

# Latviešu valoda digitālajā laikmetā - ELE projekta pētījums

Inguna Skadiņa



# Otrais Gūtenberga efekts

Valodas, kuras netiks pietiekami izmantotas modernajās tehnoloģijās, var saskarties ar digitālo izmiršanu



“Language and Technology in Europe 2000”  
(1994)



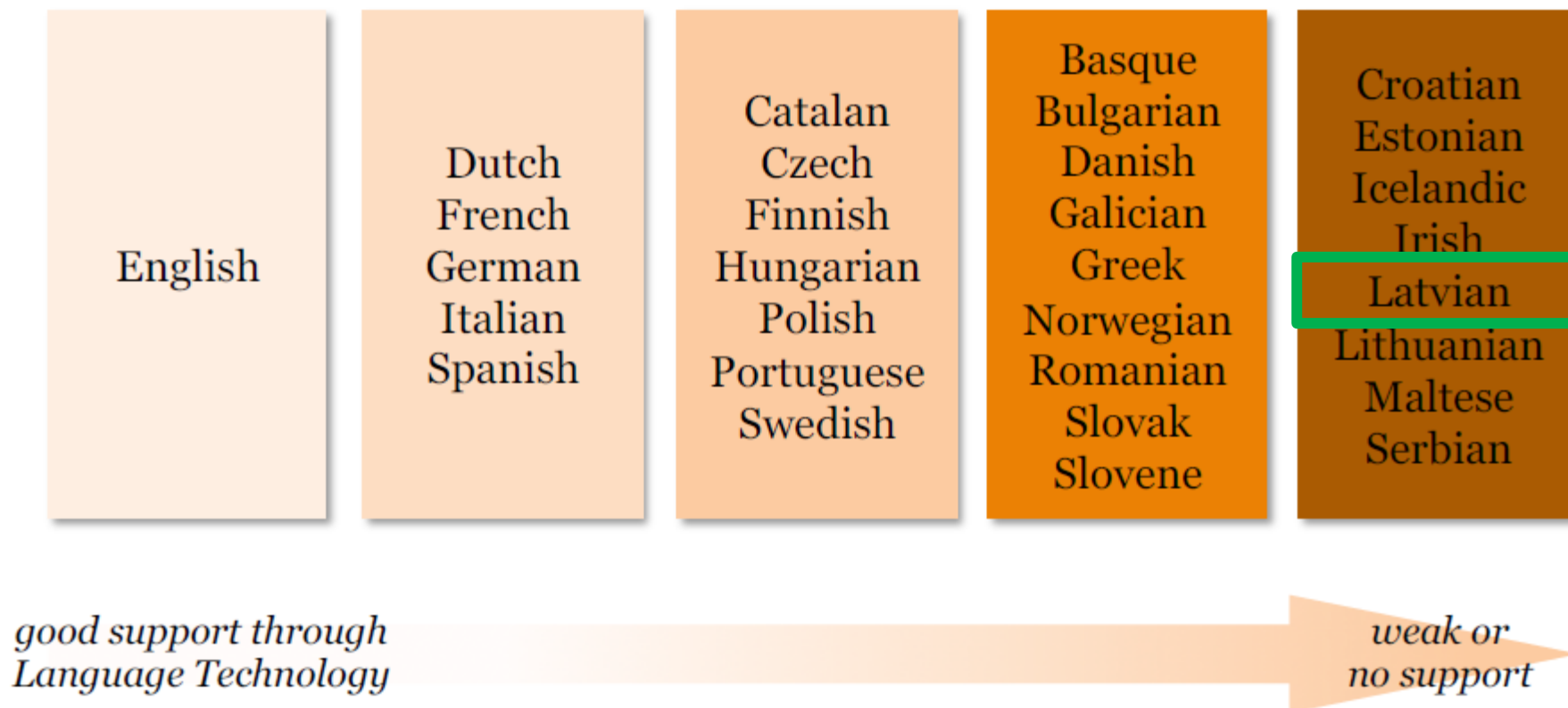
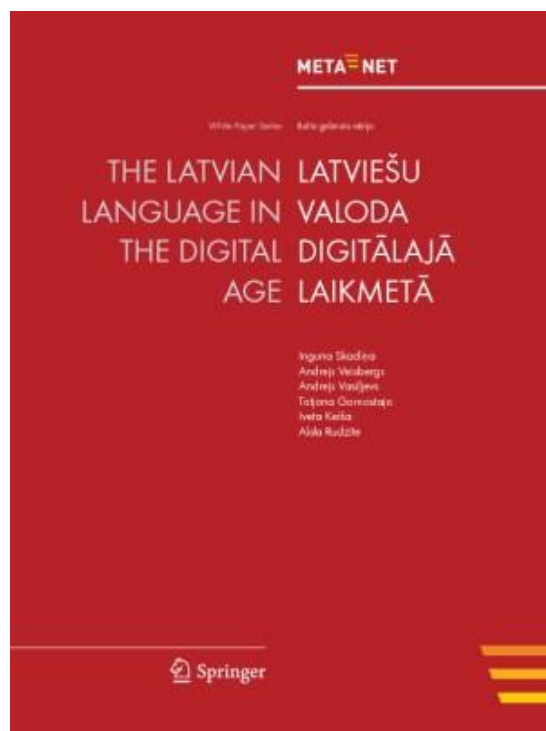
# 21

Eiropas valodas  
ilgtspēja ir apdraudēta,  
jo tām ir nepietiekams  
tehnoloģiskais  
nodrošinājums

*META-NET Language Whitepapers*

<http://www.meta-net.eu/whitepapers/overview>

# Valodas resursi un rīki Eiropas valodām



# Eiropas Parlamenta 2018. gada 11. septembra rezolūcija par valodu līdztiesību digitālajā laikmetā



BG ES CS DA DE ET EL EN FR GA HR IT LV LT HU MT NL PL PT RO SK SL FI SV

Indekss < Iepriekšējais Nākošais > Pilns teksts

Procedūra : 2018/2028(INI)			Dokumenta lietošanas cikls : A8-0228/2018	Dokumenta lietošanas cikls sēdē
Iesniegtie teksti : A8-0228/2018	Debates : PV 10/09/2018 - 27 CRE 10/09/2018 - 27	Balsojumi : PV 11/09/2018 - 6.15 Balsojumu skaidrojumi	Pieņemtie teksti : P8_TA(2018)0332	

<b>Pieņemtie teksti</b>	178k 54k
Otrdienā, 2018. gada 11. septembris - Strasbūra	
Valodu līdztiesība digitālajā laikmetā	P8_TA(2018)0332 A8-0228/2018

## Eiropas Parlamenta 2018. gada 11. septembra rezolūcija par valodu līdztiesību digitālajā laikmetā (2018/2028(INI))

Eiropas Parlaments,

- ņemot vērā Līguma par Eiropas Savienības darbību (LESD) 2. pantu un 3. panta 3. punktu,
- ņemot vērā Eiropas Savienības Pamattiesību hartas 21. panta 1. punktu un 22. pantu,
- ņemot vērā UNESCO 2003. gada Konvenciju par nemateriālā kultūras mantojuma saglabāšanu,
- ņemot vērā Eiropas Parlamenta un Padomes 2003. gada 17. novembra Direktīvu 2003/98/EK par valsts sektora informācijas atkalizmantošanu<sup>(1)</sup>,
- ņemot vērā Eiropas Parlamenta un Padomes 2013. gada 26. jūnija Direktīvu 2013/37/ES, ar ko groza Direktīvu 2003/98/EK par valsts sektora informācijas atkalizmantošanu<sup>(2)</sup>,
- ņemot vērā Eiropas Parlamenta un Padomes 2015. gada 25. novembra Lēmumu (ES) 2015/2240, ar ko izveido programmu Eiropas publiskās pārvaldes iestāžu, uzņēmumu un iedzīvotāju sadarbības risinājumu un kopīgu sistēmu nodrošināšanai (programma /SA2) kā līdzekli publiskā sektora modernizācijai<sup>(3)</sup>,

592 – par  
45- pret

[https://www.europarl.europa.eu/doceo/document/TA-8-2018-0332\\_LV.html](https://www.europarl.europa.eu/doceo/document/TA-8-2018-0332_LV.html)

# Eiropā pastāvošie šķēršļi valodu vienlīdzības panākšanai digitālajā laikmetā

- ... adekvātas politikas trūkuma dēļ Eiropā **pašlaik paplašinās tehnoloģiskā plaisa starp valodām**, kam resursu netrūkst, un valodām, kam to ir mazāk, neatkarīgi no tā, vai šīs pēdējās ir oficiālās, otrās oficiālās vai neoficiālās ES valodas; turklāt pauž nožēlu par to, ka **vairāk nekā 20 Eiropas valodām draud digitālā valodas izzušana**;
- ... **mazāk lietotajām Eiropas valodām** ir būtiski sliktāki apstākļi, jo **akūti trūkst rīku, resursu un finansējuma pētījumiem**, un tas kavē un sašaurina darbu, ko veic pētnieki, kas nespēj no valodu tehnoloģijām gūt maksimālu labumu, pat ja viņiem ir vajadzīgās tehnoloģiskās prasmes;
- ... vajadzīgi vispārīgi ES līmeņa **pasākumi** (rīcībpolitika, finansējums, izpēte un izglītošana), kas **nodrošinātu VT izstrādi mazāk lietotām oficiālajām ES valodām**, un īpaši ES līmeņa pasākumi (rīcībpolitika, finansējums, izpēte un izglītošana), kas šajā izstrādē iekļautu un popularizētu reģionālās un minoritāšu valodas;
- Atzinīgi vērtē iepriekšējo ES finansēto pētniecības tīklu, piemēram, FLaReNet, CLARIN, HBP un META-NET (tostarp META-SHARE), svarīgo celmlauža lomu, ko tie pildījuši Eiropas valodu tehnoloģiju platformas izveidē
- .....

