

Supplementary materials to the paper 'Adaptation of Russian borrowings in Andic languages'

12.04.2025

Table of contents

1	Data	1
2	Distribution of changes across languages	3
3	Modeling	6
4	Packages	12

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: unified dictionary translation
- russian_ipa: IPA transcription of russian word or part of the word with some modifications;

- target_ipa: IPA transcription of target language word;
- change: coding for the change;
- total: total number of units in the analysis;
- changes: number of observed changes;
- time_of_borrowing: time of borrowing approximated using Russian National Corpus.

Here is a most frequent values in the subsample of variables (variables dictionary_translation, 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(`(1.1) language` = language,
       `(1.2) reference` = reference,
       `(1.3) change` = change,
       `(1.4) total units per word` = total_units_per_word,
       `(1.5) changes per word` = changes_per_word,
       `(1.6) type of change` = type_of_change,
       `(1.7) 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))
```

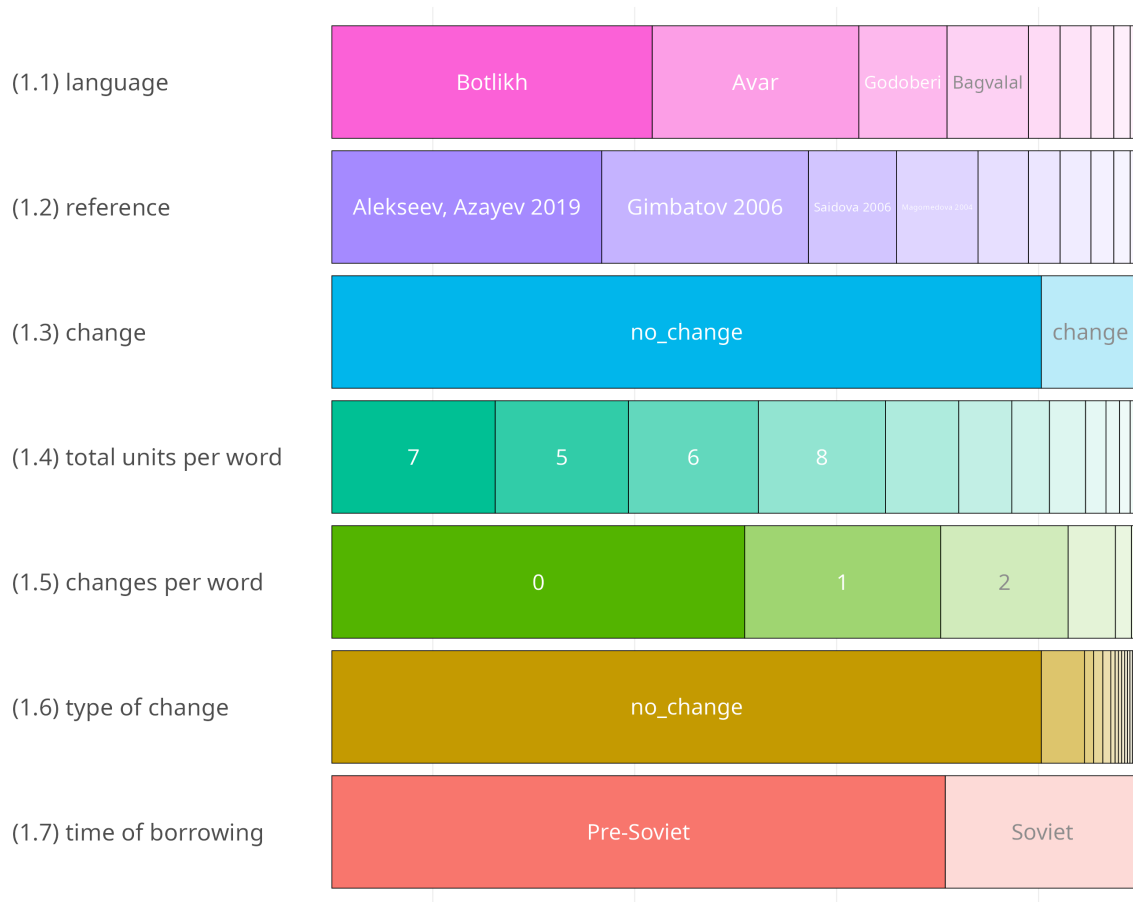


Figure 1: Frequency of values for each variable in the dataset. Gray segments are missing values.

2 Distribution of changes across languages

