



LEĻMŌ ILĻISKŪOL
LĪVŌD
INSTITŪT



Lībiešu valodas datu ieguve un tās nākotnes perspektīvas

Valts Ernštreits, LU Lībiešu institūts

Valsts pētījumu programma Atvērtas un FAIR principi atbilstīgas
digitālo humanitāro zinātņu ekosistēmas attīstība Latvijā (Dheli)



Lībiešu valodas dati

- **Leksikogrāfiskie dati** – lemmas, atbilstmes latviski, igauņiski, angļiski
- **Gramatiskā informācija** – vārdu formas un variācijas
- **Tekstuālie dati** – korpuss, t.sk. piemēri; arī paralēlkorpuss – atbilstmes latviski, igauņiski, angļiski
- **Audio dati** – lībiešu valodas lemmas un piemēri
- **Informācija par avotiem** – ziņas par avotu tapšanas apstākļiem vai saturu



Lībiešu valodas dati

- **Ģeotelpiskie dati** – kartēti lībiešu valodas vietvārdi, kas caur sasaisti ar tekstu, tā teicēju un viņa izcelsmes vietu ļauj piesaistīt valodas datus kartei
- **Ģeotelpiskos datus papildinoša informācija** – ziņas par norisēm ar vietvārdu saistītajā objektā
- **Teicēju dati** – kas caur sasaisti ar tekstu un teicēja izcelsmes vietu ļauj piesaistīt valodas datus kartei



Vārdi

kōgaztō *vj □137*

attālināties

a'igi *s *181*

sienas, dārza u.c. mala, -male

pappōl *s *219*

papele

lu'm *s *81*

sniegs

◆ **pāgiņ lūnda** daudz sniega

lībiešu ▾



ievadi atslēgas vārdu...

Meklēt





Vārdi

Formas

Teksti

Cilvēki

Avoti

Vietas

Vārdi



Aa



LI

āāāčēģīķļņōōōōōōōrštūžū'



Meklēt



Valoda

LV

EE

EN

Kārtošana

Pareiza

Apgriezta

Skaits

50

Redzami 1 - 50 no 13309 šķirkļiem

+ Pievienot



1

2

3

4

5

6

7

8

9

10



a'b ¹_s *72

palīdzība

○ pakānd | a'b kiirabi = ātrā palīdzība

◆ a'bbō andō sniegt palīdzību

a'bbō vótšō meklēt palīdzību

● A'b tu'ļ ōigiz āigal. Palīdzība nāca īstajā laikā.

Kilāro'v a'b kuo'igmīredōn tu'ļ aigstiz. Ciema ļaužu palīdzība kuģiniekiem pienāca laikā.

Kielk Tiļļōstōks āndizt t'dtuoizōn a'bbō. Ķelķis ar Tiļļu sniedza viens otram palīdzību.

Ta vótšūb mī'nstō (~ mī'n jūstō) a'bbō. KK78.3 Viņš gaida [~ meklē] no manis (~ pie manis) palīdzību.



a'b ²_s *72

plecs

◆ abīdi nustō KK77b raustīt plecus

● Pā u'm abūd va'isō sī'zzōl vie'dtōd. KK77 1 Galva ir ievilkta plecos.



Datu izmantošana: morfoloģija

- **Morfoloģiskā analīze, paradigmu piederības identificēšana un pareizrakstības rīki** – morfoloģiskā informācija (Trumses Arktiskā Universitāte un Helsinku Universitāte, Džeks Rueters un Tronds Trūsterunds)



LETMŌ IĻĪSKŪOL
LĪVŌD
INSTITŪT



Datu izmantošana: audio un paralēlkorpus

- **Mašīntulkošana** – leksikogrāfijas datubāzes, lībiešu valodas korpusa un paralēlkorpusa dati (Tartu Universitātes NLP grupa, vadītājs Marks Fišels)
- **Runas sintēze un atpazīšana** – audio dati (Tartu Universitātes NLP grupa, vadītājs Marks Fišels)



LEŅMŌ IĻĪSKŪOL
LĪVŌD
INSTITŪT



Datu izmantošana: ģeotelpiskā informācija

- **Vietvārdu dati** – kartēto lībiešu vietvārdu datubāze (SIA «Jāņa sēta»)



LETMŌ IĻĪSKŪOL
LĪVŌD
INSTITŪT



Datu izmantošana: mācību materiāli

- **Visi valodas dati un instrumenti**
– mācību programma «Op līvõ kīeldõ!» (Mācies lībiešu valodu!)

Video materiāli

[Home](#) > Video materiāli

Op līvõ kīeldõ



Es esmu / Minā um - 1. stunda



Es esmu / Minā um - 2. stunda



Sveiki! / Tēriņš! - 3. stunda



Kā klājas? / Kui jelād? - 4. stunda



Mana ģimene / Min aim - 5. stunda



Skaitļi un pulksteņa laiks / Lugūd ja kīela - 6. stunda



LEŅMŌ IĻĪSKŪOL
LĪVŌD
INSTITŪT



VPP
Valsts pētījumu
programma



Izaicinājumi datu ieguvē

- **Ierobežots datu masīvs** – īpašas pieejas datizracē, lai maksimāli izmantotu datos esošo informāciju
- **Ierobežots valodas pratēju skaits** – valodas pratēju iesaiste, mašintulkošanas, morfoloģiskās analīzes u. c. paņēmieni izmantošana



LETMŌ IĻĪSKŪOL
LĪVŌD
INSTITŪT



Izaicinājumi datu ieguvē

- **Ierobežota iepriekšējā izpēte un standartizācija** – integrēts nepabeigtības, alternatīvo iespēju un dinamisku pārmaiņu princips (piemēram, ārvietas locījumi), mikropētījumi
- **Līdzsvars un konsekvence** – citu valodu pamatleksikas datubāžu un korpusa biežuma datu izmantošana (piem., EKI)



LEŢMŌ IĻĻISKŪOL
LĪVŌD
INSTITŪT



Izaicinājumi datu ieguvē

- **Teritoriāli izkliedēti dati** – vienošanās ar datu turētājiem (Igaunijas Literatūras muzejs, Igaņu valodas institūts, Helsinku bibliotēka), starpvalstu vienošanās
- **Rakstīto tekstu griesti** – runas atpazīšanas tehnoloģiju izpēte



LETMŌ IĻĻISKŪOL
LĪVŌD
INSTITŪT





LEĶMŌ IĻĻISKŪOL
LĪVŌD
INSTITŪT



Paldies! | Tienū!