# Ieteikumi ES pētniecības politikas jomā

- Aicina Komisiju izveidot Eiropas, valsts un reģionālā līmenī īstenojamu, plašu un saskaņotu **ilgtermiņa finansēšanas programmu pētniecībai, izstrādei un inovācijai valodu tehnoloģiju jomā**, īpaši pielāgojot to Eiropas vajadzībām un prasībām;
- Uzsver, ka ar šīs programmas palīdzību būtu jācenšas sekmēt padziļinātu **dabiskās valodas izpratni** un palielināt efektivitāti, **daloties zināšanās, infrastruktūrā un resursos**, lai izstrādātu inovatīvas tehnoloģijas un pakalpojumus nolūkā sasniegt kārtējos zinātniskos pavērsienus šajā jomā un **mazināt tehnoloģisko plaisu** starp Eiropas valodām;
- Turklāt uzsver, ka šim projektam vajadzētu būt atvērtam, balstītam uz mākoņtehnoloģijām un sadarbspējīgam un **nodrošināt ļoti mērogojamus augstas veiktspējas pamatrīkus** vairākām **valodu tehnoloģiju lietojumprogrammām**;

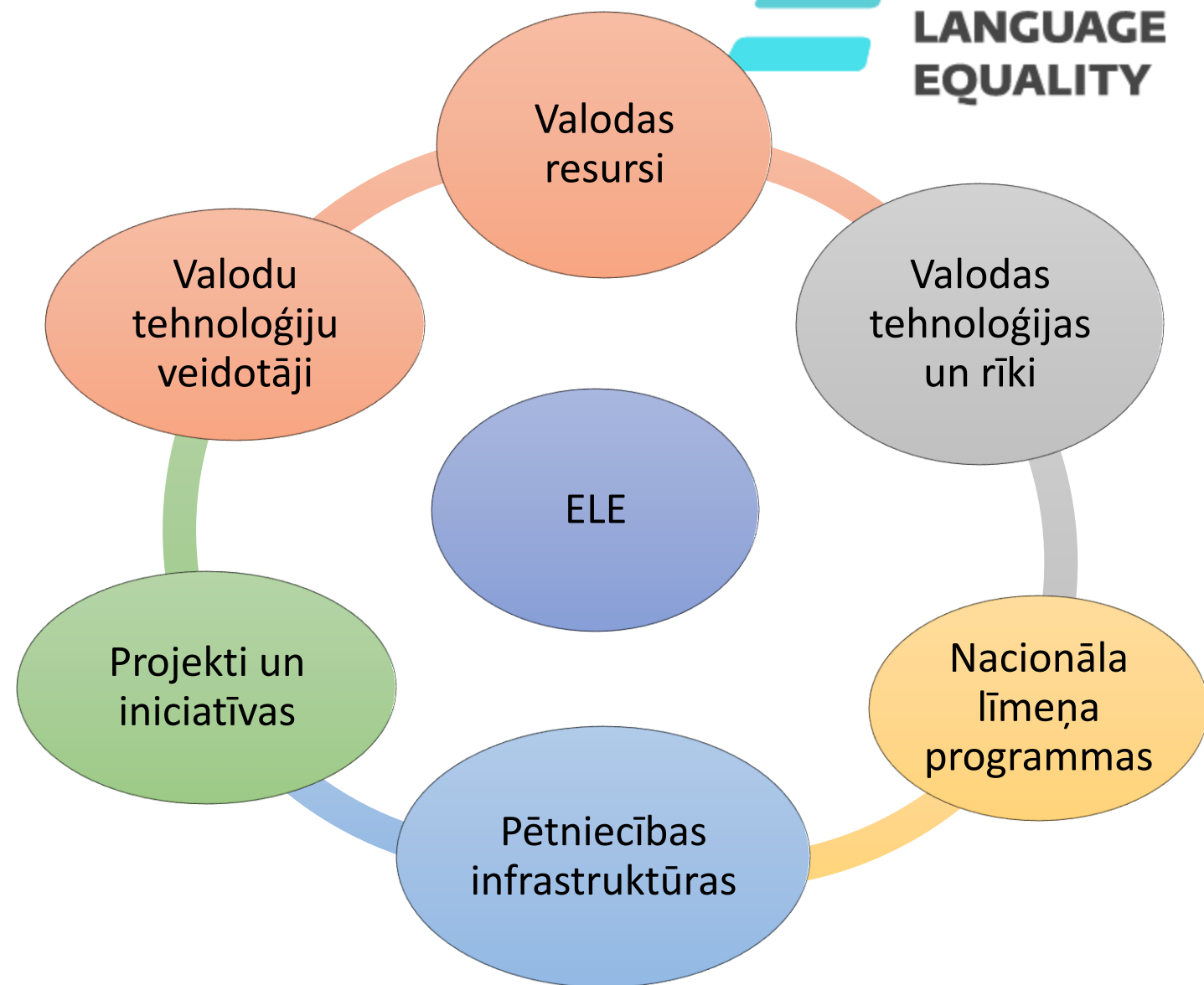
# Developing an agenda and a roadmap for achieving full digital language equality in Europe by 2030

 ELE

ELE (*European Language Equality*) projekta mērķis ir **apzināt pašreizējo Eiropas valodu stāvokli digitālajā vidē** un izstrādāt stratēģiju valodu līdztiesības nodrošināšanai 2030. gadā.



D1.17	Report on Greek
D1.18	Report on Hungarian
D1.19	Report on Icelandic
D1.20	Report on Irish
D1.21	Report on Italian
D1.22	Report on Latvian
D1.23	Report on Lithuanian
D1.24	Report on Luxembourgish



Runātās un rakstītās valodas  
resursi

# tezaurs.lv

## vārdnīcu un leksikonu platforma

**[LLVV]:** Latviešu literārās valodas vārdnīca 64 265 šķirķji  
 **[MLVV]:** Mūsdienu latviešu valodas vārdnīca 45 939 šķirķji  
 **[T]:** Rudens versija 2022 387 576 šķirķji

**valsts<sup>1</sup>**  
 valsts sieviešu dzimtes 6. deklinācijas lietvārds Locīšana

1. Politiska organizācija, ar kuru tiek realizēta sabiedrības vispārobligāta vadīšana, pastāvošas iekārtas saglabāšana;

1.1. pārnestās nozīmē **Telpa, vieta, apvidus**, arī cilvēka dzīves **joma**, kam raksturīgs kādu parādību kopums.

2. **novēcojis\_Pagasts.**  
 Saistītās nozīmes: pagasts, valstība  
 Stabils vārds savienojumi: Valsts (arī pagasta) māja  
 Tulkojumi: municipality

3. **joma: bioloģija Augstākā sistemātikas vienība bioloģijā.**  
 Saistītās nozīmes: fauna, flora, sēnes, sēne, protisti  
 Tulkojumi: kingdom

Avoti: LLVV, LE, TWN

Vārdnīcu Tehniskā Pēdējais  
 Vārdnīcu Tehniskā Pēdējais

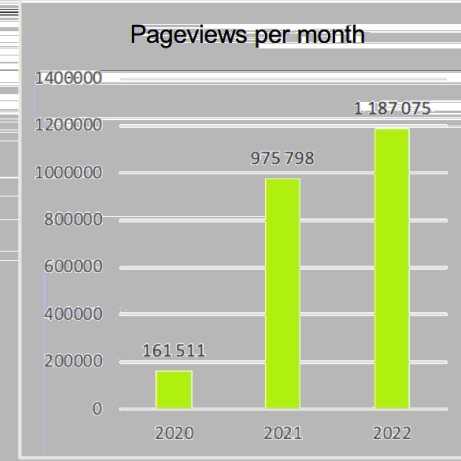
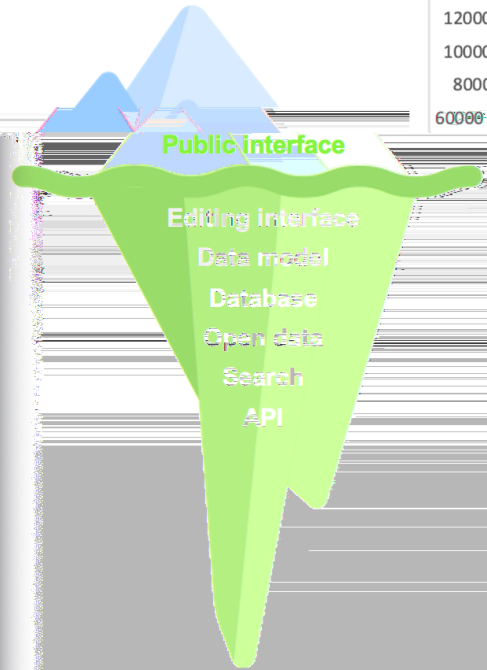
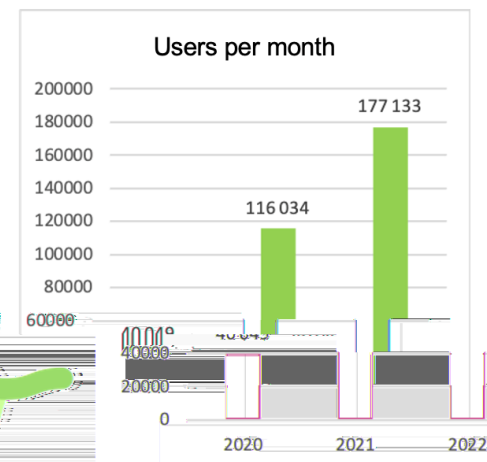
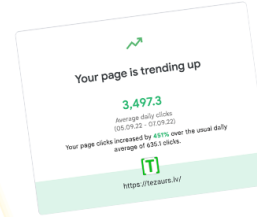
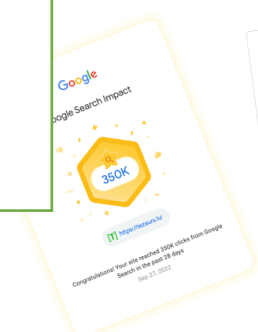
Vārdnīcu sastādījis Andrejs Spektors  
 LU MII Mākslīgā intelekta laboratorija, 2009–2022  
 Pēdējais atjauninājums: 2022. gada 23. septembrī

Kas ir Tēzauris  
 Apzīmējumi  
 Izmantotie avoti  
 Citas vārdnīcas

Android lietotne  
 Kindle vārdnīca  
 Atvērta dati  
 Tēzaura API

Autori  
 Pateicības  
 Iestātijumi  
 BUJ

Seko mums:





[T]:

vārds vai frāze



## Apkaime

Ejstrupholma

ejū

ek

EK

ēk

ek.

eka

**ēka**

eka-ekamangāns

eka-elementi

ekacēzijs

ekajods

Ekale

ekamangāns

ekārs

## ēka

ēka sieviešu dzimtes 4. deklinācijas lietvārds ▼ Locīšana

1. Celtne, ko parasti izmanto dzīvošanai, saimnieciskām, ražošanas vai sabiedriskām vajadzībām; atsevišķs objekts celtņu kompleksā.

▼ Piemēri *Lielā sarkano ķieģeļu ēka, ābeļdārza ieskauda, elpoja siltu mieru.*

▼ Saistītās nozīmes *būve, celtne*

^ Tulkojumi

*building, edifice*

a structure that has a roof and walls and stands more or less permanently in one place; "there was a three-story building on the corner"; "it was an imposing edifice"

[Princeton WordNet 3.0]

2. novecojis Klēts, vēsā ēka dažādu produktu uzglabāšanai.

▼ Saistītās nozīmes *klēts*

▼ Tulkojumi *barn*

AVOTI: LLVV, EH, TWN

KORPUSA PIEMĒRI: [šeit](#)

Ziņot

Dalīties



Vārdnīcu sastādījis Andrejs Spektors

LU MII Mākslīgā intelekta laboratorija, 2009–2022

Pēdējais atjauninājums: 2022. gada 23. septembrī

Kas ir Tēzauris

Apzīmējumi

Izmantotie avoti

Citas vārdnīcas

Android lietotne

Kindle vārdnīca

Atvērtie dati

Tēzaura API

Autori

Pateicības

Iestatījumi

Seko mums:

<http://tezauris.lv>

# WordNet

## ēka

ēka sieviešu dzimtes 4. deklinācijas lietvārds ✓ Locīšana

1. **Celtne**, ko parasti izmanto dzīvošanai, saimnieciskām, ražošanas vai sabiedriskām vajadzībām; atsevišķs objekts celtņu kompleksā.