```
df >
  select(language_ref, type_of_change) >
  filter(type_of_change != "no_change",
         type_of_change != "other") >
  mutate(type_of_change = str_split(type_of_change, ", ")) >
  unnest(type_of_change) >
  count(language_ref, type_of_change) >
  mutate(type_of_change = tidytext::reorder_within(type_of_change, n, language_ref)) >
  ggplot(aes(n, type_of_change))+
  geom_col(show.legend = FALSE)+
```

```
facet_wrap(~language_ref, scales = "free", ncol = 2)+  
tidytext::scale_y_reordered()+  
scale_x_continuous(breaks = scales::breaks_pretty()+  
labs(x = NULL,  
      y = NULL)+  
theme_minimal()
```

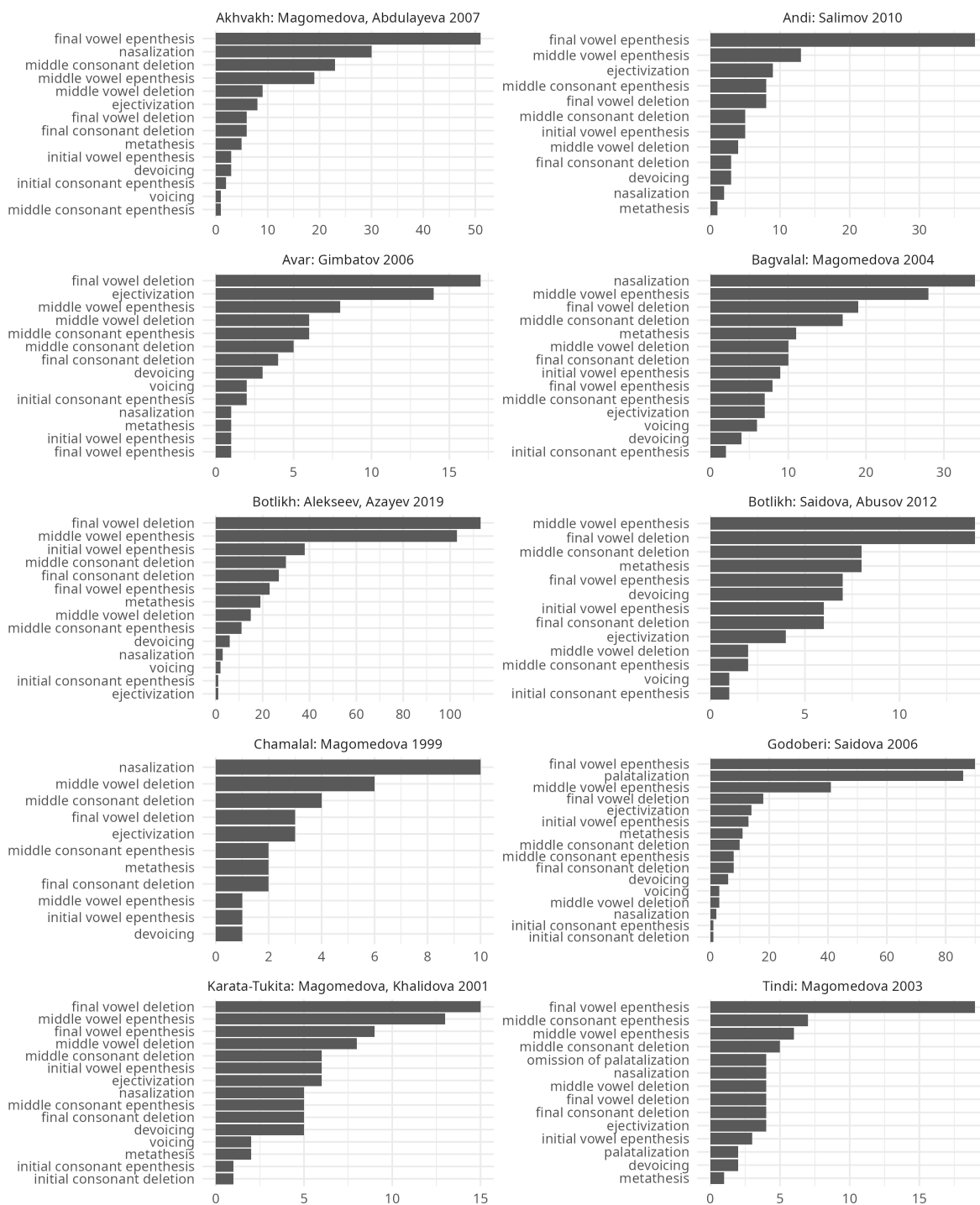


Figure 2: Frequency of each change by language

3 Modeling

We decided to create a model that predicts the average number of changes based on the resource and approximated time of borrowing. We applied a mixed effect logistic regression model. The models were generated with the R (R Core Team 2024) package `lme4` (Bates et al. 2015) with the following formula:

```
change ~ (language resource):approximated time of borrowing + (1|dictionary lemma translation)
```

The random effect of the model is unified dictionary lemma translation. We included in the model interaction of two variables: language resource and approximated time of borrowing. Since this model will compare values with some baseline Soviet borrowings from Avar dictionary (Gimbatov 2006) were used as a baseline. All language resources turn out to be statistically significantly different from Avar. Approximated time of borrowing turned out not to be statistically significant (p-value = 0.134120). Just a few interactions of language resources variable with the time of borrowing variable turn out to be statistically significant: for Andi and for 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),
         time_of_borrowing = fct_relevel(time_of_borrowing, "Soviet"),
         language_ref = fct_relevel(language_ref, "Avar: Gimbatov 2006")) |>
  lmer(change ~ time_of_borrowing*language_ref + (1|russian_source_lexeme), data = _) ->
  fit

summary(fit)
```

```

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula:
change ~ time_of_borrowing * language_ref + (1 | russian_source_lexeme)
Data:
mutate(df, change = if_else(change == "no_change", 0, 1), time_of_borrowing = fct_relevel(time_of_borrowing,
"Soviet"), language_ref = fct_relevel(language_ref, "Avar: Gimbatov 2006"))

```

REML criterion at convergence: 11134.4

Scaled residuals:

Min	1Q	Median	3Q	Max
-1.8495	-0.4589	-0.2550	-0.0416	3.3512

Random effects:

Groups	Name	Variance	Std.Dev.
russian_source_lexeme	(Intercept)	0.005455	0.07386
Residual		0.093566	0.30589

Number of obs: 21749, groups: russian_source_lexeme, 1343

Fixed effects:

	Estimate
(Intercept)	1.236e-02
time_of_borrowingPre-Soviet	1.656e-02
language_refAkhvakh: Magomedova, Abdulayeva 2007	2.415e-01
language_refAndi: Salimov 2010	1.084e-01
language_refBagvalal: Magomedova 2004	1.120e-01
language_refBotlikh: Alekseev, Azayev 2019	6.124e-02
language_refBotlikh: Saidova, Abusov 2012	8.688e-02
language_refChamalal: Magomedova 1999	2.765e-01
language_refGodoberi: Saidova 2006	1.525e-01
language_refKarata-Tukita: Magomedova, Khalidova 2001	1.675e-01
language_refTindi: Magomedova 2003	3.691e-01
time_of_borrowingPre-Soviet:language_refAkhvakh: Magomedova, Abdulayeva 2007	1.096e-02
time_of_borrowingPre-Soviet:language_refAndi: Salimov 2010	7.388e-02
time_of_borrowingPre-Soviet:language_refBagvalal: Magomedova 2004	3.784e-04
time_of_borrowingPre-Soviet:language_refBotlikh: Alekseev, Azayev 2019	1.966e-02
time_of_borrowingPre-Soviet:language_refBotlikh: Saidova, Abusov 2012	7.549e-03
time_of_borrowingPre-Soviet:language_refChamalal: Magomedova 1999	-8.966e-02
time_of_borrowingPre-Soviet:language_refGodoberi: Saidova 2006	1.782e-02
time_of_borrowingPre-Soviet:language_refKarata-Tukita: Magomedova, Khalidova 2001	3.924e-02
time_of_borrowingPre-Soviet:language_refTindi: Magomedova 2003	-1.277e-01
Std. Error	

