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

11.04.2025

Table of contents

1 Data

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

```r
library(tidyverse)

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

The dataset consists of 21751 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 ommited due to the huge amount of values).

```
library(inspectdf)

df ▷
  mutate_all(as.factor) ▷
  select(language, reference, change, total, changes, time_of_borrowing) ▷
  rename(`(1.1) language` = language,
         `(1.2) reference` = reference,
         `(1.3) change` = change,
         `(1.4) total units per word` = total,
         `(1.5) changes per word` = changes,
         `(1.6) 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))
```
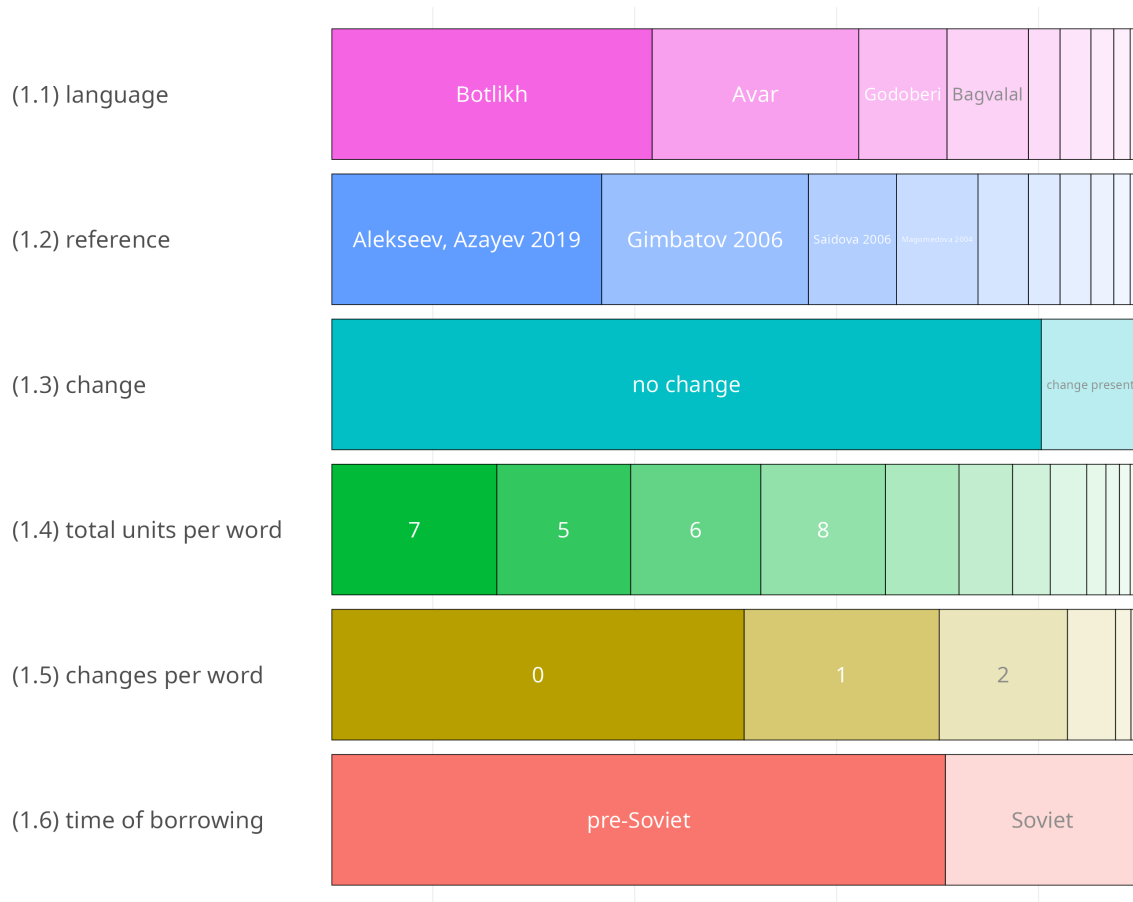
Figure 1: Frequency of values for each variable in the dataset. Gray segments are missing values.

2 Distribution of changes across languages

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 = 0.112855).

