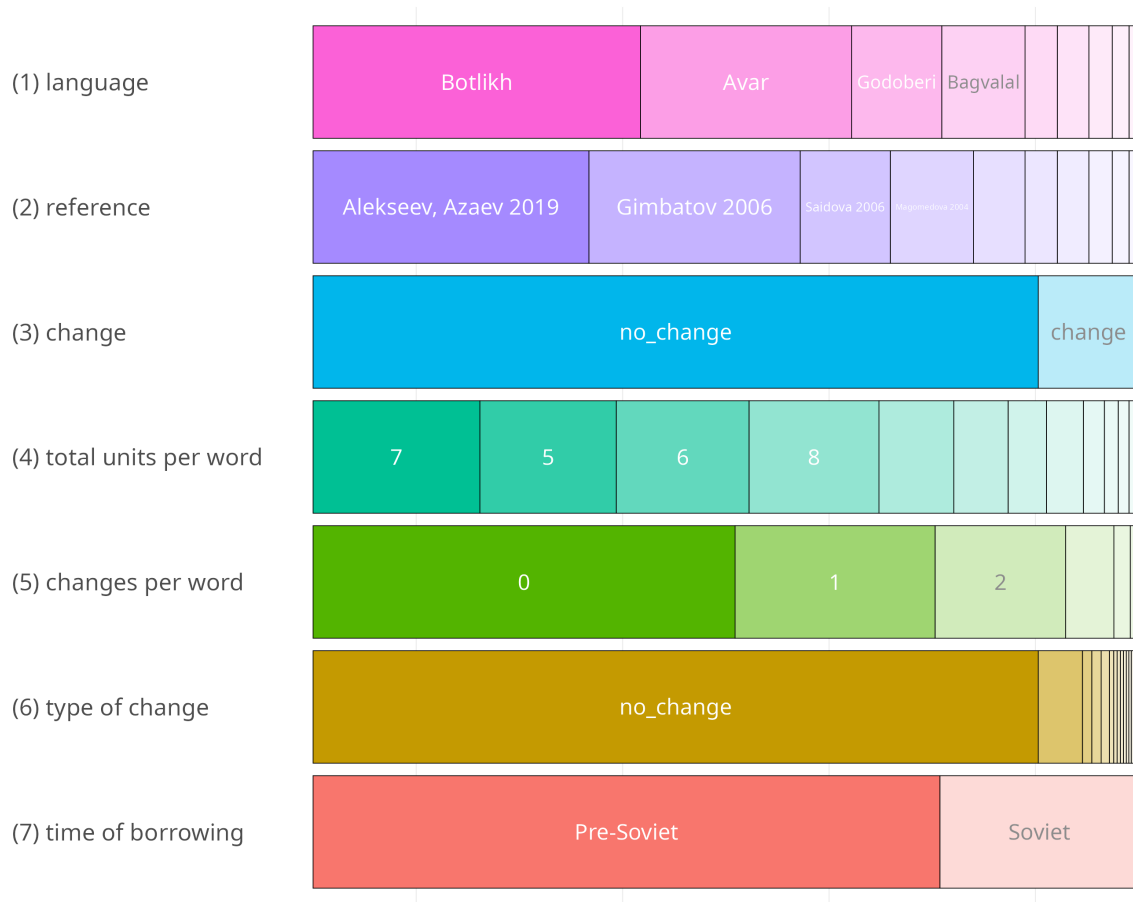
Figure 1: Frequency of values for each variable in the dataset. Gray segments are missing values.

2 Distribution of changes across languages

```
df ▷
  select(language_ref, type_of_change, time_of_borrowing) ▷
  filter(type_of_change ≠ "no_change",
         type_of_change ≠ "other") ▷
  mutate(type_of_change = str_split(type_of_change, ", ")) ▷
  unnest(type_of_change) ▷
  add_count(language_ref, type_of_change) ▷
  count(language_ref, type_of_change, time_of_borrowing, n) ▷
  mutate(type_of_change = tidytext::reorder_within(type_of_change,  n, language_ref),
         time_of_borrowing = fct_relevel(time_of_borrowing, "Soviet")) ▷
```

3

```r
ggplot(aes(nn, type_of_change, fill = time_of_borrowing))+
geom_col()+
facet_wrap(~language_ref, scales = "free", ncol = 2)+
tidytext::scale_y_reordered()+
scale_x_continuous(breaks = scales::breaks_pretty())+
labs(x = NULL,
  y = NULL,
  fill = NULL)+
theme_minimal()+
theme(legend.position = "bottom")
```

Figure 2: Frequency of each change by language

5

Here are the raw counts of all changes by language, dictionary and time of borrowing:

```
df |>
  count(language, reference, time_of_borrowing, type_of_change) |>
  arrange(language, reference, time_of_borrowing, -n) |>
  knitr::kable()
```

Table 1: Raw counts of all changes by language, dictionary and time of borrowing.