- ✓ Piemēri *Lielā sarkano ķieģeļu ēka, ābeļdārza ieskauda, elpoja siltu mieru.*
- ✓ Saistītās nozīmes *celtne*
- ✓ Tulkojumi *building, edifice*

2. novecojis **Klēts**, vēsā ēka dažādu produktu uzglabāšanai.

### ^ Saistītās nozīmes

#### Sinonīmi ⓘ

*klēts<sub>1</sub>* — Lauksaimniecības celtne graudu uzglabāšanai.

### ^ Tulkojumi

#### *barn* ↗

an outlying farm building for storing grain or animal feed and housing farm animals

[Princeton WordNet 3.0]

AVOTI: LLVV, EH, TWN



Spektors, A., Auziņa, I., Dargis, R., Grūzītis, N., Paikens, P., Pretkalniņa, L., Rituma, L., Saulīte, B. [Tēzaurs.lv: the largest open lexical database for Latvian](https://tezaurs.lv). Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC), 2016; <https://tezaurs.lv>

Lokmane, I., Rituma, L., Stāde, M., Klints, A. [Latvian WordNet and Word Sense Disambiguation: Challenges and Findings](https://wordnet.ailab.lv). Proceedings of the 7th International Conference on Electronic Lexicography (eLex), 2021; <https://wordnet.ailab.lv>



No

Visas valodas



Uz

Visas valodas

Nozare

Visas nozares



Rādīt detalizēti

Atrasts 1 rezultāts pieprasījumam "natural-language understanding"

149 līdzīgi termini

**EN natural-language understanding**

**EN natural-language comprehension**

**LV dabiskās valodas saprašana**

LZA TK ITTEA protokoli

**EN natural language**

**EN understanding**

**EN controlled natural language**

**EN natural language generation**

**EN quasi-natural language**

# EuroTermBank

**EuroTermBank.eu**

- ✓ [eurotermbank.com](http://eurotermbank.com) - single access point for terminology search
- ✓ Interlinked with other terminology banks and resources
- ✓ Aligning methodologies and standards
- ✓ Freely available under community license
- ✓ 680 collections in 45 languages, 3.9m entries, 14.5m terms



vispārīgs (6) apgūvēju (2) literārs (4) parlamentārs (1) diahronisks (1) specializēts (13) reprezentatīvs (7) paralēls (1)  
teksta (25) runas (5) morfoloģija (23) sintakse (3) semantika (1) kļūdas (2) manuāli pārbaudīts (4) tīmekļa (2)

### LVK2018

Līdzsvarotais mūsdienu latviešu valodas tekstu korpus  
2016–2018, 10 milj. vārdlietojumu (12 milj. tekstvienību)  
Izstrādātāji: LU MII

[Vairāk informācijas](#)

[Meklēt korpusā](#)

### MuLa2022

Mūsdienu latgaliešu tekstu korpus 2022  
2020–2022, 2 milj. vārdlietojumu (2,8 milj. tekstvienību)  
Izstrādātāji: RTA, LU MII

[Vairāk informācijas](#)

[Meklēt korpusā](#)

### LaVA

Latviešu valodas apgūvēju korpus  
2018–2021, 192 000 vārdlietojumu (241 000 tekstvienību)  
Izstrādātāji: LU MII

[Vairāk informācijas](#)

[Meklēt korpusā](#)

### LVTB

Latviešu valodas sintaktiski marķētais korpus  
2010–2022, 16 803 teikumi (282 167 tekstvienības) (v2.10)  
Izstrādātāji: LU MII

[Vairāk informācijas](#)

[Meklēt korpusā](#)

### Barometrs

Interneta agresivitātes indekss  
2011–2022, 26 milj. komentāru (642 milj. tekstvienību)  
Izstrādātāji: RSU, LU MII

[Vairāk informācijas](#)

[Meklēt korpusā](#)

### Disertācijas

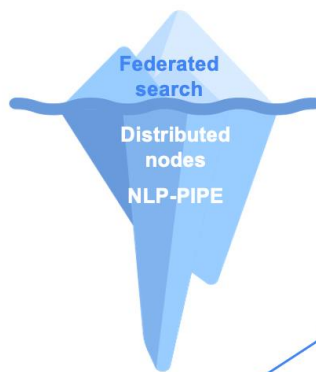
Promocijas darbu korpus  
2022, 16,7 milj. vārdlietojumu (23,4 milj. tekstvienību)  
Izstrādātāji: LU MII

[Vairāk informācijas](#)

[Meklēt korpusā](#)



# The Korpusss.lv platform



Developer: LU LFMI  
Node: nosketch.inb.lv

Developer: LU LaVI, LU MII, LU HZF

Node: nosketch.korpusss.lv

Developer: RTA, LU MII  
Node: nosketch.korpusss.lv

Nacionālā korpusu kolekcija **Meklēšana** Materiāli Par Korpusss.lv LV

sirds\*

Korpusss	Relatīvais biežums	Absolūtais biežums
	uz 1 miljonu	
<b>Pārspridumi</b> Skolēnu pārspridumu korpusss Izstrādātāji: LU MII, LiepU, RTA Mezglis: nosketch.korpusss.lv	2257	510
<b>LatSenRom</b> Latviešu senākie romāni Izstrādātāji: LNB Mezglis: nosketch.inb.lv	1420	8234
<b>Rainis</b> Raiņa darbu korpusss Izstrādātāji: LU MII Mezglis: nosketch.korpusss.lv	1196	2752
<b>IsprozaS</b> Sieviešu isproza Izstrādātāji: LU LFMI Mezglis: nosketch.inb.lv	1083	1272
<b>Senie</b> Latviešu valodas seno tekstu korpusss Izstrādātāji: LU LaVI, LU MII, LU HZF Mezglis: nosketch.korpusss.lv	765	2081
<b>LVMED</b> Radioloģisko izmeklējumu transkripciju korpusss Izstrādātāji: LU MII, RAKUS Mezglis: nosketch.korpusss.lv	616	97
<b>MuLa2012</b> Mūsdienu latgalešu tekstu korpusss 2012 Izstrādātāji: LU MII, RTA Mezglis: nosketch.korpusss.lv	426	565
<b>PanDi</b> Pandēmijas dienasgrāmatas Izstrādātāji: LU LFMI Mezglis: nosketch.inb.lv	408	289
<b>MuLa2022</b> Mūsdienu latgalešu tekstu korpusss 2022 Izstrādātāji: RTA, LU MII Mezglis: nosketch.korpusss.lv	394	1101
<b>Timeklis2020</b> Latviešu valodas timekļa korpusss Izstrādātāji: LU MII Mezglis: nosketch.korpusss.lv	334	164 504
<b>LiLa</b> Lietuviešu–latviešu–lietuviešu paralēlo tekstu korpusss Izstrādātāji: LU MII, VDU Mezglis: nosketch.inb.lv	305	1734
<b>LVK2018</b> Līdzsvarotais mūsdienu latviešu valodas tekstu korpusss Izstrādātāji: LU MII Mezglis: nosketch.korpusss.lv	270	3321
<b>Emuāri</b> Latviešu valodas emuāru korpusss Izstrādātāji: LU MII Mezglis: nosketch.korpusss.lv	256	2128

CONCORDANCE

simple sirds\* 1,272 (1,083.32 per million)

Details Left context KWIC Right context

11	doc#0	atīju tā caur savu augspīrību un aku mlēstību tāpat samaitā – par viņas	sirdi	nemaz, bet tikai par ārēju izskatu rūpēdamās. Kad Latiņa ko iedomājusēs, tu
12	doc#0	ni vien spriežot laiku pavadijām, un tu man nemaz tālāk nestāstīji, kas tavas	sirdsdrauga	un kādā vietā tas." "Viņš tagad Z. viesnīcā par sulaini," Anna stāstīja, "bet v
13	doc#0	raīga, kādā valodā vārds tiek nosaukts." "Tev gan taisnība, ka vārds cilvēka	sirdi	un dabu nepārvērs, un pa lielākai daļai tas jau no tēva mantots – jo tolaik, ke
14	doc#0	bas varētu būt arī villīgs – tik vien lūdzu nedomāt, ka tas zīmētos uz tavu	sirdsdraugu	. Tu jau gan, viņa sirdi un prātu pazīdama un no viņa krietnības pārliecināta,
15	doc#0	ik vien lūdzu nedomāt, ka tas zīmētos uz tavu sirdsdraugu. Tu jau gan, viņa	sirdi	un prātu pazīdama un no viņa krietnības pārliecināta, man par jaunu neper
16	doc#0	zar ģejiem." "Ne, ne, mīļā Trūtiņ!" sacīja Anna, Trūtes roku saķērusi un pie	sirds	spiezama. "Nevaru par to neko vairāk sacīt, – es tikai zinu, ka viņš ir skais
17	doc#0	– es tikai zinu, ka viņš ir skaists, laipns un patīkams, un ka es viņu no visas	sirds	mitu. Nekad neesmu ko dzījāki apdomājuse, kā tikai, cik jauki un laimīgi bij i
18	doc#1	i, kura veda uz ārpusi; le tas sev priekšā ieraudzīja kādu sievietes tēlu, viņa	sirds	sāka ātrāk pukstēt – vai tā varbūt nebija viņa? Šis slākaiss stāvs tam jau izī
19	doc#1	Jūsu abēju saruna, kuru nevilus dzirdēju, bija pilnens vairāk, kurš lika manai	sirdij	iet pāri. Es jūs negribēju apvainot, jo jūsu persona man sveša un vienaldzīg
20	doc#1	1, = nemandīnet mani, neļaujaties aizrautes no acumirkļa, nemodīnet mana	sirdi	jūtas, kuras es vairs nespētu apslāpēt. Es esmu dzīve tik daudz cietusi, tik c

Rows per page: 10 11–20 of 1,272

CONCORDANCE

simple sirds\* 2,081 (764.98 per million)