```r
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
```

```r
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|dictionary_translation), 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 | dictionary_translation)
   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: 11172.2

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.8506 -0.4583 -0.2549 -0.0438  3.3472
```

```
Random effects:
 Groups                    Name        Variance Std.Dev.
 dictionary_translation (Intercept) 0.00551  0.07423
 Residual                               0.09370  0.30611
Number of obs: 21751, groups:  dictionary_translation, 1343

Fixed effects:
                                                                                    Estimate

(Intercept)                                                                         1.238e-02
time_of_borrowingpre-Soviet                                                         1.755e-02
language_refAkhvakh: Magomedova, Abdulayeva 2007                                    2.311e-01
language_refAndi: Salimov 2010                                                      1.068e-01
language_refBagvalal: Magomedova 2004                                               1.111e-01
language_refBotlikh: Alekseev, Azayev 2019                                          6.043e-02
language_refBotlikh: Saidova, Abusov 2012                                           8.686e-02
language_refChamalal: Magomedova 1999                                               2.637e-01
language_refGodoberi: Saidova 2006                                                  1.496e-01
language_refKarata-Tukita: Magomedova, Khalidova 2001                               1.502e-01
language_refTindi: Magomedova 2003                                                  3.390e-01
time_of_borrowingpre-Soviet:language_refAkhvakh: Magomedova, Abdulayeva 2007        2.049e-02
time_of_borrowingpre-Soviet:language_refAndi: Salimov 2010                          7.587e-02
time_of_borrowingpre-Soviet:language_refBagvalal: Magomedova 2004                  -1.890e-04
time_of_borrowingpre-Soviet:language_refBotlikh: Alekseev, Azayev 2019              1.979e-02
time_of_borrowingpre-Soviet:language_refBotlikh: Saidova, Abusov 2012               5.920e-03
time_of_borrowingpre-Soviet:language_refChamalal: Magomedova 1999                  -7.649e-02
time_of_borrowingpre-Soviet:language_refGodoberi: Saidova 2006                      2.033e-02
time_of_borrowingpre-Soviet:language_refKarata-Tukita: Magomedova, Khalidova 2001   5.249e-02
time_of_borrowingpre-Soviet:language_refTindi: Magomedova 2003                     -9.874e-02
                                                                                    Std. Error
(Intercept)                                                                         9.595e-03
time_of_borrowingpre-Soviet                                                         1.107e-02
language_refAkhvakh: Magomedova, Abdulayeva 2007                                    3.094e-02
language_refAndi: Salimov 2010                                                      3.283e-02
language_refBagvalal: Magomedova 2004                                               1.815e-02
language_refBotlikh: Alekseev, Azayev 2019                                          1.114e-02
language_refBotlikh: Saidova, Abusov 2012                                           1.988e-02
language_refChamalal: Magomedova 1999                                               4.739e-02
language_refGodoberi: Saidova 2006                                                  1.701e-02
language_refKarata-Tukita: Magomedova, Khalidova 2001                               4.188e-02
language_refTindi: Magomedova 2003                                                  5.684e-02
time_of_borrowingpre-Soviet:language_refAkhvakh: Magomedova, Abdulayeva 2007        3.366e-02
time_of_borrowingpre-Soviet:language_refAndi: Salimov 2010                          3.542e-02
time_of_borrowingpre-Soviet:language_refBagvalal: Magomedova 2004                   2.036e-02
```