language	reference	time_of_borrowing	type_of_change	n
Akhvakh	Magomedova, Abdulayeva 2007	Pre-Soviet	no_change	499
Akhvakh	Magomedova, Abdulayeva 2007	Pre-Soviet	other	85
Akhvakh	Magomedova, Abdulayeva 2007	Pre-Soviet	final vowel epenthesis	44
Akhvakh	Magomedova, Abdulayeva 2007	Pre-Soviet	nasalization	23
Akhvakh	Magomedova, Abdulayeva 2007	Pre-Soviet	middle consonant deletion	21
Akhvakh	Magomedova, Abdulayeva 2007	Pre-Soviet	middle vowel epenthesis	16
Akhvakh	Magomedova, Abdulayeva 2007	Pre-Soviet	middle vowel deletion	8
Akhvakh	Magomedova, Abdulayeva 2007	Pre-Soviet	ejectivization	6
Akhvakh	Magomedova, Abdulayeva 2007	Pre-Soviet	final consonant deletion	6
Akhvakh	Magomedova, Abdulayeva 2007	Pre-Soviet	final vowel deletion	6
Akhvakh	Magomedova, Abdulayeva 2007	Pre-Soviet	metathesis	4
Akhvakh	Magomedova, Abdulayeva 2007	Pre-Soviet	initial vowel epenthesis	3
Akhvakh	Magomedova, Abdulayeva 2007	Pre-Soviet	devoicing	2
Akhvakh	Magomedova, Abdulayeva 2007	Pre-Soviet	initial consonant epenthesis	1
Akhvakh	Magomedova, Abdulayeva 2007	Pre-Soviet	voicing	1
Akhvakh	Magomedova, Abdulayeva 2007	Soviet	no_change	87

language	reference	time_of_borrowing	type_of_change	n
Akhvakh	Magomedova, Abdulayeva 2007	Soviet	other	13
Akhvakh	Magomedova, Abdulayeva 2007	Soviet	final vowel epenthesis	7
Akhvakh	Magomedova, Abdulayeva 2007	Soviet	nasalization	7
Akhvakh	Magomedova, Abdulayeva 2007	Soviet	middle vowel epenthesis	3
Akhvakh	Magomedova, Abdulayeva 2007	Soviet	ejectivization	2
Akhvakh	Magomedova, Abdulayeva 2007	Soviet	middle consonant deletion	2
Akhvakh	Magomedova, Abdulayeva 2007	Soviet	devoicing	1
Akhvakh	Magomedova, Abdulayeva 2007	Soviet	initial consonant epenthesis	1
Akhvakh	Magomedova, Abdulayeva 2007	Soviet	metathesis	1
Akhvakh	Magomedova, Abdulayeva 2007	Soviet	middle consonant epenthesis	1
Akhvakh	Magomedova, Abdulayeva 2007	Soviet	middle vowel deletion	1
Andi	Salimov 2010	Pre-Soviet	no_change	544
Andi	Salimov 2010	Pre-Soviet	other	82
Andi	Salimov 2010	Pre-Soviet	final vowel epenthesis	35
Andi	Salimov 2010	Pre-Soviet	middle vowel epenthesis	13
Andi	Salimov 2010	Pre-Soviet	final vowel deletion	8
Andi	Salimov 2010	Pre-Soviet	ejectivization	7
Andi	Salimov 2010	Pre-Soviet	middle consonant epenthesis	7
Andi	Salimov 2010	Pre-Soviet	initial vowel epenthesis	5
Andi	Salimov 2010	Pre-Soviet	middle consonant deletion	4
Andi	Salimov 2010	Pre-Soviet	middle vowel deletion	4
Andi	Salimov 2010	Pre-Soviet	devoicing	3
Andi	Salimov 2010	Pre-Soviet	final consonant deletion	3
Andi	Salimov 2010	Pre-Soviet	nasalization	2
Andi	Salimov 2010	Pre-Soviet	metathesis	1
Andi	Salimov 2010	Soviet	no_change	98
Andi	Salimov 2010	Soviet	other	10
Andi	Salimov 2010	Soviet	final vowel epenthesis	3
Andi	Salimov 2010	Soviet	middle consonant deletion	1
Andi	Salimov 2010	Soviet	middle consonant epenthesis	1

language	reference	time_of_borrowing	type_of_change	n
Avar	Gimbatov 2006	Pre-Soviet	no_change	4035
Avar	Gimbatov 2006	Pre-Soviet	other	55
Avar	Gimbatov 2006	Pre-Soviet	final vowel deletion	16
Avar	Gimbatov 2006	Pre-Soviet	ejectivization	11
Avar	Gimbatov 2006	Pre-Soviet	middle vowel epenthesis	7
Avar	Gimbatov 2006	Pre-Soviet	middle consonant epenthesis	6
Avar	Gimbatov 2006	Pre-Soviet	middle vowel deletion	6
Avar	Gimbatov 2006	Pre-Soviet	middle consonant deletion	5
Avar	Gimbatov 2006	Pre-Soviet	final consonant deletion	4
Avar	Gimbatov 2006	Pre-Soviet	devoicing	2
Avar	Gimbatov 2006	Pre-Soviet	initial consonant epenthesis	2
Avar	Gimbatov 2006	Pre-Soviet	voicing	2
Avar	Gimbatov 2006	Pre-Soviet	final vowel epenthesis	1
Avar	Gimbatov 2006	Pre-Soviet	initial vowel epenthesis	1
Avar	Gimbatov 2006	Pre-Soviet	metathesis	1
Avar	Gimbatov 2006	Pre-Soviet	nasalization	1
Avar	Gimbatov 2006	Soviet	no_change	1396
Avar	Gimbatov 2006	Soviet	other	8
Avar	Gimbatov 2006	Soviet	ejectivization	3
Avar	Gimbatov 2006	Soviet	devoicing	1
Avar	Gimbatov 2006	Soviet	final vowel deletion	1
Avar	Gimbatov 2006	Soviet	middle vowel epenthesis	1
Bagvalal	Magomedova 2004	Pre-Soviet	no_change	1504
Bagvalal	Magomedova 2004	Pre-Soviet	other	119
Bagvalal	Magomedova 2004	Pre-Soviet	middle vowel epenthesis	23
Bagvalal	Magomedova 2004	Pre-Soviet	nasalization	22
Bagvalal	Magomedova 2004	Pre-Soviet	final vowel deletion	18
Bagvalal	Magomedova 2004	Pre-Soviet	middle consonant deletion	14
Bagvalal	Magomedova 2004	Pre-Soviet	initial vowel epenthesis	9
Bagvalal	Magomedova 2004	Pre-Soviet	metathesis	9
Bagvalal	Magomedova 2004	Pre-Soviet	final consonant deletion	8
Bagvalal	Magomedova 2004	Pre-Soviet	middle vowel deletion	8
Bagvalal	Magomedova 2004	Pre-Soviet	ejectivization	7
Bagvalal	Magomedova 2004	Pre-Soviet	final vowel epenthesis	7
Bagvalal	Magomedova 2004	Pre-Soviet	middle consonant epenthesis	6
Bagvalal	Magomedova 2004	Pre-Soviet	voicing	5
Bagvalal	Magomedova 2004	Pre-Soviet	devoicing	3
Bagvalal	Magomedova 2004	Pre-Soviet	initial consonant epenthesis	1
Bagvalal	Magomedova 2004	Soviet	no_change	374
Bagvalal	Magomedova 2004	Soviet	other	24
Bagvalal	Magomedova 2004	Soviet	nasalization	12

