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

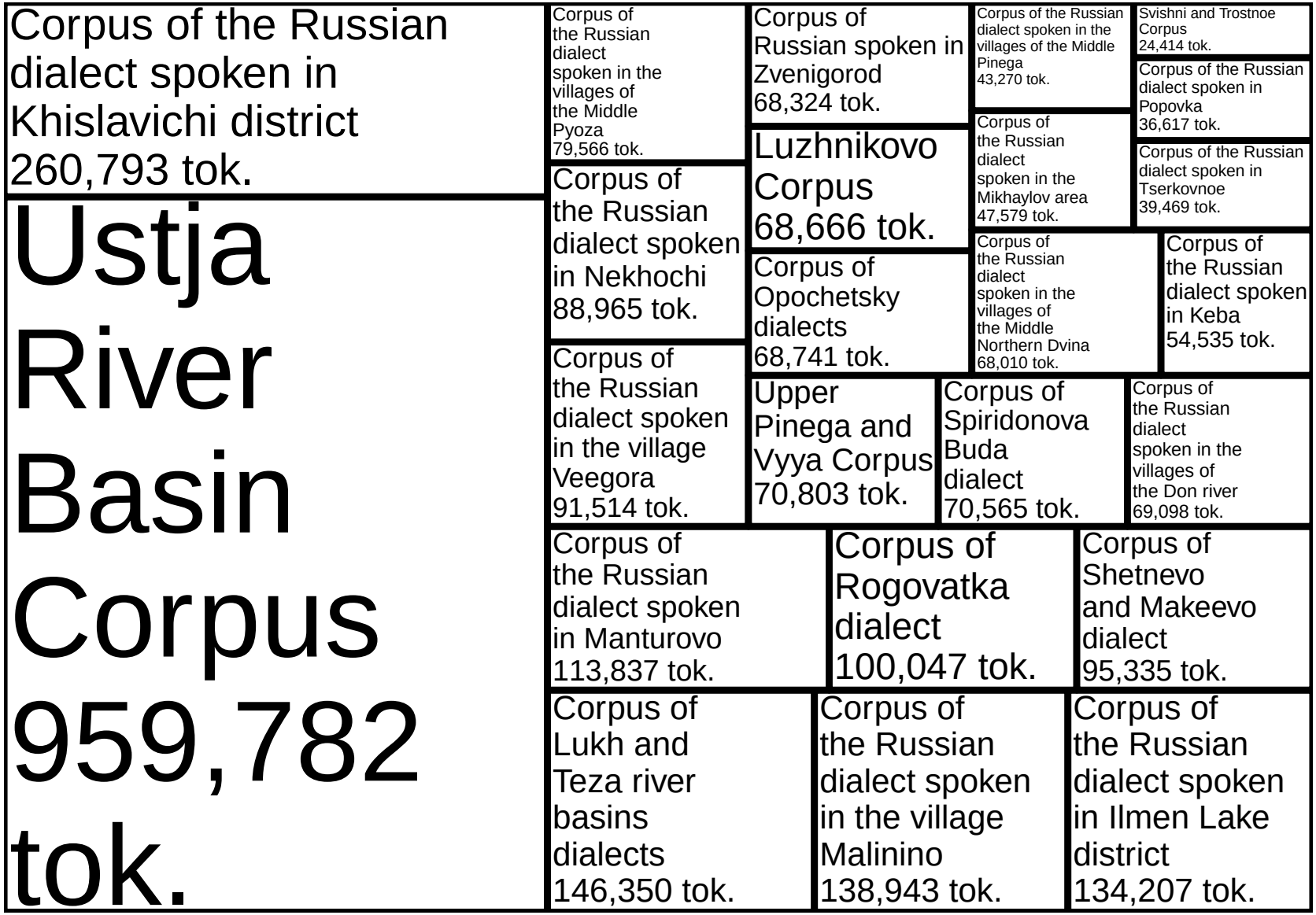
Michael Daniel