(Intercept)	9.578e-03
time_of_borrowingPre-Soviet	1.105e-02
language_refAkhvakh: Magomedova, Abdulayeva 2007	3.091e-02
language_refAndi: Salimov 2010	3.279e-02
language_refBagvalal: Magomedova 2004	1.812e-02
language_refBotlikh: Alekseev, Azayev 2019	1.113e-02
language_refBotlikh: Saidova, Abusov 2012	1.985e-02
language_refChamalal: Magomedova 1999	4.734e-02
language_refGodoberi: Saidova 2006	1.698e-02
language_refKarata-Tukita: Magomedova, Khalidova 2001	4.184e-02
language_refTindi: Magomedova 2003	5.678e-02
time_of_borrowingPre-Soviet:language_refAkhvakh: Magomedova, Abdulayeva 2007	3.362e-02
time_of_borrowingPre-Soviet:language_refAndi: Salimov 2010	3.539e-02
time_of_borrowingPre-Soviet:language_refBagvalal: Magomedova 2004	2.033e-02
time_of_borrowingPre-Soviet:language_refBotlikh: Alekseev, Azayev 2019	1.298e-02
time_of_borrowingPre-Soviet:language_refBotlikh: Saidova, Abusov 2012	2.294e-02
time_of_borrowingPre-Soviet:language_refChamalal: Magomedova 1999	5.279e-02
time_of_borrowingPre-Soviet:language_refGodoberi: Saidova 2006	1.924e-02
time_of_borrowingPre-Soviet:language_refKarata-Tukita: Magomedova, Khalidova 2001	4.438e-02
time_of_borrowingPre-Soviet:language_refTindi: Magomedova 2003	5.928e-02
	df
(Intercept)	5.411e+03
time_of_borrowingPre-Soviet	5.519e+03
language_refAkhvakh: Magomedova, Abdulayeva 2007	1.834e+04
language_refAndi: Salimov 2010	1.696e+04
language_refBagvalal: Magomedova 2004	1.869e+04
language_refBotlikh: Alekseev, Azayev 2019	1.880e+04
language_refBotlikh: Saidova, Abusov 2012	1.768e+04
language_refChamalal: Magomedova 1999	2.027e+04
language_refGodoberi: Saidova 2006	1.785e+04
language_refKarata-Tukita: Magomedova, Khalidova 2001	2.060e+04
language_refTindi: Magomedova 2003	2.002e+04
time_of_borrowingPre-Soviet:language_refAkhvakh: Magomedova, Abdulayeva 2007	1.847e+04
time_of_borrowingPre-Soviet:language_refAndi: Salimov 2010	1.727e+04
time_of_borrowingPre-Soviet:language_refBagvalal: Magomedova 2004	1.896e+04
time_of_borrowingPre-Soviet:language_refBotlikh: Alekseev, Azayev 2019	1.919e+04
time_of_borrowingPre-Soviet:language_refBotlikh: Saidova, Abusov 2012	1.814e+04
time_of_borrowingPre-Soviet:language_refChamalal: Magomedova 1999	2.028e+04
time_of_borrowingPre-Soviet:language_refGodoberi: Saidova 2006	1.836e+04
time_of_borrowingPre-Soviet:language_refKarata-Tukita: Magomedova, Khalidova 2001	2.054e+04
time_of_borrowingPre-Soviet:language_refTindi: Magomedova 2003	1.998e+04
	t value
(Intercept)	1.290