language	reference	time_of_borrowing	type_of_change	n
Bagvalal	Magomedova 2004	Soviet	middle vowel epenthesis	5
Bagvalal	Magomedova 2004	Soviet	middle consonant deletion	3
Bagvalal	Magomedova 2004	Soviet	final consonant deletion	2
Bagvalal	Magomedova 2004	Soviet	metathesis	2
Bagvalal	Magomedova 2004	Soviet	middle vowel deletion	2
Bagvalal	Magomedova 2004	Soviet	devoicing	1
Bagvalal	Magomedova 2004	Soviet	final vowel deletion	1
Bagvalal	Magomedova 2004	Soviet	final vowel epenthesis	1
Bagvalal	Magomedova 2004	Soviet	initial consonant epenthesis	1
Bagvalal	Magomedova 2004	Soviet	middle consonant epenthesis	1
Bagvalal	Magomedova 2004	Soviet	voicing	1
Botlikh	Alekseev, Azaev 2019	Pre-Soviet	no_change	4570
Botlikh	Alekseev, Azaev 2019	Pre-Soviet	other	253
Botlikh	Alekseev, Azaev 2019	Pre-Soviet	final vowel deletion	96
Botlikh	Alekseev, Azaev 2019	Pre-Soviet	middle vowel epenthesis	71
Botlikh	Alekseev, Azaev 2019	Pre-Soviet	initial vowel epenthesis	36
Botlikh	Alekseev, Azaev 2019	Pre-Soviet	middle consonant deletion	22
Botlikh	Alekseev, Azaev 2019	Pre-Soviet	final consonant deletion	21
Botlikh	Alekseev, Azaev 2019	Pre-Soviet	metathesis	17
Botlikh	Alekseev, Azaev 2019	Pre-Soviet	final vowel epenthesis	16
Botlikh	Alekseev, Azaev 2019	Pre-Soviet	middle vowel deletion	11
Botlikh	Alekseev, Azaev 2019	Pre-Soviet	middle consonant epenthesis	7
Botlikh	Alekseev, Azaev 2019	Pre-Soviet	devoicing	3
Botlikh	Alekseev, Azaev 2019	Pre-Soviet	nasalization	2
Botlikh	Alekseev, Azaev 2019	Pre-Soviet	voicing	2
Botlikh	Alekseev, Azaev 2019	Pre-Soviet	ejectivization	1
Botlikh	Alekseev, Azaev 2019	Pre-Soviet	initial consonant epenthesis	1
Botlikh	Alekseev, Azaev 2019	Soviet	no_change	1982
Botlikh	Alekseev, Azaev 2019	Soviet	other	70
Botlikh	Alekseev, Azaev 2019	Soviet	middle vowel epenthesis	32
Botlikh	Alekseev, Azaev 2019	Soviet	final vowel deletion	17
Botlikh	Alekseev, Azaev 2019	Soviet	middle consonant deletion	8
Botlikh	Alekseev, Azaev 2019	Soviet	final vowel epenthesis	7
Botlikh	Alekseev, Azaev 2019	Soviet	final consonant deletion	6
Botlikh	Alekseev, Azaev 2019	Soviet	middle consonant epenthesis	4
Botlikh	Alekseev, Azaev 2019	Soviet	middle vowel deletion	4
Botlikh	Alekseev, Azaev 2019	Soviet	devoicing	3
Botlikh	Alekseev, Azaev 2019	Soviet	initial vowel epenthesis	2
Botlikh	Alekseev, Azaev 2019	Soviet	metathesis	2
Botlikh	Alekseev, Azaev 2019	Soviet	nasalization	1
Botlikh	Saidova, Abusov 2012	Pre-Soviet	no_change	874