- 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

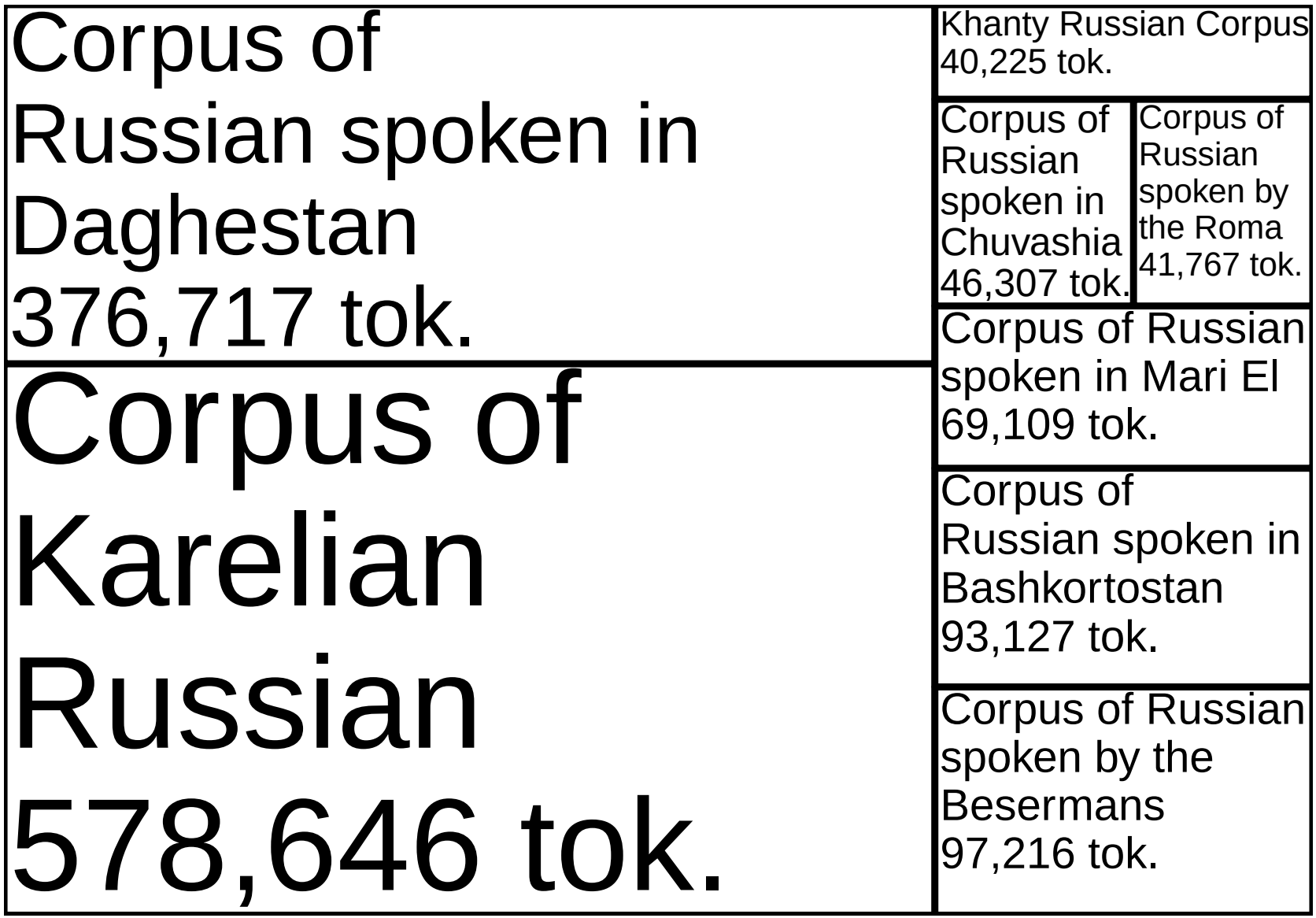
Resources of the Linguistic Convergence Laboratory

