Precursors of the project
OOOOOOOO

The DiaL2 project
OOO

The DiaL2 results
OOOOOOOOOOOOOOOO

The DiaL2 sideproject
OO

Future plans
O

References

The DiaL2 project: pipeline, results, news and future work

George Moroz   Olga Gich   Anna Grishanova   Natalia Koshelyuk
Chiara Naccarato   Anna Panova   Anastasia Yakovleva
Svetlana Zemicheva

17.09.2024

Precursors of the project

Precursors of the project

Nina Dobrushina

Michael Daniel

- Multiple sociolinguistic expeditions to Daghestan
- Several dialect expeditions to Ustya

Precursors of the project

Nina Dobrushina

Michael Daniel

- Multiple sociolinguistic expeditions to Daghestan
- Several dialect expeditions to Ustya
- Online corpora available for everyone:
  - Corpus of Russian spoken in Daghestan
  - Ustja River Basin Corpus

Precursors of the project

Nina Dobrushina

Michael Daniel

- Multiple sociolinguistic expeditions to Daghestan
- Several dialect expeditions to Ustya
- Online corpora available for everyone:
  - Corpus of Russian spoken in Daghestan
  - Ustja River Basin Corpus
  - ... and other bilingual and dialect corpora

Resources of the Linguistic Convergence Laboratory

- https://lingconlab.ru/
- 24 dialectal corpora
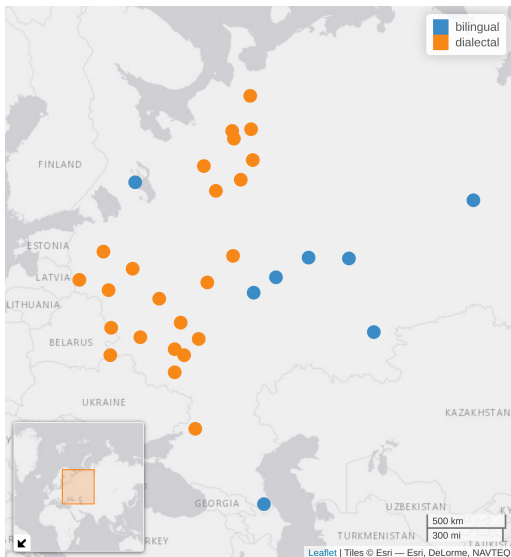- 8 bilingual corpora

Dialectal Corpora

Corpus of the Russian dialect spoken in Khislavichi district 260,793 tok.

Ustja River Basin Corpus 959,782 tok.

Corpus of the Russian dialect spoken in the villages of the Middle Pyoza 79,566 tok.

Corpus of the Russian dialect spoken in Nekhochi 88,965 tok.

Corpus of the Russian dialect spoken in the village Veegora 91,514 tok.

Corpus of Russian spoken in Zvenigorod 68,324 tok.

Luzhnikovo Corpus 68,666 tok.

Corpus of Opochetsky dialects 68,741 tok.

Upper Pinega and Vyya Corpus 70,803 tok.

Corpus of the Russian dialect spoken in the villages of the Middle Pinega 43,270 tok.

Corpus of the Russian dialect spoken in the Mikhaylov area 47,579 tok.

Corpus of the Russian dialect spoken in the villages of the Middle Northern Dvina 68,010 tok.

Corpus of Spiridonova Buda dialect 70,565 tok.

Svishni and Trostnoe Corpus 24,414 tok.

Corpus of the Russian dialect spoken in Popovka 36,617 tok.

Corpus of the Russian dialect spoken in Tserkovnoe 39,469 tok.

Corpus of the Russian dialect spoken in Keba 54,535 tok.

Corpus of the Russian dialect spoken in the villages of the Don river 69,098 tok.

Corpus of the Russian dialect spoken in Manturovo 113,837 tok.

Corpus of Rogovatka dialect 100,047 tok.

Corpus of Shetnevo and Makeevo dialect 95,335 tok.

Corpus of Lukh and Teza river basins dialects 146,350 tok.

Corpus of the Russian dialect spoken in the village Malinino 138,943 tok.

Corpus of the Russian dialect spoken in Ilmen Lake district 134,207 tok.