language	reference	time_of_borrowing	type_of_change	n
Botlikh	Saidova, Abusov 2012	Pre-Soviet	other	72
Botlikh	Saidova, Abusov 2012	Pre-Soviet	final vowel deletion	13
Botlikh	Saidova, Abusov 2012	Pre-Soviet	middle vowel epenthesis	8
Botlikh	Saidova, Abusov 2012	Pre-Soviet	middle consonant deletion	7
Botlikh	Saidova, Abusov 2012	Pre-Soviet	metathesis	6
Botlikh	Saidova, Abusov 2012	Pre-Soviet	devoicing	5
Botlikh	Saidova, Abusov 2012	Pre-Soviet	initial vowel epenthesis	5
Botlikh	Saidova, Abusov 2012	Pre-Soviet	ejectivization	4
Botlikh	Saidova, Abusov 2012	Pre-Soviet	final consonant deletion	4
Botlikh	Saidova, Abusov 2012	Pre-Soviet	final vowel epenthesis	4
Botlikh	Saidova, Abusov 2012	Pre-Soviet	middle vowel deletion	2
Botlikh	Saidova, Abusov 2012	Pre-Soviet	initial consonant epenthesis	1
Botlikh	Saidova, Abusov 2012	Pre-Soviet	middle consonant epenthesis	1
Botlikh	Saidova, Abusov 2012	Pre-Soviet	voicing	1
Botlikh	Saidova, Abusov 2012	Soviet	no_change	311
Botlikh	Saidova, Abusov 2012	Soviet	other	20
Botlikh	Saidova, Abusov 2012	Soviet	middle vowel epenthesis	6
Botlikh	Saidova, Abusov 2012	Soviet	final vowel epenthesis	3
Botlikh	Saidova, Abusov 2012	Soviet	devoicing	2
Botlikh	Saidova, Abusov 2012	Soviet	final consonant deletion	2
Botlikh	Saidova, Abusov 2012	Soviet	metathesis	2
Botlikh	Saidova, Abusov 2012	Soviet	final vowel deletion	1
Botlikh	Saidova, Abusov 2012	Soviet	initial vowel epenthesis	1
Botlikh	Saidova, Abusov 2012	Soviet	middle consonant deletion	1
Botlikh	Saidova, Abusov 2012	Soviet	middle consonant epenthesis	1
Chamalal	Magomedova 1999	Pre-Soviet	no_change	154
Chamalal	Magomedova 1999	Pre-Soviet	other	24
Chamalal	Magomedova 1999	Pre-Soviet	nasalization	6
Chamalal	Magomedova 1999	Pre-Soviet	middle vowel deletion	5
Chamalal	Magomedova 1999	Pre-Soviet	middle consonant deletion	4
Chamalal	Magomedova 1999	Pre-Soviet	final vowel deletion	3
Chamalal	Magomedova 1999	Pre-Soviet	final consonant deletion	2
Chamalal	Magomedova 1999	Pre-Soviet	middle consonant epenthesis	2
Chamalal	Magomedova 1999	Pre-Soviet	devoicing	1
Chamalal	Magomedova 1999	Pre-Soviet	ejectivization	1
Chamalal	Magomedova 1999	Pre-Soviet	initial vowel epenthesis	1
Chamalal	Magomedova 1999	Pre-Soviet	middle vowel epenthesis	1
Chamalal	Magomedova 1999	Soviet	no_change	33
Chamalal	Magomedova 1999	Soviet	other	8
Chamalal	Magomedova 1999	Soviet	nasalization	4
Chamalal	Magomedova 1999	Soviet	ejectivization	2

language	reference	time_of_borrowing	type_of_change	n
Chamalal	Magomedova 1999	Soviet	metathesis	2
Chamalal	Magomedova 1999	Soviet	middle vowel deletion	1
Godoberi	Saidova 2006	Pre-Soviet	no_change	1465
Godoberi	Saidova 2006	Pre-Soviet	other	135
Godoberi	Saidova 2006	Pre-Soviet	final vowel epenthesis	68
Godoberi	Saidova 2006	Pre-Soviet	palatalization	67
Godoberi	Saidova 2006	Pre-Soviet	middle vowel epenthesis	33
Godoberi	Saidova 2006	Pre-Soviet	final vowel deletion	17
Godoberi	Saidova 2006	Pre-Soviet	initial vowel epenthesis	12
Godoberi	Saidova 2006	Pre-Soviet	metathesis	10
Godoberi	Saidova 2006	Pre-Soviet	middle consonant deletion	9
Godoberi	Saidova 2006	Pre-Soviet	middle consonant epenthesis	8
Godoberi	Saidova 2006	Pre-Soviet	ejectivization	6
Godoberi	Saidova 2006	Pre-Soviet	final consonant deletion	6
Godoberi	Saidova 2006	Pre-Soviet	palatalization, ejectivization	5
Godoberi	Saidova 2006	Pre-Soviet	devoicing	3
Godoberi	Saidova 2006	Pre-Soviet	voicing	3
Godoberi	Saidova 2006	Pre-Soviet	middle vowel deletion	2
Godoberi	Saidova 2006	Pre-Soviet	initial consonant deletion	1
Godoberi	Saidova 2006	Pre-Soviet	initial consonant epenthesis	1
Godoberi	Saidova 2006	Pre-Soviet	nasalization	1
Godoberi	Saidova 2006	Pre-Soviet	palatalization, devoicing	1
Godoberi	Saidova 2006	Soviet	no_change	432
Godoberi	Saidova 2006	Soviet	other	32
Godoberi	Saidova 2006	Soviet	final vowel epenthesis	23
Godoberi	Saidova 2006	Soviet	palatalization	13
Godoberi	Saidova 2006	Soviet	middle vowel epenthesis	8
Godoberi	Saidova 2006	Soviet	ejectivization	3
Godoberi	Saidova 2006	Soviet	devoicing	2
Godoberi	Saidova 2006	Soviet	final consonant deletion	2
Godoberi	Saidova 2006	Soviet	final vowel deletion	1
Godoberi	Saidova 2006	Soviet	initial vowel epenthesis	1
Godoberi	Saidova 2006	Soviet	metathesis	1
Godoberi	Saidova 2006	Soviet	middle consonant deletion	1
Godoberi	Saidova 2006	Soviet	middle vowel deletion	1
Godoberi	Saidova 2006	Soviet	nasalization	1
Karata-Tukita	Magomedova, Khalidova 2001	Pre-Soviet	no_change	398
Karata-Tukita	Magomedova, Khalidova 2001	Pre-Soviet	other	72