Details Left context KWIC Right context

31	doc#17	thas paβudde pŕelchan thems. Vnde Bātcy ftarpan Bōw / nhe dedtce mufta	Szyrdtzc	exkan mums / Khad thas ar mums trefley wūrβon tho Czele / Khad thas mun
32	doc#17	aridrtczan notkumfchenne / Beth es grib yums atkal redtzczeeth / vnde yuufe	Syrdtzc	Bōw lyxmotēs / vnd yuufe Lyxmbie bus neewenam no yums yempt. Vnde py
33	doc#17	Kur tu no efle ? Beth peetzc tho / ka es to vs yums trefleys efme / gir yuufe	Syrdtzc	puelina notkumfchenne taffyff. Beth es βacke yums tho tefcham. Thas yums
34	doc#17	unne Mhere dome es yums / nhe ka tha Pafoyle dode / dome es yums. Juufe	Szyrdtzc	nhe truczenas / vnde nhe yffytas Bōw. Jus efloth zirdfelfche / ka es yums
35	doc#17	tho loune / pekarretes tam labbam. Ta Bralige myfelfibe ftarpan wen otre gyr	Szyrdtzc	duebeniga. Wens nake tham otram ar Godenāfchenne pŕelchan. Nehe efle
36	doc#19	un abhāngig.wyrsch stahwu atkahru.Er stehet abhāngig. Tu ne dshwo wis no	firds	. tu atkahru ween dshwo.du arbeitest nicht vom herten, du ar-beitest hoye,
37	doc#19	lem Ewigen leben trachten ringen. Metaph.Sich erholen. Vid. ante fol. sub F.	Sirds	atustdas hertz erholst sich Akkotsdie Spitze von einer Korn Ähren Akkotjpsc
38	doc#19	iem Ewigen leben trachten ringen. Duhkβetrūbt, abgemattet seyn.Sadukkuŕfi	Sirds	-Ein βetrūbt, untraurigesCur. hertz. Duhzamaiss,der hindern, hinterste Duhmp
39	doc#19	rsrche, die im Korn schreyet Grestauff und überlegen,apgreest,ūmbkehren,	Sirds	apgreeschas GrestSchneiden,tkehren, wenden,Tas Nasiss atgreeschas,c
40	doc#19	peem tinnahs,er gehet mit Vieh un. Klukstehit,wie eine bruhth=henne glucken	Sirds	ta Klukst. KrustHagel Kruusene, ablu dsehrensApffel Trank Kuzzenesei

Rows per page: 10 31–40 of 2,081

CONCORDANCE

simple sirds\* 1,101 (933.93 per million)

Details Left context KWIC Right context

21	doc#44	is Kuorsovā, sevi skaitu arī kai varakļoniti un cylvāku no Aglyunys. Muna	sirds	Latgolā ir Varakļonūs, Aglyunā un dzymtājā Kuorsovā. Sovim bērnim kult
22	doc#50	lpa i tai tys vysy ar golā sauguja. Sirmselba, vinkuoršelba, gaidelšona, nu	sirds	nastys duovynys – Latgola otkon i dede muzikim jaunus breinimus, par k
23	doc#50	s dzimtyš likitjos. Pomans teik rakstruots kai lšokts na tik Latgolyš ļaužu	sirds	i dvēselis dzēivē, un arī Latvejīs viesituris krystpunktkūs – skalteituojam bys
24	doc#56	īmār pateik bērnim ruodeit, kai tei kasdīna var nūitk stuođojūt. Maņ suop	sirds	, ka mes drupeit par vālu asam saprotuši soidaukyūs školyš vajadzēibu. ņ
25	doc#57	, kū stuosteit; .β0. gođūs populāri beja balderiani, niule niivnam navajaga,	sirdsmāteri	padavai. Dažus augus poši padarejom populārus, pirmāram, lofantu – Kore
26	doc#70	sātu, a sastajam navarieju nūsacēit, kas valinis. Smird alkohols, dreizuoka	sirdsdarbelba	, a pats ni par kū nasasydz. Es pasaceju, ka jū vajag nūvārot. Vītu nav, i
27	doc#70	hit kū cytu. Interesantouk raksteit i skateit, atkļuojit nazynomos faktus. Nu	sirds	vieleju Osyuna pamatškolys nūvodpieticeibys ustoba dabūt Janinys atr
28	doc#70	busu atbrakušēit, eipāši baļtinavīšis, sacējuši: „Es pi jums atjuoju.“ Jai suop	sirds	, ka Osyuna vaira nazynūts, ka pareizais, eisti latgaliskais ituos vītyš nūsai
29	doc#75	vede juos izšyvyumus. Par tū jei lelu nauđu saņēme. Pats Dymša nūmyra ar	sirdstriku	. Ar sunēiti izgojua pasastaigaut, izskreja leļi suni. Cikom jis sovu kuceiti g
30	doc#75	te atguojuse da babenis atsavadeit. A kungs šylymā teinurēpa nūmyra ar	sirdstriku	. Naspēja puordzeivout, ka juopamat Rušyuna.“ Rušyuns luocs Antoneja:

Rows per page: 10 31–40 of 1,101

- Šobrīd NKK ir **30 korpusi**, to izstrādē piedalījušās **13 institūcijas**.
- Lielākā daļa NKK korpusu ir gramatiski marķēti un iekļauti vienotās meklēšanas sistēmā.
- Vienotā meklēšana iespējama **24 korpusos** (2,1 miljardi tekstvienību).

# Mūsdienu latgaliešu runas korpuss

MuLaR

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

**Query:** *[word='kaķis'%c]*

**Number of results:** 0



**Recording**

**Transcription**

▶ 0:00 / 0:00 — 🔊 ⋮



*F216-xxxx:* kai tev pasceit nu skrytuleits itaids vo pastateits špūli lik taidi vēļ div pagalis vot itaidys i nu šenis ▶ skrytuleiti ībuoz vydā tai vo tymā skrytulī vot īlaiž vēļ dzeļža taidu īlaiž i iz četru kuojeņu tys ratineš ▶

*130-M:* aha aha da ▶



Hugo.lv

## Hugo.lv paralēlie korpusi

teksta (25)

vispārīgs (6)

---

Korpusa apjoms 10,5 milj. vārdlietojumu

---

Izstrādes periods 2018

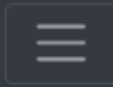
---

Izstrādātāji Kultūras informācijas sistēmu centrs



Mākslīgā intelekta laboratorija | [info@korpuss.lv](mailto:info@korpuss.lv)

Latvijas Universitātes Matemātikas un informātikas institūts



# Hugo.lv

## Hugo.lv par

teksta (25)

vispārīgā

Korpusa apjoms

Izstrādes periods

Izstrādātāji

Valodu rīki ▾ Par Hugo Palīdzība

Sākums / Meklēšana korpusā

## Meklēšana korpusā

Atslēgvārds



Wikipedia.com.lv-ru

Culture

1698720 Vārdi

fold.lv.en-lv

Culture

468137 Vārdi

OpenSubtitles2016.en-lv

Culture

5925977 Vārdi

CultureWebSites.lv-ru

Culture

909485 Vārdi

Wikipedia.com.en-lv

Culture

3276964 Vārdi

OpenSubtitles2018.lv-ru

Culture

4624795 Vārdi

LatviaWebSitesOpen.lv-ru

General

2434817 Vārdi

Latviešu ▾

Valoda

Meklēt




Mākslīgā intelekta laboratorija | Latvijas Universitātes Matemātik

<https://hugo.lv/lv/CorpusSearch>

# Meklēšana korpusā

Atslēgvārds


 

Korpuss

Valoda

Meklēt

Meklēšanas rezultāti: 58  [Lejupielādēt](#)

 [Pārslēgties uz rindu skatu](#) **Lapas garums**

200 

15 . Palielinās bērnudārzu grupu skaits , kurās mācību	valoda	ir latviešu
as domei saziņā ar iedzīvotājiem sociālajos tīklos būtu jāizmanto arī krievu	valoda	.
Bet tur ir ļoti daudz tūristu un angļu	valoda	tur ir dzirdama biežāk , nekā latviešu .
Darba	valoda	mums projektā ir angļu , bet veikalos un daudz kur citur , man vienmēr bija
Diemžēl pietrūkst laika mācībām , turklāt angļu	valoda	vietējiem ir tik augstā līmenī , ka motivācija un nepieciešamība apgūt valsts
Domāju , ka	valoda	ir ļoti sarežģīta – gramatika pilnīgi nelīdzinās angļu valodai .
Dzīvojam šeit , tāpēc jāmacās	valoda	.
Emīlija un Alija mācās parastā skolā , tatāru	valoda	viņas vēl nepārvalda ļoti labi , bet ar savu izcelsmi viņas ļoti lepojas un reizi
saprast rakstītu latviešu valodu , piemēram , kad lasu restorāna ēdienkarti ,	valoda	nedaudz atgādina zviedru , somu un vācu valodu .

# Paralēlie korpusi

Filter by:

Language

- Latvian (137)
- English (79)
- German (31)
- French (30)
- Polish (29)

more

Resource Type

Media Type

Licence

Conditions of Use

Linguality Type

Multilinguality Type

Data Format

Domain





Appropriateness For DSI

Funding Project

137 Language Resources (Page 1 of 7)

« Previous | Next »

Order by: Resource Name A-Z

 <b>Anonymised ParaCrawl release 7 Latvian-English</b>	↓ 0	👁 22
English   Latvian	CC0-1.0	
 <b>Anonymised ParaCrawl release 8 Latvian-English</b>	↓ 0	👁 10
English   Latvian	CC0-1.0	
 <b>Anonymised ParaCrawl release 9 English-Latvian</b>	↓ 0	👁 12
English   Latvian	CC0-1.0	
<b>ab</b> <b>Arheologia terminibaas</b>	↓ 0	👁 8
Dutch; Flemish   English   Estonian   Finnish   German   Latvian   Polish   Russian   Swedish	CC-BY-4.0	
 <b>Bilingual corpus from the European Vaccination Portal (LV-EN)</b>	↓ 17	👁 45
English   Latvian	CC-BY-4.0	

# Parlamentu datu korpuss

Saeimas stenogrammas [Meklēšana](#)

## Par vietni

Šī vietne ir izveidota, lai atvieglotu darbu žurnālistiem, politologiem un citiem, kas interesējas par Saeimas darbu un deputātu viedokļiem. Ir izveidots Saeimas stenogrammu 11. Saeimai ieskaitot) korpuss, kas ļauj meklēt gan vienlaicīgi visās Saeimas stenogrammas, gan arī konkrētās sēdēs, laika posmos un pat konkrētu deputātu runā.

Lai darbu atvieglotu, katrai konkordancei ir norādīts to teikušais deputāts, viņa dzimšanas gads, frakcija, kurā deputāts ir bijis, kā arī sēdes laiks un tips. Noklikšķinot uz re, parādīts plašāks konteksts.