Bilingual Corpora

Corpus of Russian spoken in Daghestan 376,717 tok.	Khanty Russian Corpus 40,225 tok.
	Corpus of Russian spoken in Chuvashia 46,307 tok. / Corpus of Russian spoken by the Roma 41,767 tok.
Corpus of Karelian Russian 578,646 tok.	Corpus of Russian spoken in Mari El 69,109 tok.
	Corpus of Russian spoken in Bashkortostan 93,127 tok.
	Corpus of Russian spoken by the Besermans 97,216 tok.

Precursors of the project · ○○○○○●○○
The DiaL2 project · ○○○
The DiaL2 results · ○○○○○○○○○○○○○○○○○○
The DiaL2 sideproject · ○○
Future plans · ○
References

Bilingual and Dialectal Corpora

Created with `lingtypology` [Moroz, 2017]

Can we analyze variation of linguistic features across all corpora?

Can we analyze variation of linguistic features across all corpora?
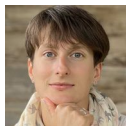
What are the factors that influence variation?

Can we analyze variation of linguistic features across all corpora?

What are the factors that influence variation?

Can we find different variation patterns?

Previous publications

- Daghestanian Russian [Daniel et al., 2010, Panova and Philippova, 2021]
- Russian of Erzya speakers [Shagal, 2016]
- Russian of Kazakh speakers [Rakhilina and Kazkenova, 2018]
- Russian of Nganasan speakers [Khomchenkova, 2020]
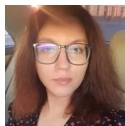- Russian of Nanai and Ulcha speakers [Stoynova, 2021]

Precursors of the project
00000000

The DiaL2 project
●00

The DiaL2 results
0000000000000000

The DiaL2 sideproject
00

Future plans
0

References

The DiaL2 project

The DiaL2 team

Maria Ermolova

Anna Grishanova
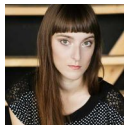

Natalia Koshelyuk

George Moroz

Chiara Naccarato
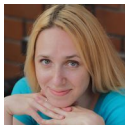
Anastasia Yakovleva

Svetlana Zemicheva

The DiaL2 pipeline

The DiaL2 results

Some results

- Non-standard numeral constructions in L2 Russian
- Propositional Drop
- Propositional Drop in Chuvash
- Dialect Genitive Plural Forms of Masculine and Neuter Nouns in Numeral Constructions
- Negative Existential Constructions

Non-standard numeral constructions in L2 Russian

Chiara Naccarato

George Moroz

Non-standard numeral constructions in L2 Russian

- Variation in numeral constructions (NCs) in bilingual corpora
  - e.g. dva brat vs. dva brata
- Previous research on other L2 Russian varieties
  - Stoynova (2021) on Nanai and Ulcha Russian: evidence for pattern borrowing
- Also mentioned by
  - Shagal (2016: 369-370) for Erzya Russian
  - Rakhilina & Kazkenova (2018: 610) for Kazakh Russian

Research questions

- Does the amount of variation in NCs differ across corpora and/or among speakers of the same variety?
- Can variation in NCs be explained in terms of contact influence?
- Do other factors promote or hinder variation in NCs?

Precursors of the project
The DiaL2 project
The DiaL2 results
The DiaL2 sideproject
Future plans
References

The database and parameters of data annotation

4,144 observations

(1.1) corpora: 7	Karelian · Daghestan · Mari
(1.2) speakers: 181	
(1.3) gender	f · m
(1.4) year of birth	
(1.5) L1: 21	Karelian · Mari
(1.6) L1 family	Uralic · East Caucasian · Turkic
(1.7) education	lower · higher
(1.8) standardness of the speaker	fully_standard · non_standard
(2.1) marking	standard
(2.2) numeral	dva · tri
(2.3) noun token	let
(2.4) noun lemma	god
(2.5) numeral type	other · paucal
(2.6) noun type	
(2.7) ambiguous	no · yes
(2.8) dice coefficient	9.224