language	reference	time_of_borrowing	type_of_change	n
Karata-Tukita	Magomedova, Khalidova 2001	Pre-Soviet	final vowel deletion	14
Karata-Tukita	Magomedova, Khalidova 2001	Pre-Soviet	middle vowel epenthesis	12
Karata-Tukita	Magomedova, Khalidova 2001	Pre-Soviet	final vowel epenthesis	8
Karata-Tukita	Magomedova, Khalidova 2001	Pre-Soviet	middle vowel deletion	8
Karata-Tukita	Magomedova, Khalidova 2001	Pre-Soviet	initial vowel epenthesis	6
Karata-Tukita	Magomedova, Khalidova 2001	Pre-Soviet	middle consonant deletion	5
Karata-Tukita	Magomedova, Khalidova 2001	Pre-Soviet	middle consonant epenthesis	5
Karata-Tukita	Magomedova, Khalidova 2001	Pre-Soviet	nasalization	5
Karata-Tukita	Magomedova, Khalidova 2001	Pre-Soviet	devoicing	4
Karata-Tukita	Magomedova, Khalidova 2001	Pre-Soviet	ejectivization	4
Karata-Tukita	Magomedova, Khalidova 2001	Pre-Soviet	final consonant deletion	4
Karata-Tukita	Magomedova, Khalidova 2001	Pre-Soviet	metathesis	2
Karata-Tukita	Magomedova, Khalidova 2001	Pre-Soviet	voicing	2
Karata-Tukita	Magomedova, Khalidova 2001	Pre-Soviet	initial consonant deletion	1
Karata-Tukita	Magomedova, Khalidova 2001	Pre-Soviet	initial consonant epenthesis	1
Karata-Tukita	Magomedova, Khalidova 2001	Soviet	no_change	49
Karata-Tukita	Magomedova, Khalidova 2001	Soviet	other	7
Karata-Tukita	Magomedova, Khalidova 2001	Soviet	ejectivization	2
Karata-Tukita	Magomedova, Khalidova 2001	Soviet	devoicing	1
Karata-Tukita	Magomedova, Khalidova 2001	Soviet	final consonant deletion	1

language	reference	time_of_borrowing	type_of_change	n
Karata-Tukita	Magomedova, Khalidova 2001	Soviet	final vowel deletion	1
Karata-Tukita	Magomedova, Khalidova 2001	Soviet	final vowel epenthesis	1
Karata-Tukita	Magomedova, Khalidova 2001	Soviet	middle consonant deletion	1
Karata-Tukita	Magomedova, Khalidova 2001	Soviet	middle vowel epenthesis	1
Tindi	Magomedova 2003	Pre-Soviet	no_change	278
Tindi	Magomedova 2003	Pre-Soviet	other	67
Tindi	Magomedova 2003	Pre-Soviet	final vowel epenthesis	16
Tindi	Magomedova 2003	Pre-Soviet	middle consonant epenthesis	7
Tindi	Magomedova 2003	Pre-Soviet	middle vowel epenthesis	6
Tindi	Magomedova 2003	Pre-Soviet	final consonant deletion	4
Tindi	Magomedova 2003	Pre-Soviet	final vowel deletion	4
Tindi	Magomedova 2003	Pre-Soviet	middle consonant deletion	4
Tindi	Magomedova 2003	Pre-Soviet	middle vowel deletion	4
Tindi	Magomedova 2003	Pre-Soviet	nasalization	4
Tindi	Magomedova 2003	Pre-Soviet	omission of palatalization	4
Tindi	Magomedova 2003	Pre-Soviet	initial vowel epenthesis	3
Tindi	Magomedova 2003	Pre-Soviet	ejectivization	2
Tindi	Magomedova 2003	Pre-Soviet	palatalization	2
Tindi	Magomedova 2003	Pre-Soviet	devoicing	1
Tindi	Magomedova 2003	Pre-Soviet	metathesis	1
Tindi	Magomedova 2003	Soviet	no_change	20
Tindi	Magomedova 2003	Soviet	other	8
Tindi	Magomedova 2003	Soviet	final vowel epenthesis	3
Tindi	Magomedova 2003	Soviet	ejectivization	2
Tindi	Magomedova 2003	Soviet	devoicing	1
Tindi	Magomedova 2003	Soviet	middle consonant deletion	1

3 Modeling

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:

```
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.

```
library(lme4)
library(lmerTest)
```

```
Attaching package: 'lmerTest'

The following object is masked from 'package:lme4':

    lmer

The following object is masked from 'package:stats':

    step
```

```
df |>
  mutate(change = if_else(change == "no_change", 0, 1),
         lemma_frequency_ipm = if_else(is.na(lemma_frequency_ipm), log(0.0001), log(lemma_frequency_ipm)),
         time_of_borrowing = fct_relevel(time_of_borrowing, "Soviet"),
         language_ref = fct_relevel(language_ref, "Avar: Gimbatov 2006")) |>
  glmer(change ~ time_of_borrowing:language_ref+lemma_frequency_ipm + (1|russian_source_lexeme),
        family = "binomial",
        control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=9000)),
        data = _) ->
  fit
```