Stenogrammas iegūtas no [Saeima.lv](#)

## Daži meklējumu piemēri:

### „no zāles”

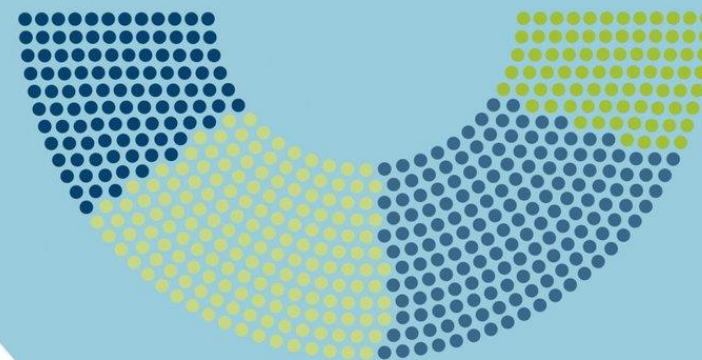
...Par - 77, pret un atturas - nav. Likumprojekts par steidzamu atzīti. (	<b>No zāles</b>	dep. A.Bērziņš: "Veiksmīgs skaitlis!") Balsosim par likumprojekta "Gr...
...u svētkiem piešķirt brīvdienu. Tāpēc komisijā tas netika atbalstīts. (	<b>No zāles</b>	dep. Dz.Ābiķis: "Lauku sievietes, kas dejo un dzied!")...
...4, pret un atturas - nav. Likumprojekts pirmajā lasījumā atbalstīts. (	<b>No zāles</b>	dep. J.Ādamsons: "Uz Gruziju!" Lūdzu noteikt priekšlikumu iesniegšana.
 Augusts Brigmanis (1952) ZZS 2014/10/02 kārtējā sēde		rlietu komisijas vārdā - deputāts Ojā...
...kos. Pēc šīs sēdes tātad... Nu, pareizi Čepānes kundze mani palabojā (	<b>No zāles</b>	dep. I.Čepāne: "Es neko neteicu!", ka tātad dalībniekiem paši Dziesm...
...im par atzīmējamu dienu. Tātad - vīrieši, atcerieties par sievietēm! (	<b>No zāles</b>	dep. Dz.Ābiķis: "Dziersim katru dienu!")...
...umprojektā, bet es šo priekšlikumu neatbalstīšu, godātie vīrieši, jo (	<b>No zāles</b>	deputāti vīrieši: "Āāā...!" Zālē smieklī.) es tiešām uzskatu: mums šo...
... dienu. Un šeit nav runa par 23.novembri, kā mēs runājam... (Aplausi.	<b>No zāles</b>	: "Ooo!" Šeit nav runa... Es vēlreiz uzsveru: šeit nav runa, es atvai...

### aizver muti

 Ingmārs Līdaka (1966)  
ZZS  
2009/10/15 kārtējā sēde

... . Katrā valstī valsts Neatkarības diena ir svinamā... (Dep. V.Buzajeva starpsauciens no zāles.) **Aizver muti!**... Katrā valstī Neatkarības diena ir svinamā diena! Tātad - ko Kabanova kungs domāja ar 11.....

# ParlaMint



# ParlaMint: UD-annotated comparable corpora



ParlaMint-LV 2.1 (Latvian parliament)

Query **NOUN, iobj, VERB, palīdzēt** 618 (75.61 per million) **i**

Page  of 31  [Next](#) | [Last](#)

2016-06-16,ZariņšIvars	tirgu , un tieši mēs palīdzējām ar savu	<b>kompetenci</b>	. Tas pats ir ar gāzes tirgu , par kuru mums būs
2019-12-19,StepaņenkoJūlija	, kas atrisinās šo problēmu un palīdzēs tiem	<b>cilvēkiem</b>	, kas ir savākuši šos parakstus ... vairāk nekā 50
2016-12-15,ŠicsMārtiņš	mērķis ir palīdzēt arī tiem otrās vai trešās grupas	<b>invalidiem</b>	, kuri var piestrādāt atbilstoši saglabātajām
2016-12-15,KirsisVilnis	šī riska kapitālista kompetence un spēja palīdzēt	<b>produktam</b>	nonākt globālajā tirgū . Primāri svarīgas ir šīs
2017-02-09,Bitelnga	par kopīgu eksportu , lai palīdzētu Latvijas	<b>uzņēmējiem</b>	atrast noieta tirgus savai produkcijai Kanādā .
2020-10-15,DombrovskisVjačeslavs	to PVN samazinātā likme kā instruments var palīdzēt	<b>nozārēm</b>	, tā vairs nav nekāda svētā govs ... ko nekādā veidā
2020-10-15,DombrovskisVjačeslavs	uzskatīja , ka tas var palīdzēt augļu un dārzeņu	<b>nozarei</b>	palielināt apgrozījumu , samazināt ēnu ekonomiku
2020-10-15,Reizniece-OzolaDana	nodokli , mēs faktiski palīdzam mazāk turīgajiem	<b>iedzīvotājiem</b>	. Tas Latvijā ir aktuāli , jo , kā mēs zinām , mums ir
2020-10-15,LiepiņaLiene	, ka šīs izmaiņas ir ļoti būtiskas , lai veselai	<b>virknei</b>	uzņēmēju palīdzētu iznākt ārā no pelēkās zonas .
2020-10-15,StepaņenkoJūlija	saprast , vai tiešām šis samazinātais PVN palīdzēs	<b>uzņēmējiem</b>	. Tas bija galvenais jautājums ; mēs nerunājam
2020-10-15,DobravaJānis	stimulējošie pasākumi , kā mēs varam palīdzēt	<b>sabiedrībai</b>	pārbūvēt esošās automašīnas par videi
2020-09-10,ZariņšIvars	no medijiem dzirdējuši ... ka no sākuma palīdzējām <b>[PER:</b>	<b>Lukašenko</b>	<b>]</b> , lai viņš varētu izrēķināties ar savu opozīciju .
2020-09-10,GobzemsAldis	bet izlietota ! - tādā veidā , lai tas palīdzētu visai	<b>sabiedrībai</b>	un Latvijas konkurētspējai ... Latvijas spējai
2019-01-31,Ločmele-LuņovaRegīna	, ka tas varētu arī rūdīt raksturu un palīdzētu	<b>vīrietim</b>	kļūt vēl vīrišķīgākam , - mazliet paciesties .
2019-01-31,ZariņšIvars	tribīnes esmu apsollījis , es palīdzēju ekonomikas	<b>ministram</b>	, ja viņš vēlēties tikt galā ar šo afēru . Arī tad , ja
2019-01-31,ZariņšIvars	jautājumi ir konkrēti , lai palīdzētu ekonomikas	<b>ministram</b>	saprast , kas tad viņam būtu jāpajautā saviem
2020-10-29,StepaņenkoJūlija	ļoti jāsteidzas sodīt cilvēkus , bet palīdzēt šiem	<b>cilvēkiem</b>	... nu , nevajag steigties . Nu tāda ir loģika . Es
2020-10-28,KariņšArtursKrišjānis	, kuri ar savu piemēru un argumentāciju arī palīdz	<b>sabiedrībai</b>	izvairīties no slimības izplatības un faktiski
2020-10-28,KariņšArtursKrišjānis	ka valdības ekonomiskais stimulants var palīdzēt mūsu	<b>ekonomikai</b>	būt izturīgai šajos grūtajos apstākļos . Mēs
2020-10-28,KariņšArtursKrišjānis	bīstama , kā to vispār varētu ārstēt un kā palīdzēt	<b>pacientiem</b>	... valdībā mēs pieņemām lēmumu - un paldies , ka jūs ,

CQL query:

```
[pos="NOUN"  
& dep="iobj"  
& dep_head_pos="VERB"  
& dep_head_lemma=  
"palīdzēt"]
```

Erjavec, T., Ogrodniczuk, M., ..., Dargis, R., ..., Krilavičius, T., et al. [ParlaMint: Comparable Corpora of European Parliamentary Data](#). Proceedings of the CLARIN Annual Conference, 2021

<https://www.clarin.eu/content/parlamint-towards-comparable-parliamentary-corpora>





**TEHNOLOĢIJAS UN RĪKI**

# NLP-PIPE: Latvian NLP Pipeline as a Service

Select a sample text

2022. gada 30. novembrī Rīgā notiek seminārs "Latviešu valoda digitālā Eiropā", kurā Kristofers Skots no Tulkošanas ģenerāldirektorāta (DGT) iepazīstina ar eTranslation platformu.



Go

tokenizer x

ner x

morpho x

parser x

Visual  CoNLL  JSON

2022. gada 30. novembrī time Rīgā GPE notiek seminārs "Latviešu valoda digitālā Eiropā" event, kurā Kristofers Skots person no Tulkošanas ģenerāldirektorāta organization (DGT organization) iepazīstina ar eTranslation platformu entity.

<http://nlp.ailab.lv>

INDEX	FORM	LEMMA	UPOSTAG	XPOSTAG	FEATS	HEAD	DEPREL
#text=2022. gada 30. novembrī Rīgā notiek seminārs "Latviešu valoda digitālā Eiropā", kurā Kristofers Skots no Tulkošanas ģenerāldirektorāta ( DGT ) iepazīstina ar eTranslation platformu .							
1	2022.	2022.	ADJ	xo	NumType=Ord ← ▶	2	amod
2	gada	gads	NOUN	ncmsg1	Case=Gen Gender ← ▶	4	nmod
3	30.	30.	ADJ	xo	NumType=Ord ← ▶	4	amod
4	novembrī	novembris	NOUN	ncmsl2	Case=Loc Gender ← ▶	6	obl
5	Rīgā	Rīga	PROPN	npfsl4	Case=Loc Gender ← ▶	6	obl
6	notiek	notikt	VERB	vmnipi130an	Evident=Fh Mood= ← ▶	0	root
7	seminārs	seminārs	NOUN	ncmsn1	Case=Nom Gende ← ▶	10	nmod
8	"	"	PUNCT	zq	- ← ▶	10	punct
9	Latviešu	latvietis	NOUN	ncmpg2	Case=Gen Gender ← ▶	10	nmod
10	valoda	valoda	NOUN	ncfsn4	Case=Nom Gende ← ▶	6	nsubj
11	digitālā	digitāls	ADJ	affslnp	Case=Loc Dgree= ← ▶	12	amod
12	Eiropā	Eiropa	PROPN	npfsl4	Case=Loc Gender ← ▶	10	acl
13	"	"	PUNCT	zq	- ← ▶	10	punct
14	,	,	PUNCT	zc	- ← ▶	24	punct
15	kurā	kurš	PRON	pr0msln	Case=Loc Gender ← ▶	24	obl
16	Kristofers	Kristofers	PROPN	npmsn1	Case=Nom Gende ← ▶	24	nmod
17	Skots	skots	NOUN	ncmsn1	Case=Nom Gende ← ▶	16	flat.name
18	no	no	ADP	spsg	- ← ▶	20	case
19	Tulkošanas	tulkošana	NOUN	ncfsg4	Case=Gen Gender ← ▶	20	nmod
20	ģenerāldirektorāta	ģenerāldirektorāts	NOUN	ncmsg1	Case=Gen Gender ← ▶	16	nmod

# WMT2017-2019

Trīs gadus pēc kārtas ziņu tulkošanas uzdevumā starp angļu valodu un Baltijas valstu valodām (latviešu, [lietuviešu](#) un [igauņu](#)) sabiedrība “Tilde” uzrādījusi dalītus labākos rezultātus prestižajās mašīntulkošanas darbsemināra (*Workshop on Statistical Machine Translation*; kopš 2016. gada – *Conference on Machine Translation*) rīkotajās sacensībās (*shared task*).