```
time_of_borrowingpre-Soviet:language_refBotlikh: Alekseev, Azayev 2019                      1.300e-02
time_of_borrowingpre-Soviet:language_refBotlikh: Saidova, Abusov 2012                       2.297e-02
time_of_borrowingpre-Soviet:language_refChamalal: Magomedova 1999                           5.285e-02
time_of_borrowingpre-Soviet:language_refGodoberi: Saidova 2006                              1.927e-02
time_of_borrowingpre-Soviet:language_refKarata-Tukita: Magomedova, Khalidova 2001           4.443e-02
time_of_borrowingpre-Soviet:language_refTindi: Magomedova 2003                              5.934e-02
                                                                                                  df
(Intercept)                                                                                5.407e+03
time_of_borrowingpre-Soviet                                                                5.514e+03
language_refAkhvakh: Magomedova, Abdulayeva 2007                                           1.837e+04
language_refAndi: Salimov 2010                                                             1.700e+04
language_refBagvalal: Magomedova 2004                                                      1.870e+04
language_refBotlikh: Alekseev, Azayev 2019                                                 1.883e+04
language_refBotlikh: Saidova, Abusov 2012                                                  1.765e+04
language_refChamalal: Magomedova 1999                                                      2.030e+04
language_refGodoberi: Saidova 2006                                                         1.782e+04
language_refKarata-Tukita: Magomedova, Khalidova 2001                                      2.062e+04
language_refTindi: Magomedova 2003                                                         2.006e+04
time_of_borrowingpre-Soviet:language_refAkhvakh: Magomedova, Abdulayeva 2007               1.849e+04
time_of_borrowingpre-Soviet:language_refAndi: Salimov 2010                                 1.732e+04
time_of_borrowingpre-Soviet:language_refBagvalal: Magomedova 2004                          1.896e+04
time_of_borrowingpre-Soviet:language_refBotlikh: Alekseev, Azayev 2019                     1.922e+04
time_of_borrowingpre-Soviet:language_refBotlikh: Saidova, Abusov 2012                      1.814e+04
time_of_borrowingpre-Soviet:language_refChamalal: Magomedova 1999                          2.030e+04
time_of_borrowingpre-Soviet:language_refGodoberi: Saidova 2006                             1.834e+04
time_of_borrowingpre-Soviet:language_refKarata-Tukita: Magomedova, Khalidova 2001          2.056e+04
time_of_borrowingpre-Soviet:language_refTindi: Magomedova 2003                             2.001e+04
                                                                                             t value
(Intercept)                                                                                    1.290
time_of_borrowingpre-Soviet                                                                    1.586
language_refAkhvakh: Magomedova, Abdulayeva 2007                                               7.469
language_refAndi: Salimov 2010                                                                 3.252
language_refBagvalal: Magomedova 2004                                                          6.121
language_refBotlikh: Alekseev, Azayev 2019                                                     5.424
language_refBotlikh: Saidova, Abusov 2012                                                      4.370
language_refChamalal: Magomedova 1999                                                          5.564
language_refGodoberi: Saidova 2006                                                             8.794
language_refKarata-Tukita: Magomedova, Khalidova 2001                                          3.586
language_refTindi: Magomedova 2003                                                             5.965
time_of_borrowingpre-Soviet:language_refAkhvakh: Magomedova, Abdulayeva 2007                   0.609
time_of_borrowingpre-Soviet:language_refAndi: Salimov 2010                                     2.142
time_of_borrowingpre-Soviet:language_refBagvalal: Magomedova 2004                             -0.009
time_of_borrowingpre-Soviet:language_refBotlikh: Alekseev, Azayev 2019                         1.523
```

```
time_of_borrowingpre-Soviet:language_refBotlikh: Saidova, Abusov 2012           0.258
time_of_borrowingpre-Soviet:language_refChamalal: Magomedova 1999                -1.447
time_of_borrowingpre-Soviet:language_refGodoberi: Saidova 2006                    1.055
time_of_borrowingpre-Soviet:language_refKarata-Tukita: Magomedova, Khalidova 2001 1.181
time_of_borrowingpre-Soviet:language_refTindi: Magomedova 2003                   -1.664
                                                                                 Pr(>|t|)
(Intercept)                                                                      0.197184
time_of_borrowingpre-Soviet                                                      0.112855
language_refAkhvakh: Magomedova, Abdulayeva 2007                                 8.42e-14
language_refAndi: Salimov 2010                                                   0.001148
language_refBagvalal: Magomedova 2004                                            9.47e-10
language_refBotlikh: Alekseev, Azayev 2019                                       5.90e-08
language_refBotlikh: Saidova, Abusov 2012                                        1.25e-05
language_refChamalal: Magomedova 1999                                            2.68e-08
language_refGodoberi: Saidova 2006                                                < 2e-16
language_refKarata-Tukita: Magomedova, Khalidova 2001                            0.000337
language_refTindi: Magomedova 2003                                               2.48e-09
time_of_borrowingpre-Soviet:language_refAkhvakh: Magomedova, Abdulayeva 2007     0.542658
time_of_borrowingpre-Soviet:language_refAndi: Salimov 2010                       0.032237
time_of_borrowingpre-Soviet:language_refBagvalal: Magomedova 2004                0.992590
time_of_borrowingpre-Soviet:language_refBotlikh: Alekseev, Azayev 2019           0.127848
time_of_borrowingpre-Soviet:language_refBotlikh: Saidova, Abusov 2012            0.796610
time_of_borrowingpre-Soviet:language_refChamalal: Magomedova 1999                0.147807
time_of_borrowingpre-Soviet:language_refGodoberi: Saidova 2006                   0.291322
time_of_borrowingpre-Soviet:language_refKarata-Tukita: Magomedova, Khalidova 2001 0.237452
time_of_borrowingpre-Soviet:language_refTindi: Magomedova 2003                   0.096143