time_of_borrowingPre-Soviet	1.498
language_refAkhvakh: Magomedova, Abdulayeva 2007	7.814
language_refAndi: Salimov 2010	3.305
language_refBagvalal: Magomedova 2004	6.177
language_refBotlikh: Alekseev, Azayev 2019	5.502
language_refBotlikh: Saidova, Abusov 2012	4.377
language_refChamalal: Magomedova 1999	5.841
language_refGodoberi: Saidova 2006	8.982
language_refKarata-Tukita: Magomedova, Khalidova 2001	4.003
language_refTindi: Magomedova 2003	6.501
time_of_borrowingPre-Soviet:language_refAkhvakh: Magomedova, Abdulayeva 2007	0.326
time_of_borrowingPre-Soviet:language_refAndi: Salimov 2010	2.088
time_of_borrowingPre-Soviet:language_refBagvalal: Magomedova 2004	0.019
time_of_borrowingPre-Soviet:language_refBotlikh: Alekseev, Azayev 2019	1.515
time_of_borrowingPre-Soviet:language_refBotlikh: Saidova, Abusov 2012	0.329
time_of_borrowingPre-Soviet:language_refChamalal: Magomedova 1999	-1.698
time_of_borrowingPre-Soviet:language_refGodoberi: Saidova 2006	0.926
time_of_borrowingPre-Soviet:language_refKarata-Tukita: Magomedova, Khalidova 2001	0.884
time_of_borrowingPre-Soviet:language_refTindi: Magomedova 2003	-2.154
	Pr(> t)
(Intercept)	0.196933
time_of_borrowingPre-Soviet	0.134120
language_refAkhvakh: Magomedova, Abdulayeva 2007	5.84e-15
language_refAndi: Salimov 2010	0.000951
language_refBagvalal: Magomedova 2004	6.65e-10
language_refBotlikh: Alekseev, Azayev 2019	3.80e-08
language_refBotlikh: Saidova, Abusov 2012	1.21e-05
language_refChamalal: Magomedova 1999	5.26e-09
language_refGodoberi: Saidova 2006	< 2e-16
language_refKarata-Tukita: Magomedova, Khalidova 2001	6.28e-05
language_refTindi: Magomedova 2003	8.19e-11
time_of_borrowingPre-Soviet:language_refAkhvakh: Magomedova, Abdulayeva 2007	0.744532
time_of_borrowingPre-Soviet:language_refAndi: Salimov 2010	0.036844
time_of_borrowingPre-Soviet:language_refBagvalal: Magomedova 2004	0.985153
time_of_borrowingPre-Soviet:language_refBotlikh: Alekseev, Azayev 2019	0.129883
time_of_borrowingPre-Soviet:language_refBotlikh: Saidova, Abusov 2012	0.742083
time_of_borrowingPre-Soviet:language_refChamalal: Magomedova 1999	0.089451
time_of_borrowingPre-Soviet:language_refGodoberi: Saidova 2006	0.354412
time_of_borrowingPre-Soviet:language_refKarata-Tukita: Magomedova, Khalidova 2001	0.376630
time_of_borrowingPre-Soviet:language_refTindi: Magomedova 2003	0.031274
(Intercept)	
time_of_borrowingPre-Soviet	

```

language_refAkhvakh: Magomedova, Abdulayeva 2007      ***
language_refAndi: Salimov 2010                        ***
language_refBagvalal: Magomedova 2004                 ***
language_refBotlikh: Alekseev, Azayev 2019            ***
language_refBotlikh: Saidova, Abusov 2012             ***
language_refChamalal: Magomedova 1999                 ***
language_refGodoberi: Saidova 2006                   ***
language_refKarata-Tukita: Magomedova, Khalidova 2001 ***
language_refTindi: Magomedova 2003                   ***
time_of_borrowingPre-Soviet:language_refAkhvakh: Magomedova, Abdulayeva 2007
time_of_borrowingPre-Soviet:language_refAndi: Salimov 2010      *
time_of_borrowingPre-Soviet:language_refBagvalal: Magomedova 2004
time_of_borrowingPre-Soviet:language_refBotlikh: Alekseev, Azayev 2019
time_of_borrowingPre-Soviet:language_refBotlikh: Saidova, Abusov 2012
time_of_borrowingPre-Soviet:language_refChamalal: Magomedova 1999 .
time_of_borrowingPre-Soviet:language_refGodoberi: Saidova 2006
time_of_borrowingPre-Soviet:language_refKarata-Tukita: Magomedova, Khalidova 2001
time_of_borrowingPre-Soviet:language_refTindi: Magomedova 2003  *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Correlation matrix not shown by default, as $p = 20 > 12$.

Use `print(x, correlation=TRUE)` or
`vcov(x)` if you need it

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 = 16),
      legend.position = "bottom")