# EU PRESIDENCY

## Challenges:

- Facilitating multilingual communication in Europe
- Massive amounts of official documentation & texts
- Fast communication of the Presidency work & activities
- Delegates, Journalists and Presidency staff synergies
- Challenging languages

## Solutions:

- pan-European MT infrastructure
- Customized NMT to the Presidency requirements
- Integration of relevant in-domain data & terminology
- Workspace integration of NMT systems





## Runas tehnoloģijas

**Runas atpazinējs** nodrošina runas pārvēršanu tekstā no iepriekš ierakstīta audiofaila vai nodiktēta teksta.

**Runas sintēzes tehnoloģijas** sniedz iespēju rakstītu tekstu izrunāt balsī.

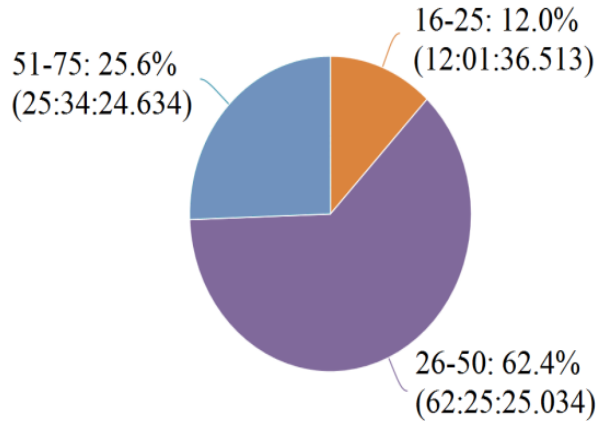


# Runātās valodas korpuss

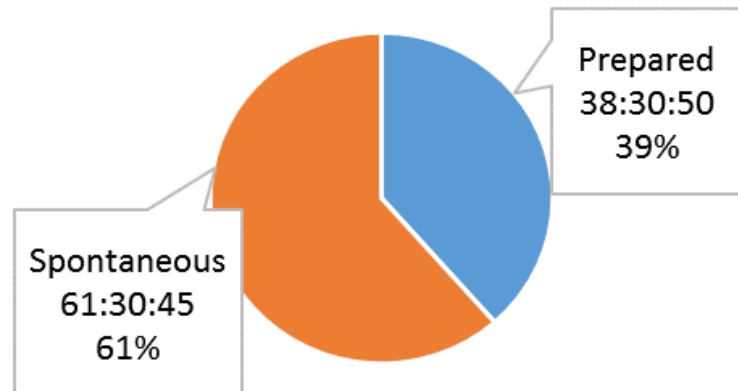
Ortogrāfiski marķēts latviešu valodas runas korpuss:

Ortogrāfiski marķēti dati (100 stundas)

Fonētiski marķēti dati (4 stundas)

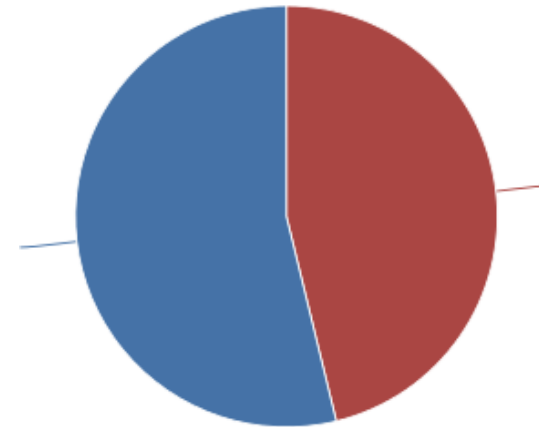


*Data distribution with respect to the age of speakers*



*Data distribution with respect to the gender of speakers*

Men: 53.8%  
(53:49:13.904)



Women: 46.2%  
(46:12:12.277)

**Duration of segmented recordings (hh:mm)**

	Total dur of recordings	Word- and segment
Dial.	76:41	74:13
Monol.	12:51	9:05
Total	89:32	83:19

# Izmēģini Tildes runas tehnoloģijas

Izrunāt tekstu

Diktēšana

Ievadi tekstu, lai noklausītos balss sintēzes tehnoloģijas demonstrāciju. Atceries, ka šo pakalpojumu var lietot tikai atbilstoši tā lietošanas noteikumiem (skatīt sadaļu "Lietošanas noteikumi")! To nav atļauts izmantot tekstiem, kas satur aizskarošus, nepieklājīgus, apmelojošus, maldinošus, naidu vai diskrimināciju veicinošus apgalvojumus. Par lietošanas noteikumu pārkāpumu var iestāties LR likumos noteiktā atbildība.



- Pašlaik runas atpazīšanas tehnoloģijas var atpazīt apmēram 80 % no ierunātā, ja audioieraksta kvalitāte ir pietiekami laba, un sasniedz pat 94 % precizitāti specializētos lietojumos.
- Runas atpazīšanā tiek pētīti un veidoti specializēti risinājumi tādām jomām kā [medicīna](#), telekomunikācijas, klientu apkalpošana u. c.

# RUTA:MED

## medicīnisko diktātu pusautomātiska transkribēšana

Close ASR Preview

MR prostatai ar kontrastvielu prostata trīs komats deviņi reiz trīs komats trīs reiz trīs komats astoņi centimetri tilpums divdesmit seši komats piecpadsmīt kubikcentimetri PSA līmenis nav zināms

dziedzera pārejas daļā redzami

<unk> labdabīgas hiperplāzijas mezgli

urīnpūslis gludām sieniņām bez intralūmenāliem ieslēgumiem patoloģiska lieluma limfmezglus nesaskatu un rakstam lūdzu slēdzienu labdabīga prostatas hiperplāzija

pašreiz nav datu par augstas malignitātes tumora audiem



00:01:10 / 00:01:32

Verify

Edit

Submit to TC

Download

⋮

MR prostatai ar k/v.

Prostata 3,9 x 3,3 x 3,8 cm, tilpums 26,15 cm<sup>3</sup>.

PSA līmenis nav zināms.

Dziedzera pārejas daļā redzami labdabīgas hiperplāzijas mezgli.

Urīnpūslis gludām sieniņām, bez intralūmenāliem ieslēgumiem.

Patoloģiska lieluma l/m nesaskatu.

### Slēdziens

Labdabīga prostatas hiperplāzija.

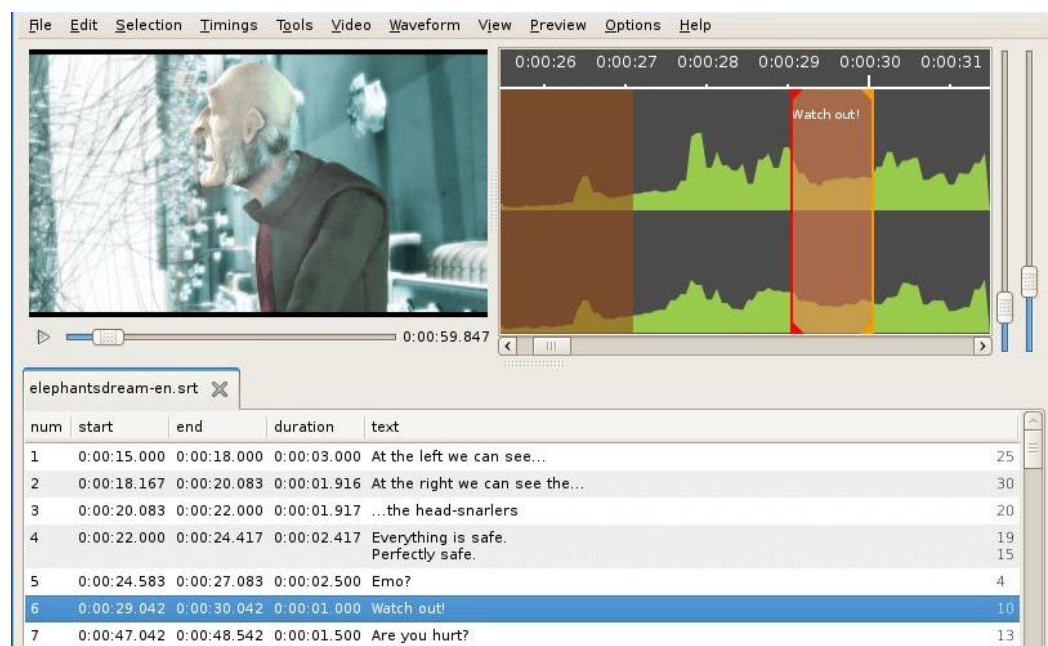
Pašreiz nav datu par augstas malignitātes tumora audiem



00:01:10 / 00:01:32



# Automatizēta subtitru veidošana un subtitru mašīntulkošana



The screenshot shows a video editing software interface. The top menu bar includes File, Edit, Selection, Timings, Tools, Video, Waveform, View, Preview, Options, and Help. The main window is divided into a video preview area on the left and a waveform area on the right. The video preview shows a man speaking. The waveform area displays audio levels with a red vertical line indicating a specific time point. Below the video player, a subtitle list is visible for the file 'elephantsdream-en.srt'. The list contains the following entries:

num	start	end	duration	text	
1	0:00:15.000	0:00:18.000	0:00:03.000	At the left we can see...	25
2	0:00:18.167	0:00:20.083	0:00:01.916	At the right we can see the...	30
3	0:00:20.083	0:00:22.000	0:00:01.917	...the head-snarlers	20
4	0:00:22.000	0:00:24.417	0:00:02.417	Everything is safe. Perfectly safe.	19 15
5	0:00:24.583	0:00:27.083	0:00:02.500	Emo?	4
6	0:00:29.042	0:00:30.042	0:00:01.000	Watch out!	10
7	0:00:47.042	0:00:48.542	0:00:01.500	Are you hurt?	13



Reālā laika runas atpazīšana festivālā «Lampa»

# Virtuālie sarunu biedri



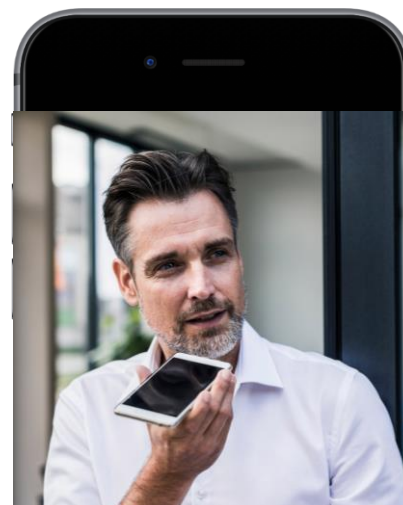
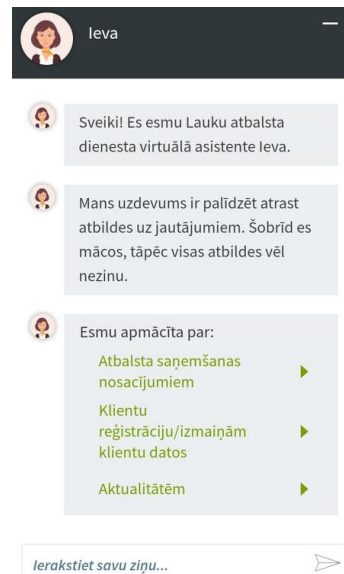
Latviski runājošie virtuālie sarunu biedri “strādā” valsts pārvaldes iestādēs, tiek izmantoti izglītībā, uzziņām un lietotāja atbalsta servisos