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

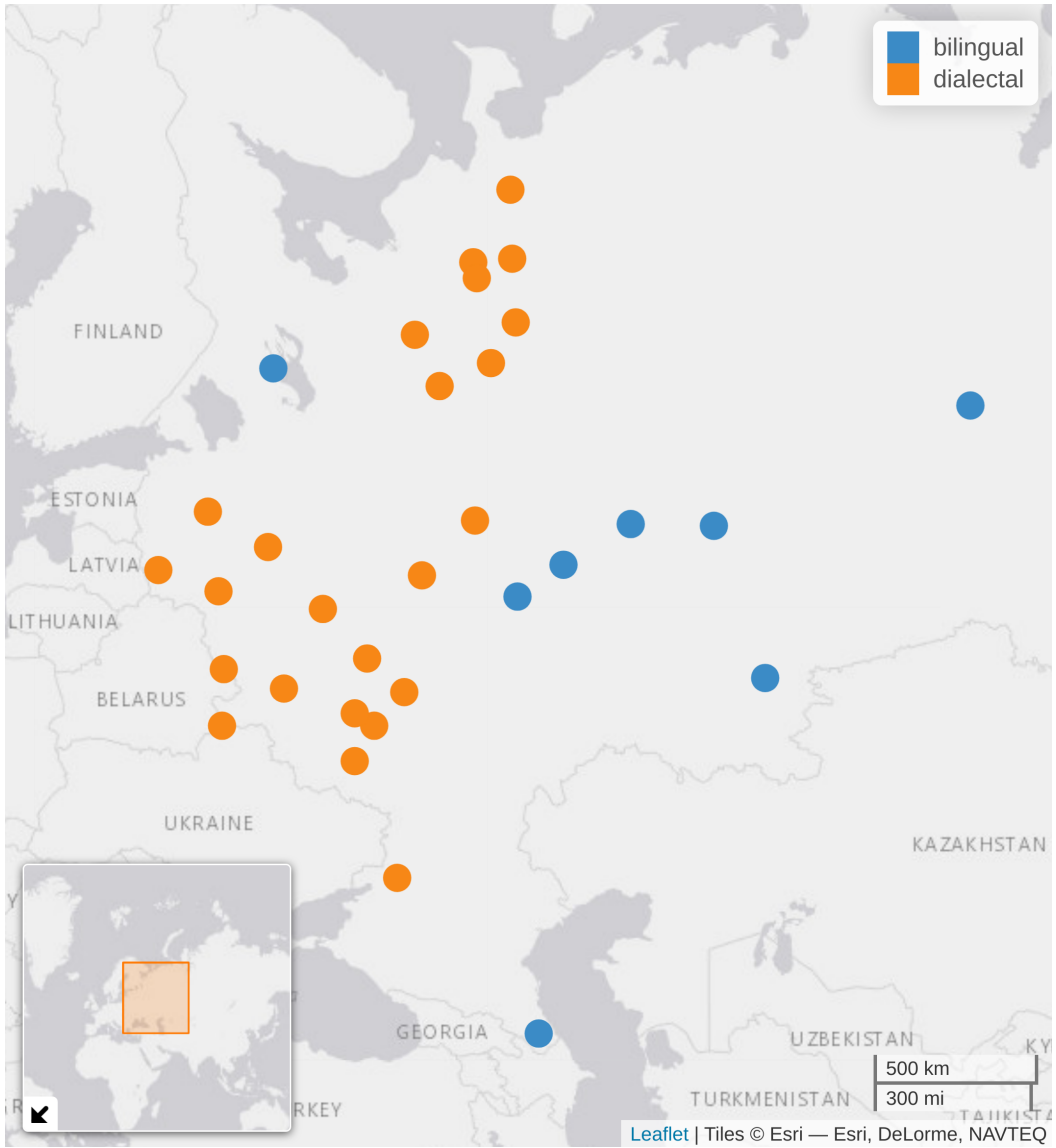
Dialectal Corpora



Bilingual Corpora



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?

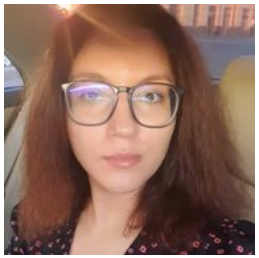
Can we find different variation patterns?

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

The DiaL2 sideproject

Future plans

References I

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