```
fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

```
summary(fit)
```

```
Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [glmerMod]
 Family: binomial  ( logit )
```

```
Formula: change ~ time_of_borrowing:language_ref + lemma_frequency_ipm +
    (1 | russian_source_lexeme)
   Data:
mutate(df, change = if_else(change == "no_change", 0, 1), lemma_frequency_ipm = if_else(is.na(lemma_frequency_ipm),
    log(1e-04), log(lemma_frequency_ipm)), time_of_borrowing = fct_relevel(time_of_borrowing,
    "Soviet"), language_ref = fct_relevel(language_ref, "Avar: Gimbatov 2006"))
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 9000))


     AIC      BIC    logLik -2*log(L)  df.resid
  14204.7  14380.4  -7080.3   14160.7     21727


Scaled residuals:
    Min     1Q  Median     3Q     Max
-1.3892 -0.3706 -0.2647 -0.1341  8.8987


Random effects:
 Groups               Name        Variance Std.Dev.
 russian_source_lexeme (Intercept) 0.6014   0.7755
Number of obs: 21749, groups:  russian_source_lexeme, 1343


Fixed effects:
                                                                                Estimate

(Intercept)                                                                     -1.26161
lemma_frequency_ipm                                                             -0.02144
time_of_borrowingSoviet:language_refAvar: Gimbatov 2006                         -3.56512
time_of_borrowingPre-Soviet:language_refAvar: Gimbatov 2006                     -2.47470
time_of_borrowingSoviet:language_refAkhvakh: Magomedova, Abdulayeva 2007        -0.22147
time_of_borrowingPre-Soviet:language_refAkhvakh: Magomedova, Abdulayeva 2007     0.11087
time_of_borrowingSoviet:language_refAndi: Salimov 2010                          -0.97695
time_of_borrowingPre-Soviet:language_refAndi: Salimov 2010                      -0.27689
time_of_borrowingSoviet:language_refBagvalal: Magomedova 2004                   -0.92745
time_of_borrowingPre-Soviet:language_refBagvalal: Magomedova 2004               -0.70873
time_of_borrowingSoviet:language_refBotlikh: Alekseev, Azaev 2019              -1.50719
time_of_borrowingPre-Soviet:language_refBotlikh: Alekseev, Azaev 2019           -0.99937
time_of_borrowingSoviet:language_refBotlikh: Saidova, Abusov 2012              -1.14449
time_of_borrowingPre-Soviet:language_refBotlikh: Saidova, Abusov 2012           -0.88194
time_of_borrowingSoviet:language_refChamalal: Magomedova 1999                  -0.04652
time_of_borrowingPre-Soviet:language_refChamalal: Magomedova 1999               -0.26149
time_of_borrowingSoviet:language_refGodoberi: Saidova 2006                      -0.56304
time_of_borrowingPre-Soviet:language_refGodoberi: Saidova 2006                  -0.28145
time_of_borrowingSoviet:language_refKarata-Tukita: Magomedova, Khalidova 2001   -0.60688
time_of_borrowingPre-Soviet:language_refKarata-Tukita: Magomedova, Khalidova 2001 -0.17274
time_of_borrowingSoviet:language_refTindi: Magomedova 2003                       0.35786
```

	Std. Error
(Intercept)	0.13444
lemma_frequency_ipm	0.01166
time_of_borrowingSoviet:language_refAvar: Gimbatov 2006	0.30481
time_of_borrowingPre-Soviet:language_refAvar: Gimbatov 2006	0.15920
time_of_borrowingSoviet:language_refAkhvakh: Magomedova, Abdulayeva 2007	0.27497
time_of_borrowingPre-Soviet:language_refAkhvakh: Magomedova, Abdulayeva 2007	0.15327
time_of_borrowingSoviet:language_refAndi: Salimov 2010	0.35228
time_of_borrowingPre-Soviet:language_refAndi: Salimov 2010	0.15935
time_of_borrowingSoviet:language_refBagvalal: Magomedova 2004	0.21907
time_of_borrowingPre-Soviet:language_refBagvalal: Magomedova 2004	0.14507
time_of_borrowingSoviet:language_refBotlikh: Alekseev, Azaev 2019	0.16453
time_of_borrowingPre-Soviet:language_refBotlikh: Alekseev, Azaev 2019	0.13793
time_of_borrowingSoviet:language_refBotlikh: Saidova, Abusov 2012	0.23821
time_of_borrowingPre-Soviet:language_refBotlikh: Saidova, Abusov 2012	0.16196
time_of_borrowingSoviet:language_refChamalal: Magomedova 1999	0.38446
time_of_borrowingPre-Soviet:language_refChamalal: Magomedova 1999	0.22138
time_of_borrowingSoviet:language_refGodoberi: Saidova 2006	0.19729
time_of_borrowingPre-Soviet:language_refGodoberi: Saidova 2006	0.14043
time_of_borrowingSoviet:language_refKarata-Tukita: Magomedova, Khalidova 2001	0.36721
time_of_borrowingPre-Soviet:language_refKarata-Tukita: Magomedova, Khalidova 2001	0.16377
time_of_borrowingSoviet:language_refTindi: Magomedova 2003	0.44747