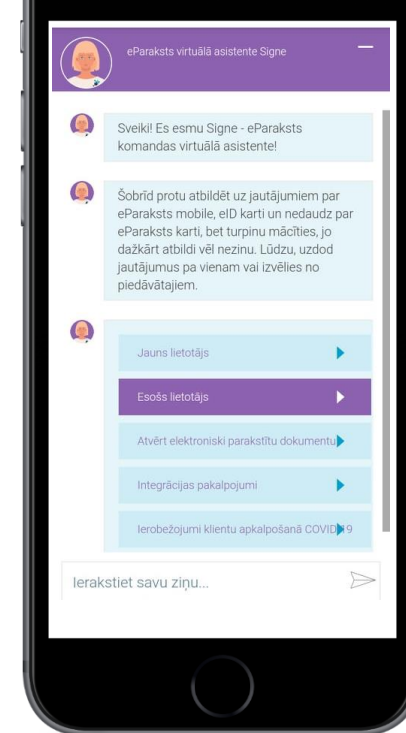
**110+** organizācijas izmanto virtuālos sarunu biedrus Latvijā



**5 000+** cilvēku komunicē ar Tildes VA ik dienu



**Jaunums –  
telefonijas bots**



Virtuālais asistents informācijai par palīdzības saņemšanu un sniegšanu Ukrainai

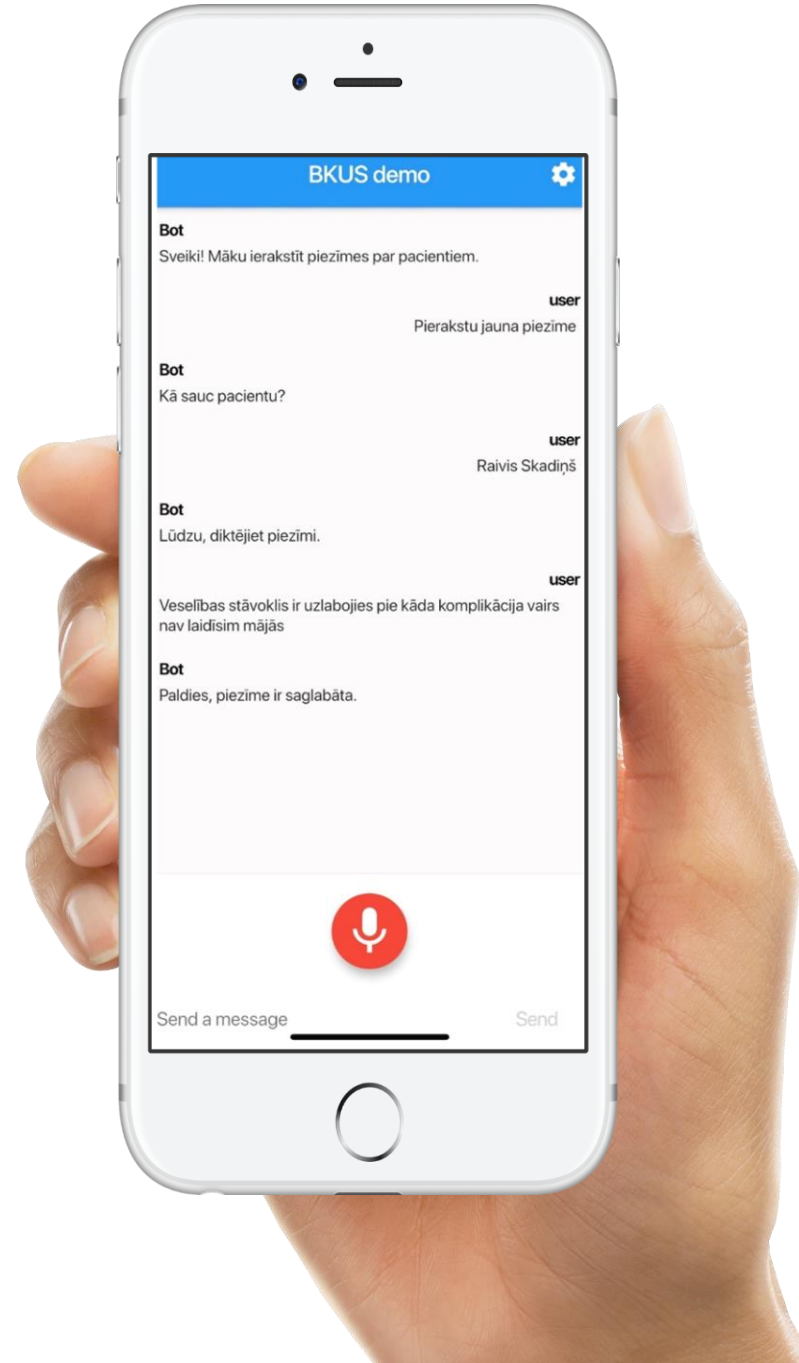


Use Case

# Doctors' virtual assistant for Children Clinical University Hospital



Bērnu klīniskā  
universitātes  
slimnīca





Literatūras darbi

Mācies interaktīvi

Virtuālie asistenti

Pedagogiem

letonika.lv

# Kauja pie Knipskas



Kas ir virtuālais asistents?

Cibiņš

Skroderiene

Zvejnieks

Uzmini žanru vai darbu!

**Autors:**

Jānis Poruks

[skatīt autora biogrāfiju](#)

**Pirmpublicējums:**

1897

**Literatūras veids:**

epika jeb proza

[skatīt literatūras veidus](#)

