



MŪSDIENU LATGALIEŠU RUNAS KORPUSA (MuLaR) IZVEIDE MAZĀK LIETOTO VALODU DOKUMENTĒŠANAS KONTEKSTĀ

ANTRA.KLAVINSKA@RTA.LV

SEMINĀRS «RUNAS KORPUSU IZVEIDE UN TO IZMANTOJUMS» 18.05.2023.

- Latgaliešu valoda ar 150 000 uzrādītiem lietotājiem 2013. gadā atzīta par vienu no pasaules apdraudētajām valodām, kas pakļauta riskam (UNESCO, Moseley 2020).
- ES direktīvas un ieteikumi reģionālo un riskam pakļauto valodu saglabāšanai, izpētei un attīstībai; Valsts valodas politikas nostādnes 2021. – 2027. gadam attiecībā uz daudzveidīgu tekstu korpusu izstrādi, publiskošanu tīmeklī un pieejamību.
- Mazāk lietoto valodu dokumentēšanas priekšrocības, veidojot runas korpusus: pieejamība un ilgtspējība; liels oriģinālu, digitāli arhivētu, transkribētu, anotētu un katalogizētu datu apjoms; daudzfunkcionalitāte; kontekstualizācija u.tml. (Rießler, Wilbur 2017: 60).

KORPUSA IZVEIDE

- Projekts TrimCo “Triangulation Approach for Modelling Convergence with a High Zoom-In Factor” ir realizēts, pateicoties prof. Dr. Björn Wiemer (German Research Foundation, Johannes-Gutenberg-University Mainz, projekta numurs: WI 1286/16-1).
- Projekts “Humanitāro zinātņu digitālie resursi: integrācija un attīstība” (Nr. VPP-IZM-DH-2020/1-0001) no 2020.gada oktobra līdz 2022.gada oktobrim tika īstenots Valsts pētījumu programmā “Humanitāro zinātņu digitālie resursi”.
- Valsts pētījumu programmas projekts “Latviešu valodas daudzveidība laikā un telpā (LaDi)” (Nr. VPP-LETONIKA-2021/4-0003) tiek īstenots no 20.12.2021. līdz 19.12.2024.

DATU IEGUVE

- Dažādos Latgales novados 2009. – 2021. gadā ierakstītas intervijas, tostarp “TrimCo” projekta dati.
- Sibīrijā dzīvojošo latgaliešu runas ieraksti (lauka pētījumi 2017. – 2018. gadā, kas līdz šim daļēji bija ievietoti austrumu diasporas latgalisko izlokšņu datubāzē (www.reiti.rta.lv).
- TV un radio ieraksti (2018. –2020. gadā Latgales Reģionālajā televīzijā veiktas intervijas; 2020.-2023.g. Latvijas radio Latgales studijas ieraksti).

DATU APSTRĀDE

00:00:38.099

Time Interval: 00:00:36.760 - 00:00:39.340 2580

Volume 100

Rate 100

Settings

- ☒ Automatic playback of media
- ☒ Create missing annotations
- ☒ Show tier names
 - ☐ Colors only on "No." column
- ☐ Navigate across column
- ☐ Scroll current annotation to center

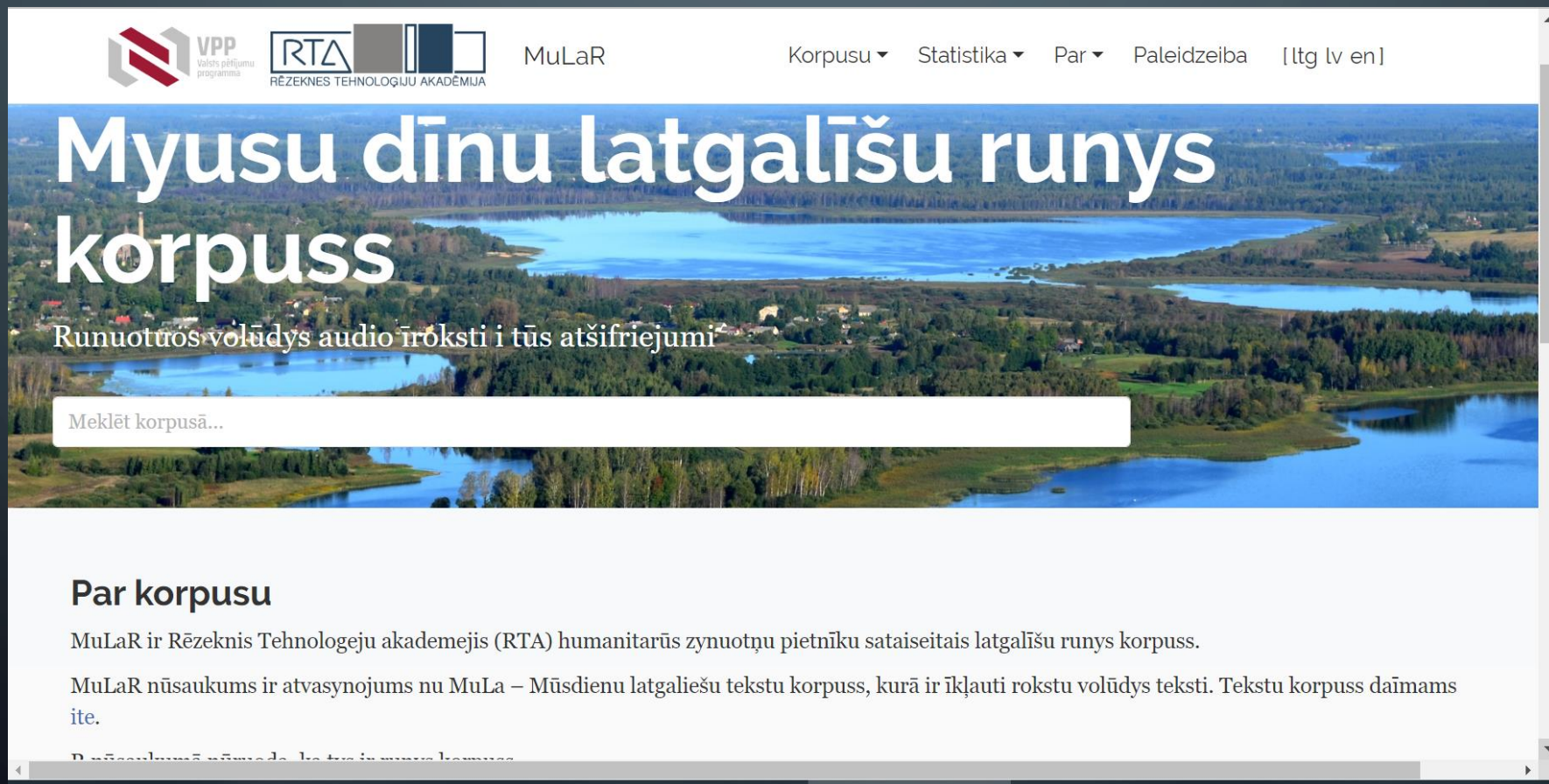
Configure...