(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, dictionary_translation) ▷
    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")
```
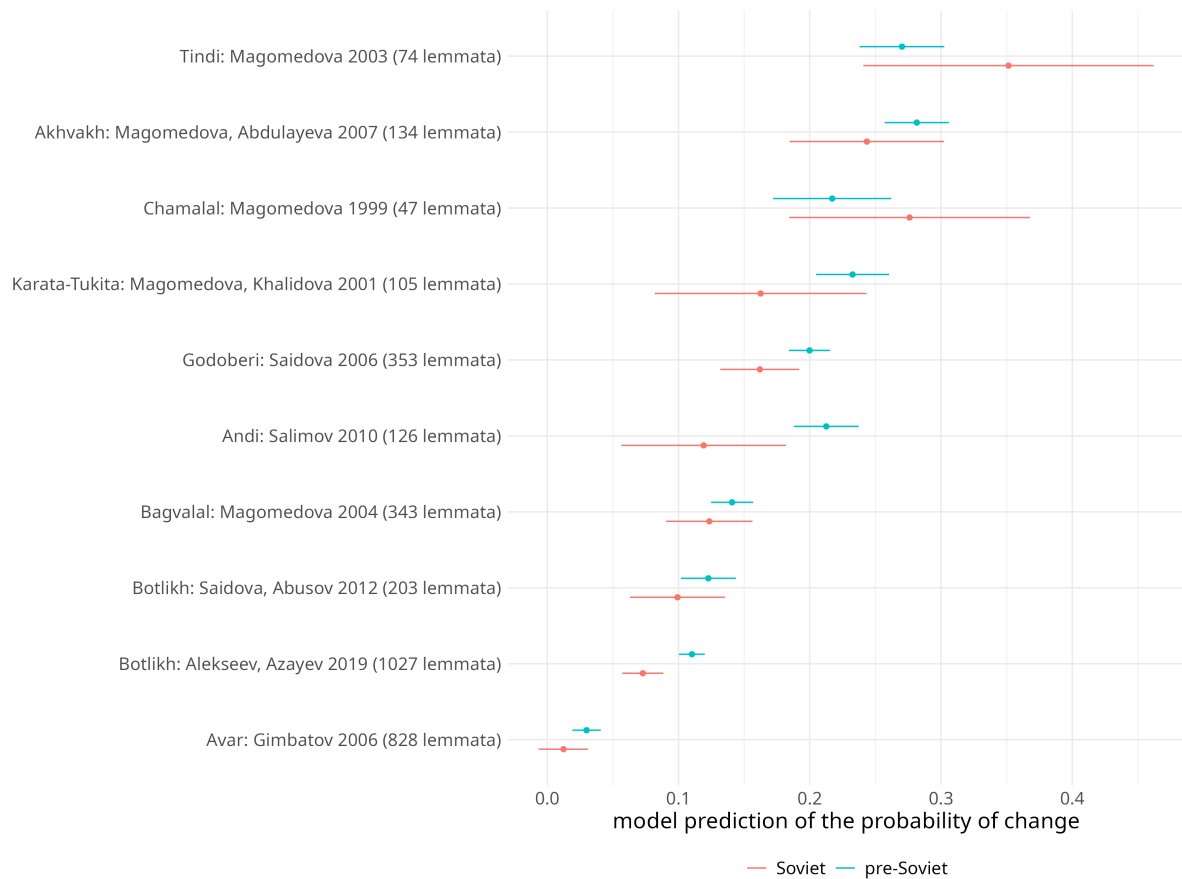
Figure 2: Probabilities of change by language, source and approximated time of borrowing

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)
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.

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.