**Literatūras žanrs:**

stāsts

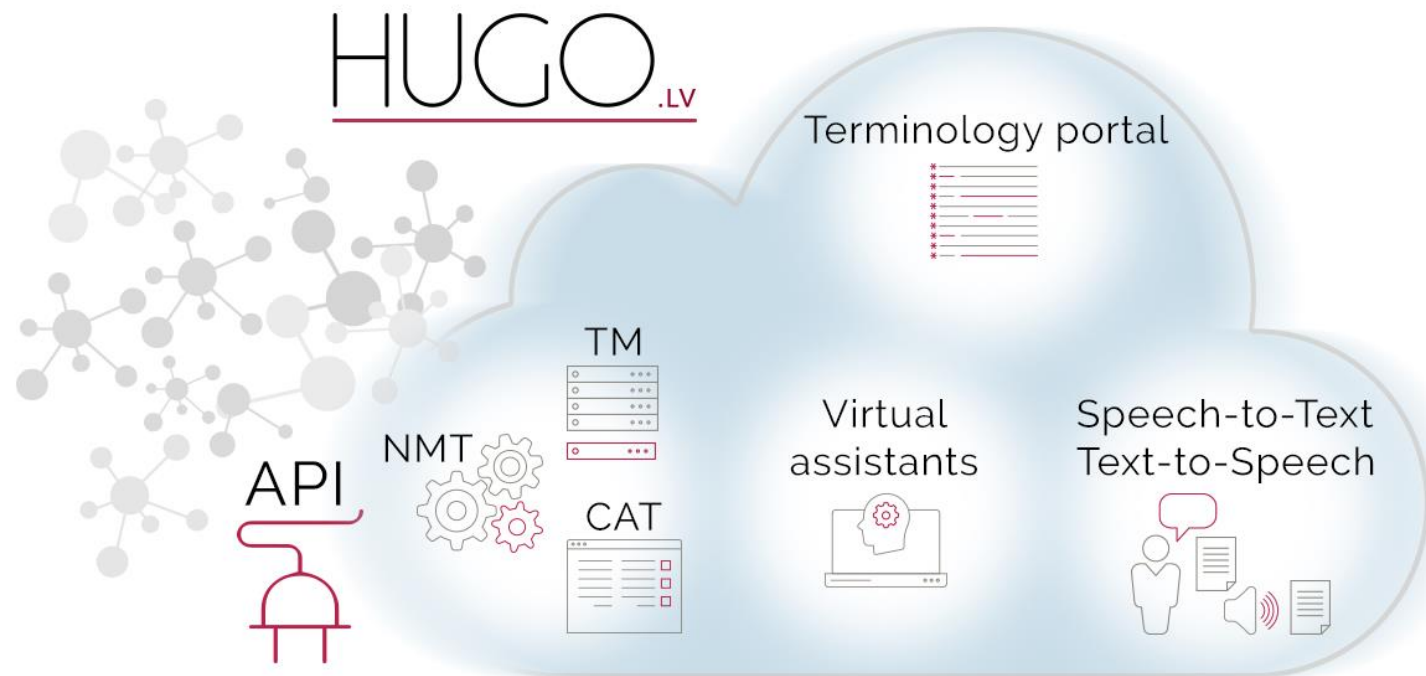


Tikko

Ierakstiet ziņojumu



# HUGO.LV – Latvijas valsts pārvaldes valodas tehnoloģiju platforma



**20+**

**Virtuālie asistenti**

nodrošina nepārtrauktu komunikāciju ar klientiem

**4,000+**

**cilvēku**

Satunājas ar virtuālajiem sarunu biedriem ik dienu

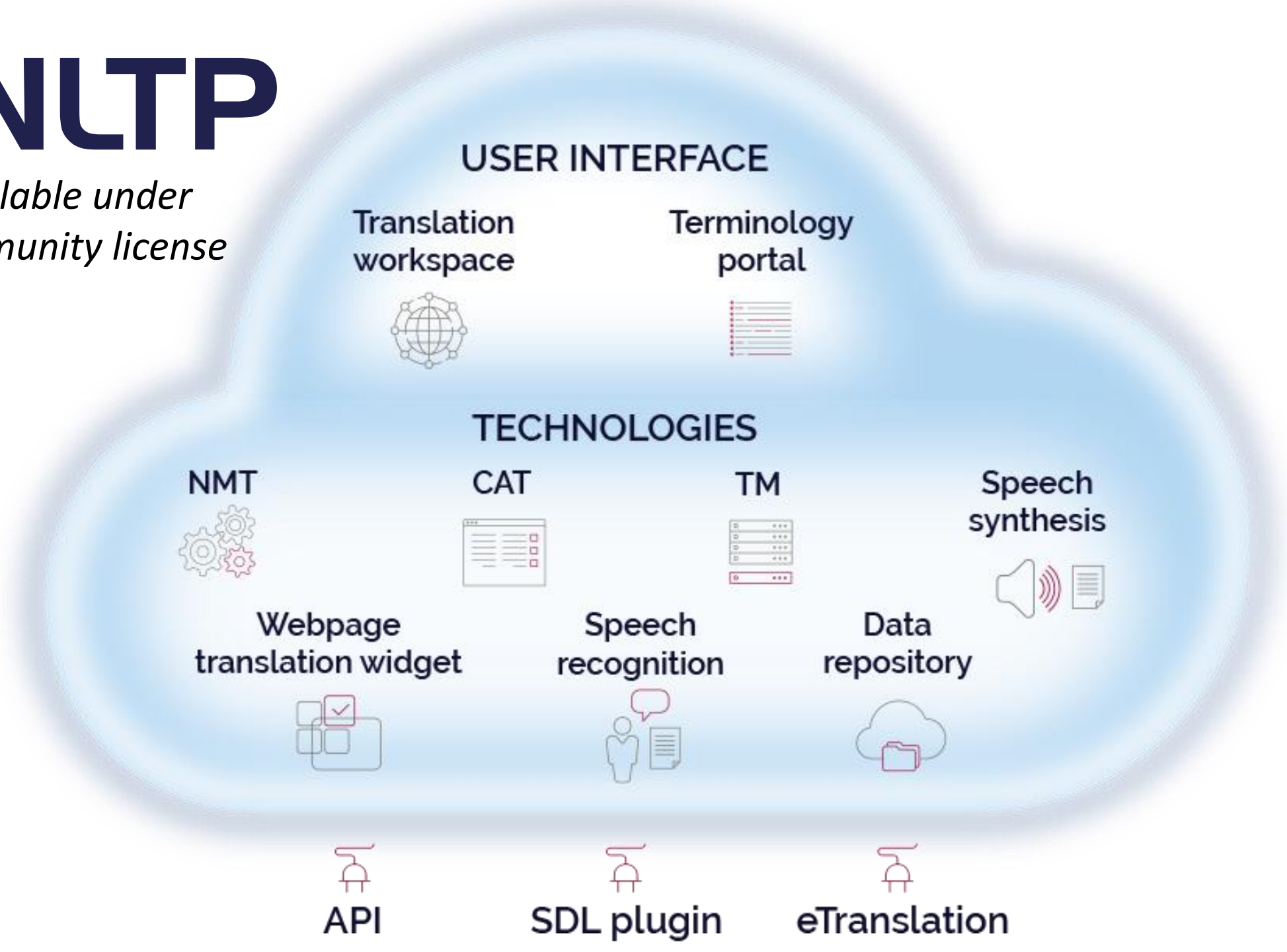
**12,000+**

**ziņu**

katru dienu nosūta Tildes čatboti visā Baltijā

# NLTP

*Freely available under  
open community license*



# CLARIN-LV valodas resursu un rīku repozitorijs

## Kas jauns

### ToolService

CLARIN Centre Of Latvian Language Resources And Tools

#### RUTA:MED – Dual Workflow Medical Speech Transcription Pipeline and Editor



##### Autors (-i):

Znotiņš, Artūrs ; Darģis, Roberts ; Grūzītis, Normunds ; Bārzdīņš, Guntis and Goško, Didzis

##### Apraksts:

In the medical domain, various approaches are used to produce examination reports and other medical records. Depending on the language-specific technology support, the type of examination, the size of the hospital or clinic, ...

Šajā vienumā nav failu.

### Corpus

CLARIN Centre Of Latvian Language Resources And Tools

#### Corpus of Latvian Women Writers' Short Fiction



##### Autors (-i):

Kārkla, Zīta and Matulis, Haralds

##### Apraksts:

The corpus consists of short fiction by Latvian women writers published from 1893 to 2002.

Šajā vienumā nav failu.

### LexicalConceptualResource

CLARIN Centre Of Latvian Language Resources And Tools

#### LVMED: Latvian Pronunciation Dictionary of the Medical Domain



##### Autors (-i):

Darģis, Roberts ; et al.

► show everyone

##### Apraksts:

A machine-readable pronunciation dictionary of the medical domain derived from a large text corpus of historical medical records. Consists of 109k entries in the CSV format: first column - a wordform; second column - its ...

Šajā vienumā ir 1 fails (665.59 KB).

Publicly Available

🔍 Pārlūkot

> Visi repozitoriji

👤 Mans kots

➡ Pieteikšanās

📄 Vispārīga informācija

📄 Deponēt

🗨 Citēt

🔄 Iesniegšanas dzīves cikls

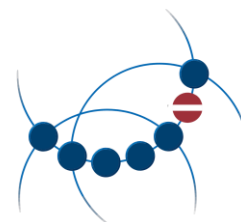
? Bieži uzdotie jautājumi

📢 Par

✉ Palīdzības dienests

📡 RSS Feed

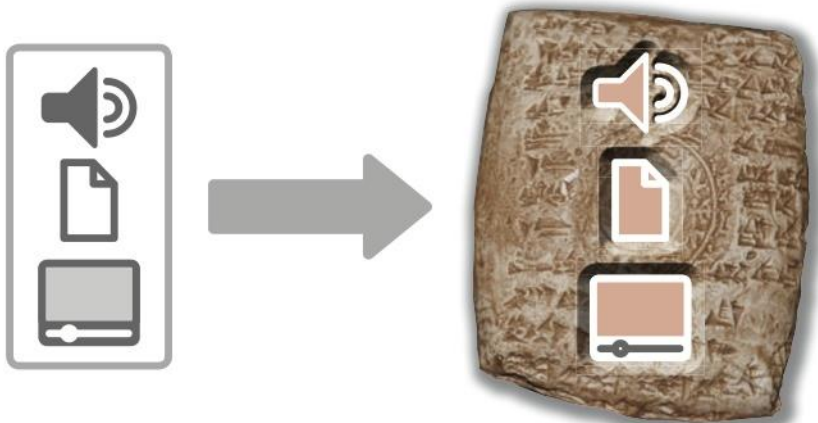
CLARIN-LV



<https://repository.clarin.lv/>

# Atbalsts atvērtai zinātnei: datu uzglabāšanas servisi

Viens no CLARIN pamatzdevumiem ir nodrošināt, ka valodas resursi tiek arhivēti, tie ir **pieejami** un tie ir **uzticami**. Tāpēc CLARIN centri piedāvā valodas resursu uzglabāšanas (deponēšanas) servisu.



- **F** – *findability*: existence at a data **provider**, discovery, **persistence**, long-term **preservation**
- **A** – *accessibility*: **open** data / access
- **I** – *interoperability*: common **terminology**, data **models** and **formats** (for both humans & machines)
- **R** – *reusability*: content, **documentation**, provenance, **versioning**, licencing



# Valodu digitālās līdztiesības metrika

## Technological Factors

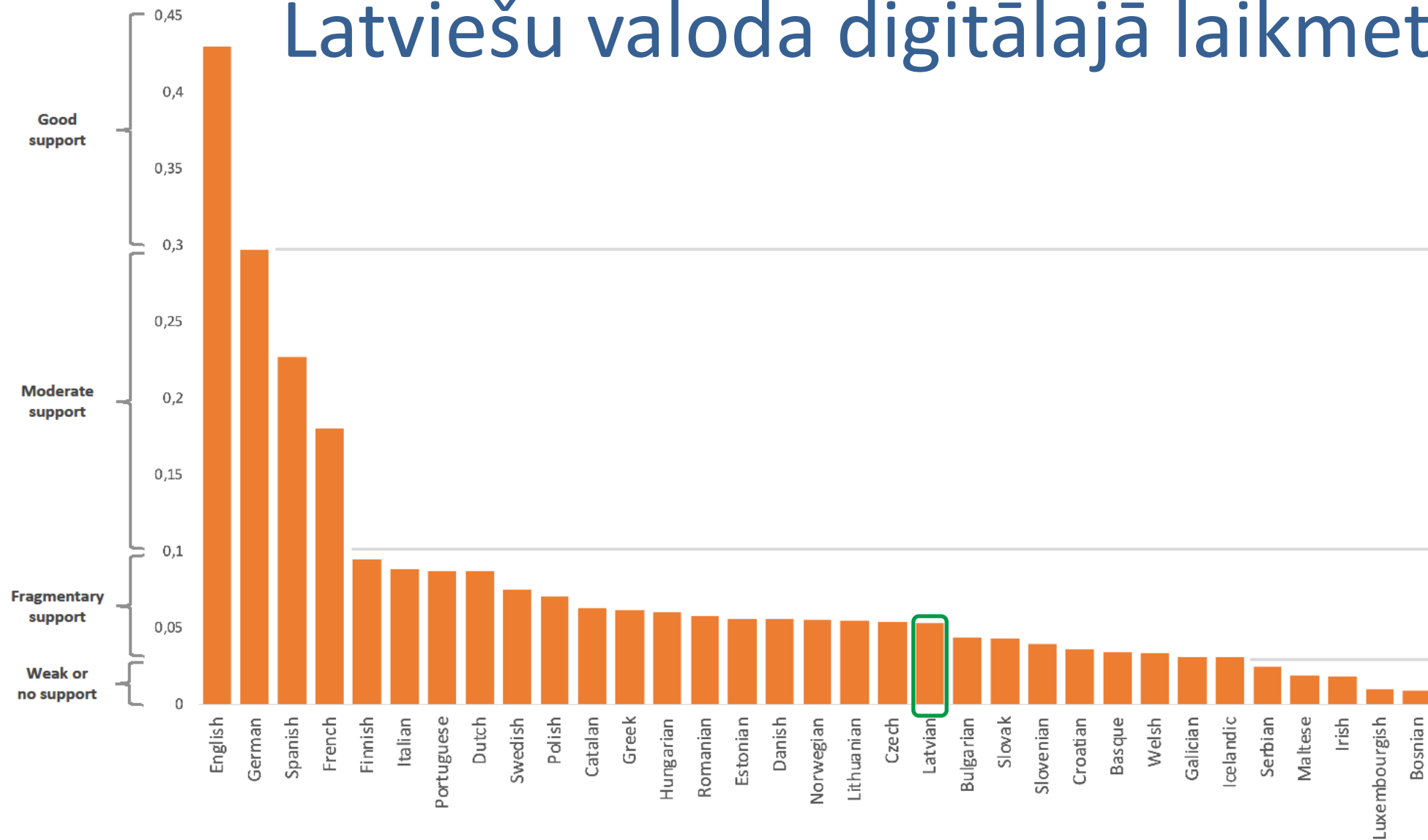
- Tools and Services
- Corpora
- Language Descriptions and Models
- Lexical and Conceptual Resources
- Projects
- Organisations



## Contextual Factors

- Industry
- Online
- Research & Development & Innovation
- Technology
- Economy
- Education
- Funding
- Law
- Media
- Policy
- Public administration
- Society

# Latviešu valoda digitālajā laikmetā: 2022



# Secinājumi

- Digitālajā vidē latviešu valodas pozīcijas pēdējā desmitgadē ir būtiski uzlabojušās. Radīti jauni resursi un rīki rakstītās un runātas valodas apstrādei, cilvēka-datora saziņai un mašīntulkošanai.
- Kaut gan kopš META-NET Balto grāmatu publicēšanas 2012. gadā latviešu valodas resursi un rīki ir ievērojami uzlabojušies, tomēr daudz **plašāk lietotām valodām** (piem., angļu, franču vai vācu) **tehnoloģiskais nodrošinājums** kopumā vēl arvien **ir ievērojami attīstītāks**.
- Lai latviešu valoda arī turpmāk būtu dzīvotspējīga attīstītā digitālajā pasaulē, tai **jābūt pieejamiem atbilstošiem resursiem un tehnoloģiskajiem risinājumiem**.
- Tikai **sistemātisks un mērķtiecīgs valsts atbalsts ilgtspējīgiem pētījumiem valodas resursu un rīku izveidē un uzturēšanā** var nodrošināt valodu līdztiesību ikdienas lietojumā un digitālajā vidē.

**PALDIES PAR  
UZMANĪBU!**



VPP-IZM-DH-2020/1-0001  
VPP-LETONIKA-2021/1-0006



**CLARIN-LV**

1.1.1.5/18/I/016