No	Type 1 : default-It
	F421-2007
10	vacuoki visi runuo ar mani latgaliski
11	i vecmammys vectāvi tože
	i20-F
12	a ar klasibīdrim kai sasarunojīt voi školys bīdrim puļceņu bīdrim?
	F421-2007
13	eee školā ar visim runojam latviski
14	ir meitine kas runoj latgaliski bet maņ ir kaut kai nn naīrostai ar jū runuot školā latgaliski he
15	deļtuo es vīnmār atbyldu latviski jai
	i20-F
16	a ot parkū školā tai līkās naīrasti runuot latgaliski?
	F421-2007
17	nazynu kaut kai nu pyrmuos klasīs ir pīrosts jau un tagad navaru īsēst tymā

KORPUSA PLATFORMAS IZVEIDE

[HTTPS://MULARKORPUSS.RTA.LV](https://mularkorpuss.rta.lv)

Tīmekļa saskarne SpoCo (Waldenfels, Woźniak 2016).



KORPUSA IESPĒJAS

MuLaR

Corpus ▾ Statistics ▾ About ▾ Help [ltg lv en]

▶ 0:00 / 0:00 — 🔊 ⋮

a ka īrauga **tāvs** sasītu tiuleņ giva rūkuos mītu

Person ID: F124-1929
Source: L

Year of birth: 1929

Sex: f

Place: Cibla



▶ 0:00 / 0:00 — 🔊 ⋮

F124-1929: mobilizēja syutēja a - ha tod nū - nūskrēja ▶ no tuo i atgruo - i pajēme meitu vot i navajadzēja jai iz tom šahtom braukt ▶

io8-F: a kai jius nūklivot Krīvejī ? ▶

F124-1929: **nu tāvs bāga nu vuocišim** ▶ nabeja vēl partejī īleids bet jau kandidatūs ▶ pats bogotuokīs i pats komunists beja īstuostēja kai ▶ nu tāvs beja jis redzi tymā ▶

Person ID: F124-1929
Source: L

Year of birth: 1929

Sex: f

Place: Cibla



▶ 0:00 / 0:00 — 🔊 ⋮

nu **tāvs** beja jis redzi tymā

Person ID: F124-1929
Source: L

Year of birth: 1929

Sex: f

Place: Cibla



SECINĀJUMI UN IECERES

- Iespējas: pētnieki var radoši izpausties, pētīt un izvēlēties viņiem saistošāko korpusa dizainu, datorprogrammas, iegūt pieredzi no ārzemju kolēģiem, pētniekiem u. tml.
- Izaicinājumi: darbs pie runas korpusa izveides ir komplicēts un laikietilpīgs.
- Korpusa apjoms (05.2023.): 172 309 vārdlietojumi (~20 h audio ierakstu transkripcijas). Mērķis – 2024.g. korpusā ievietotas ~25 h audio ierakstu transkripcijas.
- Korpusa popularizēšana: semināri, videolekcijas skolotājiem:
https://ltg.korpuss.rta.lv/video_lekcijas/
- Vairāk par MuLaR korpusa izveidi: Juško-Štekele, Angelika, Kļavinska, Antra (2022). Mūsdienu latgaliešu valodas runas korpusa izveide mazāk lietoto valodu dokumentēšanas kontekstā. *Letonica*, 47. 226.-242.lpp.
http://lulfmi.lv/files/letonica/Letonica_47.pdf

LITERATŪRA

- Moseley, Christopher (ed.) (2010). *Atlas of the World's Languages in Danger*, 3rd edn. Paris, UNESCO Publishing.
- Rießler, Michael, Wilbur, Joshua (2017). Documenting endangered oral histories of the Arctic: A proposed symbiosis for endangered language documentation and oral history research, illustrated by Saami and Komi examples. Kasten, Erich, Roller, Katja, Wilbur, Joshua (eds). *Oral History Meets Linguistics*. Fürstenberg, Havel: Kulturstiftung Sibirien, pp. 31–64.
- von Waldenfels, Ruprecht, Woźniak, Michał (2016). SpoCo – a simple and adaptable web interface for dialect corpora. Kupietz, Marc, Geyken, Alexander (eds.). *Journal for Language Technology and Computational Linguistics. Corpus Linguistic Software Tools*. Vol.31, Number 1, 155-170.