Precursors of the project ○○○○○○○○

The DiaL2 project ○○○

The DiaL2 results ○○○○○○●○○○○○○○○

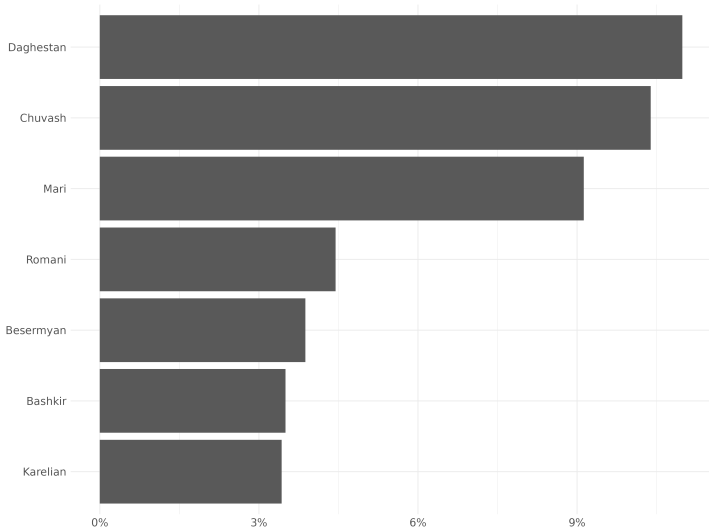The DiaL2 sideproject ○○

Future plans ○
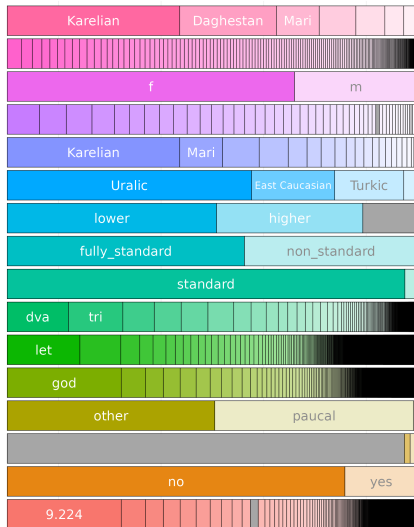
References

# Fully standard (71.3%) vs. non-standard speakers (28.7%)
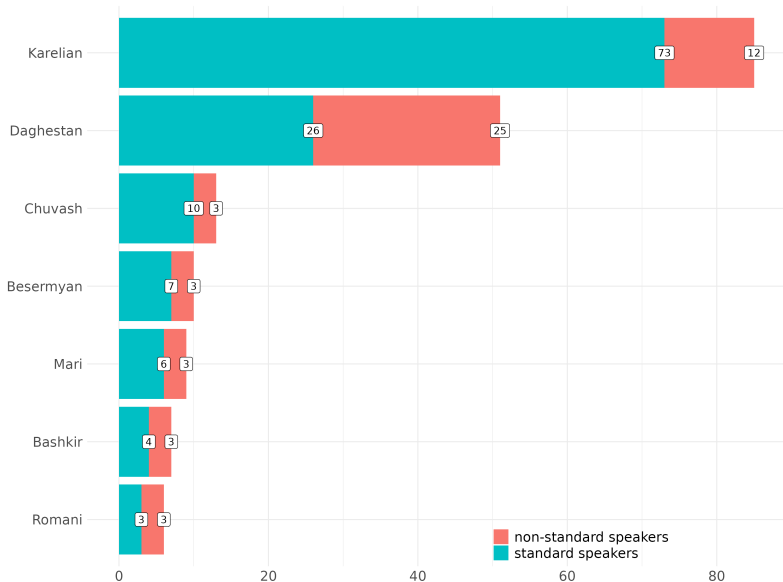
# Proportion of non-standard occurrences per corpus

1,748 observations

Distribution of n-std forms with different types of numerals

- NOM instead of GEN is frequent both with paucals and other numerals
- n-std GEN is attested sporadically
- other case forms are even less frequent
- only ~45% of n-std expressions could in principle be explained by L1 pattern borrowing
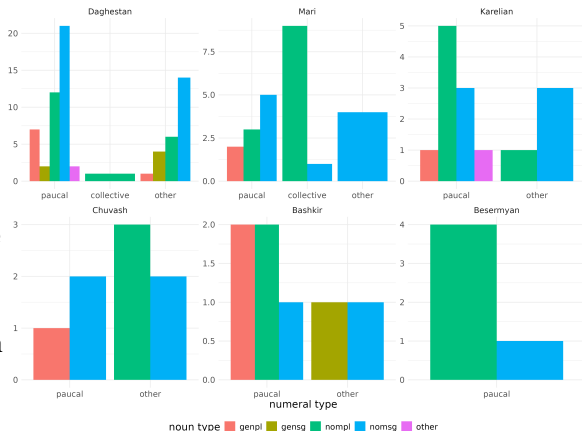
Statistical modelling
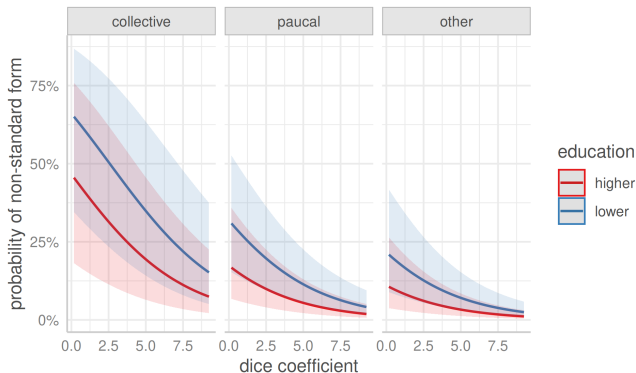
- Logistic regression: `standardness ~ Dice coefficient + year of birth + education + numeral type + gender + (1|L1 family/speaker id)`
- Conditional importance of the variables in our model (generalized R squared): collocationality (Dice coefficient) > education > year of birth > numeral type > gender

