



Project «Spatiotemporal model design and application for spoken Latgalian language corpus entry-data multifaceted analysis» contributes to the advancement of digital humanities in Latvia

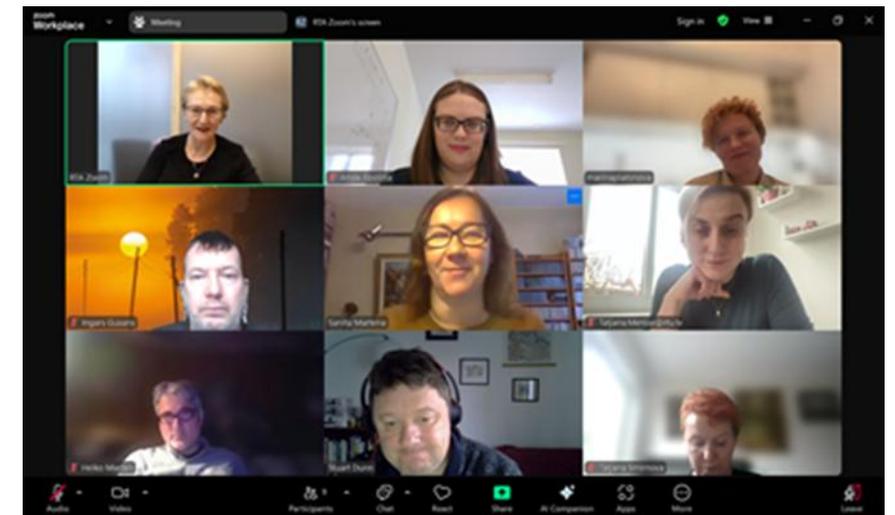
Dr.philol. Antra Kļavinska
15.01.2026

Izstrādāts ar ES ANM atbalstu projekta Nr. 5.2.1.1.i.0/2/24/I/CFLA/003 P&A grantā "Spatiotemporālā modeļa izstrāde un pielietošana Latgaliešu runas korpusa datu daudzveidīgas analīzes nolūkos", Nr. RTU-PA-2024/1-0063

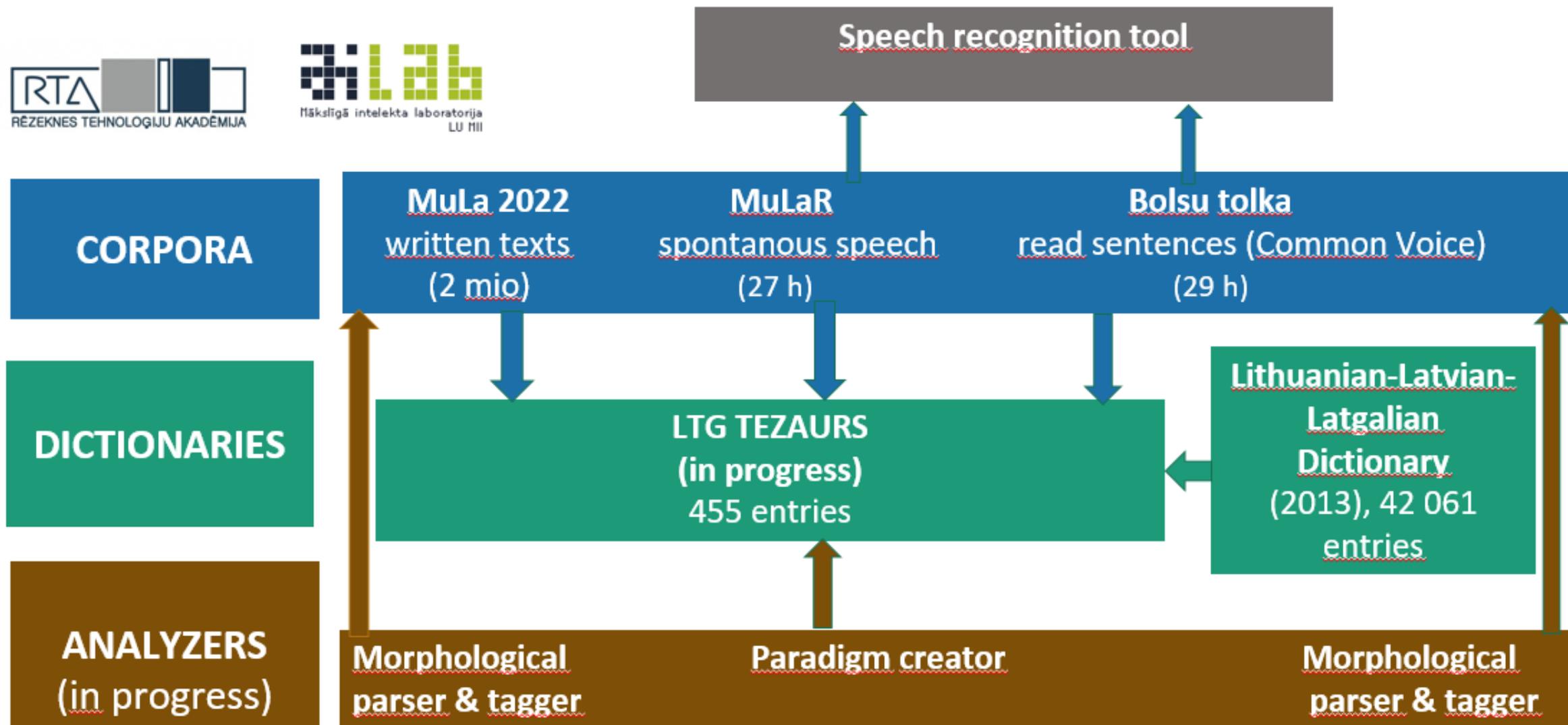
R&D grant «Spatiotemporal model design and application for spoken Latgalian language corpus entry-data multifaceted analysis»