```

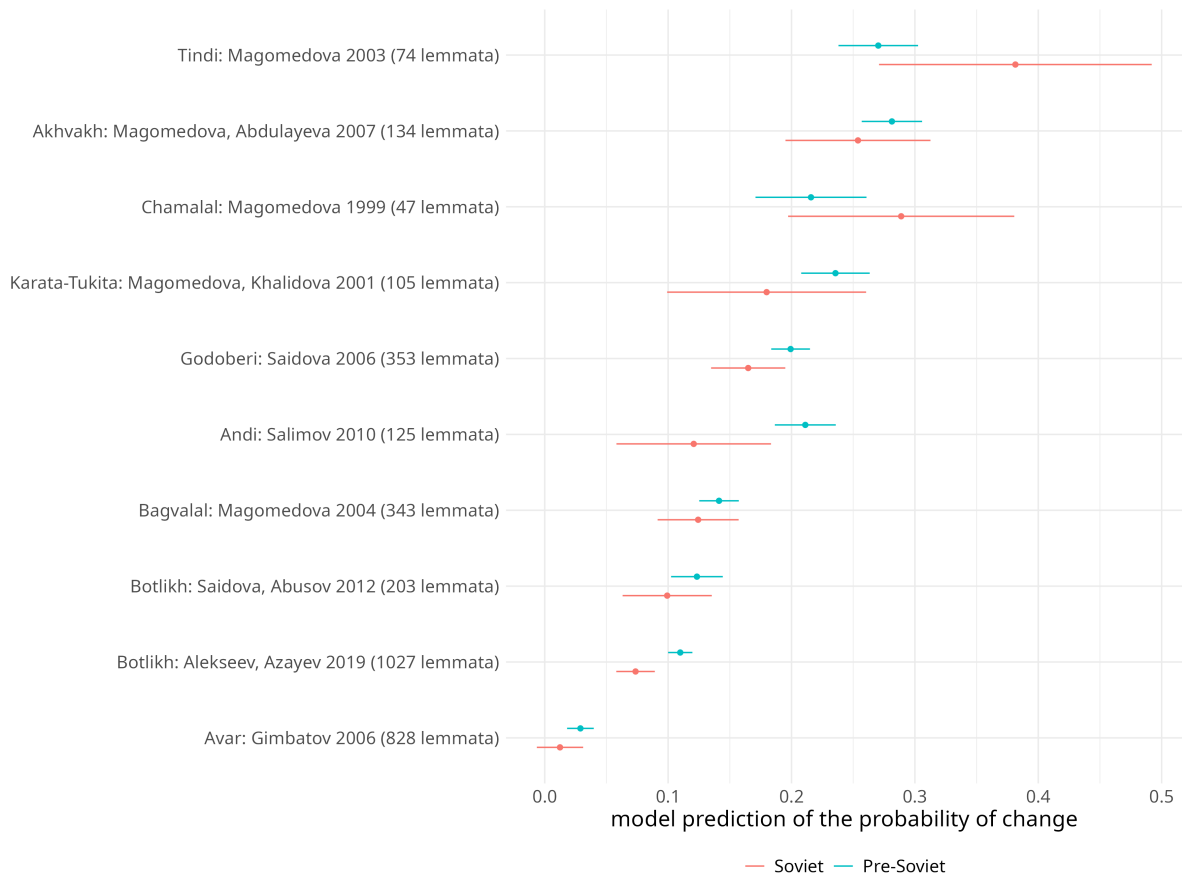


Figure 3: Probabilities of change by language, source and approximated time of borrowing

4 Packages

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

Table 1: 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 Golemund, 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>.