	z value
(Intercept)	-9.384
lemma_frequency_ipm	-1.838
time_of_borrowingSoviet:language_refAvar: Gimbatov 2006	-11.696
time_of_borrowingPre-Soviet:language_refAvar: Gimbatov 2006	-15.545
time_of_borrowingSoviet:language_refAkhvakh: Magomedova, Abdulayeva 2007	-0.805
time_of_borrowingPre-Soviet:language_refAkhvakh: Magomedova, Abdulayeva 2007	0.723
time_of_borrowingSoviet:language_refAndi: Salimov 2010	-2.773
time_of_borrowingPre-Soviet:language_refAndi: Salimov 2010	-1.738
time_of_borrowingSoviet:language_refBagvalal: Magomedova 2004	-4.234
time_of_borrowingPre-Soviet:language_refBagvalal: Magomedova 2004	-4.886
time_of_borrowingSoviet:language_refBotlikh: Alekseev, Azaev 2019	-9.161
time_of_borrowingPre-Soviet:language_refBotlikh: Alekseev, Azaev 2019	-7.246
time_of_borrowingSoviet:language_refBotlikh: Saidova, Abusov 2012	-4.805
time_of_borrowingPre-Soviet:language_refBotlikh: Saidova, Abusov 2012	-5.446
time_of_borrowingSoviet:language_refChamalal: Magomedova 1999	-0.121
time_of_borrowingPre-Soviet:language_refChamalal: Magomedova 1999	-1.181
time_of_borrowingSoviet:language_refGodoberi: Saidova 2006	-2.854
time_of_borrowingPre-Soviet:language_refGodoberi: Saidova 2006	-2.004
time_of_borrowingSoviet:language_refKarata-Tukita: Magomedova, Khalidova 2001	-1.653
time_of_borrowingPre-Soviet:language_refKarata-Tukita: Magomedova, Khalidova 2001	-1.055

```
time_of_borrowingSoviet:language_refTindi: Magomedova 2003                              0.800
                                                                                      Pr(>|z|)
(Intercept)                                                                            < 2e-16
lemma_frequency_ipm                                                                    0.06601
time_of_borrowingSoviet:language_refAvar: Gimbatov 2006                                < 2e-16
time_of_borrowingPre-Soviet:language_refAvar: Gimbatov 2006                            < 2e-16
time_of_borrowingSoviet:language_refAkhvakh: Magomedova, Abdulayeva 2007               0.42056
time_of_borrowingPre-Soviet:language_refAkhvakh: Magomedova, Abdulayeva 2007           0.46945
time_of_borrowingSoviet:language_refAndi: Salimov 2010                                 0.00555
time_of_borrowingPre-Soviet:language_refAndi: Salimov 2010                             0.08227
time_of_borrowingSoviet:language_refBagvalal: Magomedova 2004                          2.30e-05
time_of_borrowingPre-Soviet:language_refBagvalal: Magomedova 2004                      1.03e-06
time_of_borrowingSoviet:language_refBotlikh: Alekseev, Azaev 2019                      < 2e-16
time_of_borrowingPre-Soviet:language_refBotlikh: Alekseev, Azaev 2019                  4.31e-13
time_of_borrowingSoviet:language_refBotlikh: Saidova, Abusov 2012                      1.55e-06
time_of_borrowingPre-Soviet:language_refBotlikh: Saidova, Abusov 2012                  5.16e-08
time_of_borrowingSoviet:language_refChamalal: Magomedova 1999                          0.90370
time_of_borrowingPre-Soviet:language_refChamalal: Magomedova 1999                      0.23754
time_of_borrowingSoviet:language_refGodoberi: Saidova 2006                             0.00432
time_of_borrowingPre-Soviet:language_refGodoberi: Saidova 2006                         0.04506
time_of_borrowingSoviet:language_refKarata-Tukita: Magomedova, Khalidova 2001          0.09839
time_of_borrowingPre-Soviet:language_refKarata-Tukita: Magomedova, Khalidova 2001      0.29154
time_of_borrowingSoviet:language_refTindi: Magomedova 2003                             0.42386

(Intercept)                                                                           ***
lemma_frequency_ipm                                                                   .
time_of_borrowingSoviet:language_refAvar: Gimbatov 2006                                ***
time_of_borrowingPre-Soviet:language_refAvar: Gimbatov 2006                            ***
time_of_borrowingSoviet:language_refAkhvakh: Magomedova, Abdulayeva 2007
time_of_borrowingPre-Soviet:language_refAkhvakh: Magomedova, Abdulayeva 2007
time_of_borrowingSoviet:language_refAndi: Salimov 2010                                 **
time_of_borrowingPre-Soviet:language_refAndi: Salimov 2010                             .
time_of_borrowingSoviet:language_refBagvalal: Magomedova 2004                          ***
time_of_borrowingPre-Soviet:language_refBagvalal: Magomedova 2004                      ***
time_of_borrowingSoviet:language_refBotlikh: Alekseev, Azaev 2019                      ***
time_of_borrowingPre-Soviet:language_refBotlikh: Alekseev, Azaev 2019                  ***
time_of_borrowingSoviet:language_refBotlikh: Saidova, Abusov 2012                      ***
time_of_borrowingPre-Soviet:language_refBotlikh: Saidova, Abusov 2012                  ***
time_of_borrowingSoviet:language_refChamalal: Magomedova 1999
time_of_borrowingPre-Soviet:language_refChamalal: Magomedova 1999
time_of_borrowingSoviet:language_refGodoberi: Saidova 2006                             **
time_of_borrowingPre-Soviet:language_refGodoberi: Saidova 2006                         *
time_of_borrowingSoviet:language_refKarata-Tukita: Magomedova, Khalidova 2001          .
```

```
time_of_borrowingPre-Soviet:language_refKarata-Tukita: Magomedova, Khalidova 2001
time_of_borrowingSoviet:language_refTindi: Magomedova 2003
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



Correlation matrix not shown by default, as p = 21 > 12.
Use print(x, correlation=TRUE)  or
    vcov(x)         if you need it


fit warnings:
fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

The model predictions are visualized with effect plots.