Project partners:

- **RTU Rēzekne, Centre for Education, Languages and Social Technologies:** Antra Kļavinska, Sanita Martena, Heiko Marten, Ingars Gusāns, Anda Āboliņa;
- **RTU Institute of Digital Humanities:** Marina Platonova, Tatjana Smirnova, Tatjana Menise, Zane Seņko, Anna Pačerska;
- **King's College London:** Stuart Dunn.

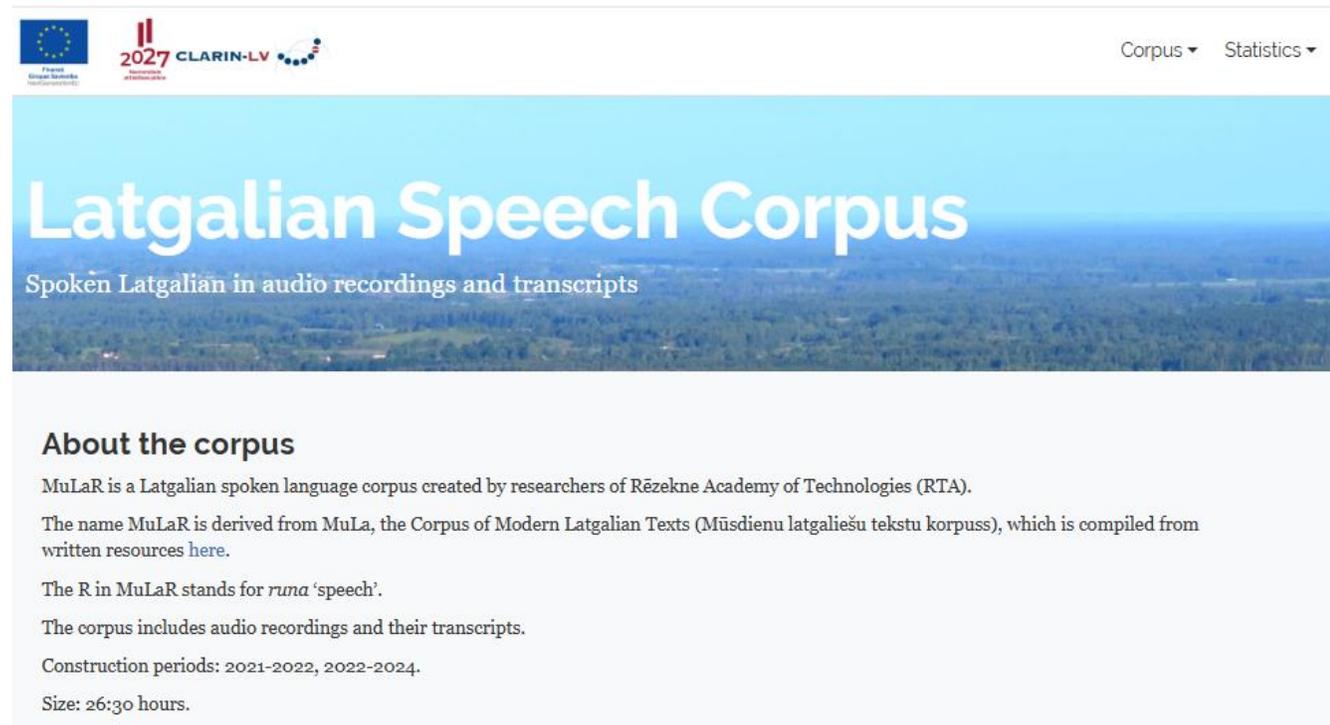


Latgalian language digital resources and tools



The goal: to design a prototype of a spatiotemporal model for conducting efficient multifaceted analysis of a corpus of contemporary Latgalian speech (MuLaR)

- <https://mularkorpuss.rta.lv>
- <https://korpuss.lv/id/MuLaR>
- <https://repository.clarin.lv>



The screenshot shows the website for the Latgalian Speech Corpus. At the top, there are logos for the European Union, the 2027 National Development Plan, and CLARIN-LV. The main heading is "Latgalian Speech Corpus" with the subtitle "Spoken Latgalian in audio recordings and transcripts". Below this, there is a section titled "About the corpus" which provides details about the corpus's origin, name, and construction.

Corpus ▾ Statistics ▾

Latgalian Speech Corpus

Spoken Latgalian in audio recordings and transcripts

About the corpus

MuLaR is a Latgalian spoken language corpus created by researchers of Rēzekne Academy of Technologies (RTA).

The name MuLaR is derived from MuLa, the Corpus of Modern Latgalian Texts (Mūsdienu latgališu tekstu korpuss), which is compiled from written resources [here](#).

The R in MuLaR stands for *runa* 'speech'.