Conclusions

- Variation in NCs is attested in all L2 corpora, but not to the same extent in each of them
- Daghestanian Russian as a more uniform variety, probably due to a lower pervasiveness of Russian in every-day life, especially in the more isolated communities of the highlands
- The variables that turned out to be statistically significant are all logically related to L2 proficiency and exposure to the input, but there is no robust evidence for a contact explanation
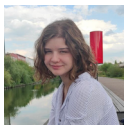
Propositional Drop
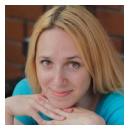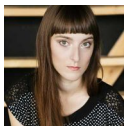

Anastasia Yakovleva
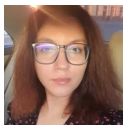
Natalia Koshelyuk

George Moroz

Propositional Drop in Chuvash

Anna Grishanova

Dialect Genitive Plural Forms in Numeral Constructions

Svetlana Zemicheva

Chiara Naccarato

George Moroz

# Negative Existential Constructions

Chiara Naccarato

George Moroz

Negative Existential Constructions

- Existential negation = negation strategies used in existential sentences of the type *there is/are no* X (*somewhere*), in which the subject is typically non-referential
- We use the terms "existential negation" and "negative existential constructions" (NECs) in a wider sense to include constructions that are sometimes referred to as "locative negation" (*X is/are not in some place*, in which X is a definite subject) and "possessive negation" (*Y does/do not have X*); cf. [Veselinova, 2013, 110–111]
- All of them predicate absolute absence rather than relative absence, and Russian employs one and the same strategy in all three cases, which is different from the strategy employed in standard negation, i.e. negation of overt verb predicates

The DiaL2 sideproject

The DiaL2 sideproject
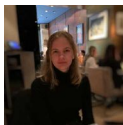
Anna Panova

Olga Gich

George Moroz

Future plans

References I

M. Daniel, N. Dobrushina, and S. Knyazev. Highlanders' Russian: Case study in bilingualism and language interference in Central Daghestan. *Slavica Helsingiensia*, 40:65–93, 2010.

I. Khomchenkova. Contact-induced features in the Russian speech of Nganasans. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics*, 11(2):13–37, 2020.

George Moroz. *lingtypology: easy mapping for Linguistic Typology*, 2017. URL https://CRAN.R-project.org/package=lingtypology.

A. Panova and T. Philippova. When a cross-linguistic tendency marries incomplete acquisition: Preposition drop in Russian spoken in Daghestan. *International Journal of Bilingualism*, 25(3):640–667, 2021.

References II

E. V. Rakhilina and A. K. Kazkenova. Zametki o russkom čisle [Notes on Russian number]. *Russian Journal of Linguistics*, 22(3):605–627, 2018.

K. Shagal. Contact-induced grammatical phenomena in the Russian of Erzya Speakers. In *Mordvin languages in the field*, pages 363–377. University of Helsinki, 2016.

N. Stoynova. Nestandartnye količestvennye konstrukcii v russkoj reči nositelej nanajskogo i ul'čskogo jazykov [Non-standard syntax of numerals in the Russian of Nanai and Ulcha speakers]. *Russian Linguistics*, 45(3):305–334, 2021.

L. Veselinova. Negative existentials: A cross-linguistic study. *Rivista di linguistica*, 25(1):107–145, 2013.