```
library(ggeffects)

df |>
    distinct(language_ref, russian_source_lexeme) |>
    count(language_ref) |>
    rename(x = language_ref,
        word_list_size = n) →
    word_list_size

fit |>
  ggpredict(terms = c("language_ref", "time_of_borrowing")) |>
  as_tibble() |>
  left_join(word_list_size) |>
  mutate(x = str_c(x, " (", word_list_size, " lemmata)"),
       x = fct_reorder(x, predicted)) |>
  ggplot(aes(predicted, x, color = group))+
  geom_linerange(aes(xmin = conf.low, xmax = conf.high), position = position_dodge(width = 0.5)) +
  geom_point(show.legend = FALSE, position = position_dodge(width = 0.5))+
  theme_minimal()+
  labs(x = "model prediction of the probability of change",
    y = NULL,
    color = NULL)+
  theme(text = element_text(size = 19),
       legend.position = "bottom")
```
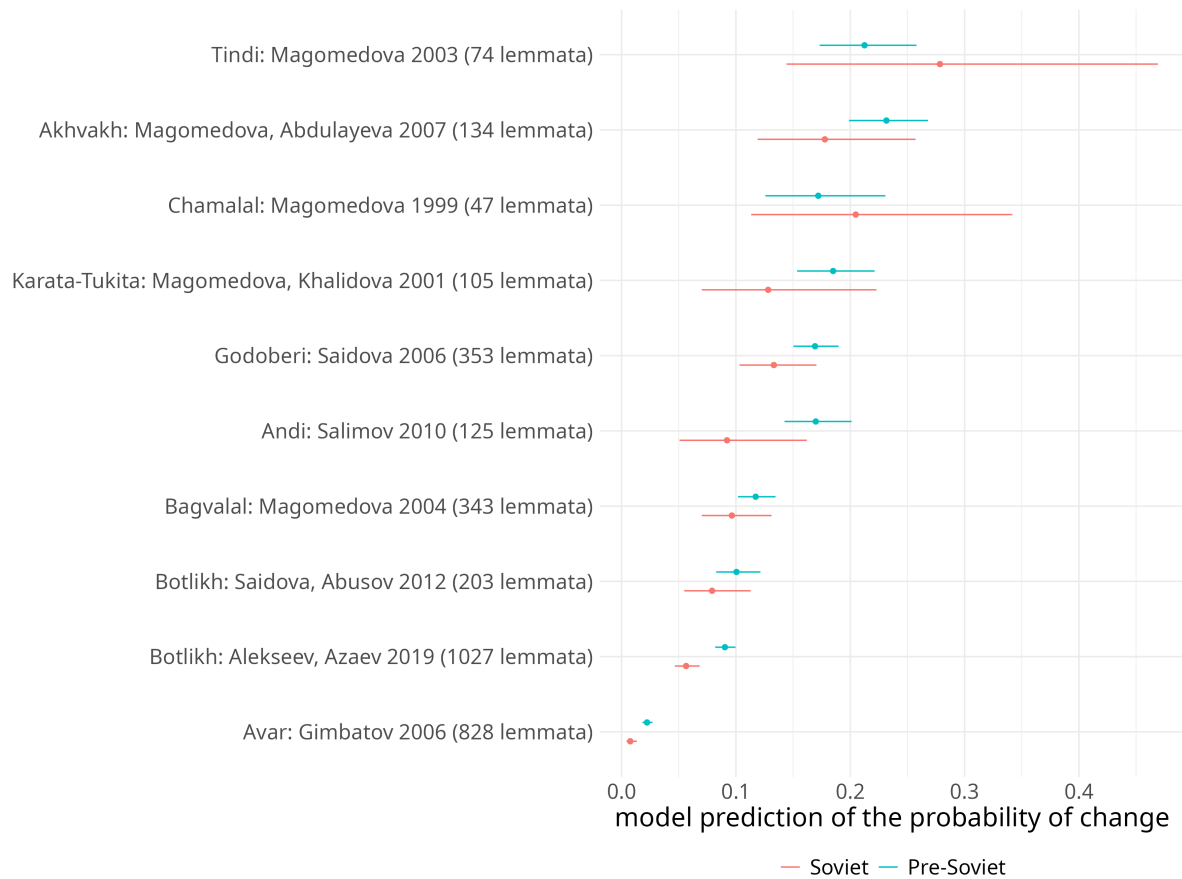
Figure 3: Probabilities of change by dictionary and approximate time of borrowing with 95% confidence intervals.

4 Packages

In the following table, we list all R packages and R version used in the project:

Table 2: The list of versions of R packages used in the project

package	version	citation
ggeffects	2.2.1	Lüdecke (2018)
inspectdf	0.0.12.1	Rushworth (2024)
lme4	1.1.37	Bates (2015)
quarto	1.4.4	Allaire (2024)
scales	1.3.0	Wickham (2023)
tidytext	0.4.2	Silge (2016)
tidyverse	2.0.0	Wickham (2019)
R	4.3.3	R Core Team (2024)

Allaire, JJ, and Christophe Dervieux. 2024. *Quarto: R Interface to 'Quarto' Markdown Publishing System.* https://CRAN.R-project.org/package=quarto.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Lüdecke, Daniel. 2018. "Ggeffects: Tidy Data Frames of Marginal Effects from Regression Models." *Journal of Open Source Software* 3 (26): 772. https://doi.org/10.21105/joss.00772.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rushworth, Alastair. 2024. *Inspectdf: Inspection, Comparison and Visualisation of Data Frames.* https://CRAN.R-project.org/package=inspectdf.

Silge, Julia, and David Robinson. 2016. "Tidytext: Text Mining and Analysis Using Tidy Data Principles in r." *JOSS* 1 (3). https://doi.org/10.21105/joss.00037.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Thomas Lin Pedersen, and Dana Seidel. 2023. *Scales: Scale Functions for Visualization.* https://CRAN.R-project.org/package=scales.