The corpus includes audio recordings and their transcripts.

Construction periods: 2021-2022, 2022-2024.

Size: 26:30 hours.

Significant achievements

- Enhancement of the MuLaR corpus with 6 hours of audio recordings and their transcriptions.
- Application of a new methodology for corpus data visualization, leading to the creation of a spatiotemporal model prototype.

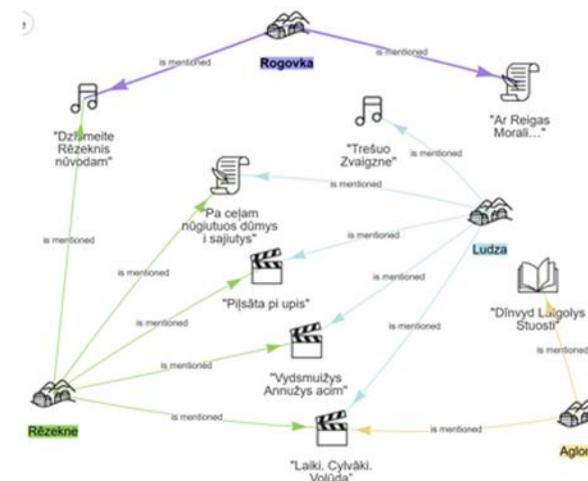


Figure 3. A model illustrating the web of texts that immerse the Latgalian region.

Impact:

- Recognition of the role of the Latgalian language as a member of the family of lesser-used European regional languages and acknowledgment of the spatiotemporal aspects of its use will allow developing of a more comprehensive model of the European and global linguistic landscape.
- The spatiotemporal prototype designed as one of the project deliverables is going to become the first speech corpus visualisation model in Latvia.
- The expansion of MuLaR and the integration of spatiotemporal metadata into the corpus will open wider opportunities for its further use as a resource in geospatial, ethnographic, and linguistic research.



Future collaboration opportunities

Horizon project application “Vigorous and Open Interdisciplinary Cooperation Ecosystem (VOICE)” (under evaluation), participation in the National Research Programme “Digital Humanities” etc.



Finansē
Eiropas Savienība
NextGenerationEU



2027
Nacionālais
attīstības plāns



1862
RĪGAS TEHNISKĀ
UNIVERSITĀTE

Antra.Klavinska@rtu.lv