

Latvian Language in the Digital Age: The Main Achievements in the Last Decade

*Inguna Skadiņa, Baiba Saulīte, Ilze Auziņa, Normunds Grūzītis, Andrejs
Vasiļjevs, Raivis Skadiņš, Mārcis Pinnis*



Background

- Language technologies in Latvia have a long history, starting at the end of the 50s
- Overview of HLT in Latvia from 1984 has been presented at several Baltic language technology events:
 - “Language and Technology in Europe 2000” in 1994
 - First Baltic conference on Human Language Technologies in 2004
 - Fourth Baltic conference on Human Language Technologies in 2010
 - Sixth conference on Human Language Technologies in 2014
 - Seventh conference on Human Language Technologies in 2016



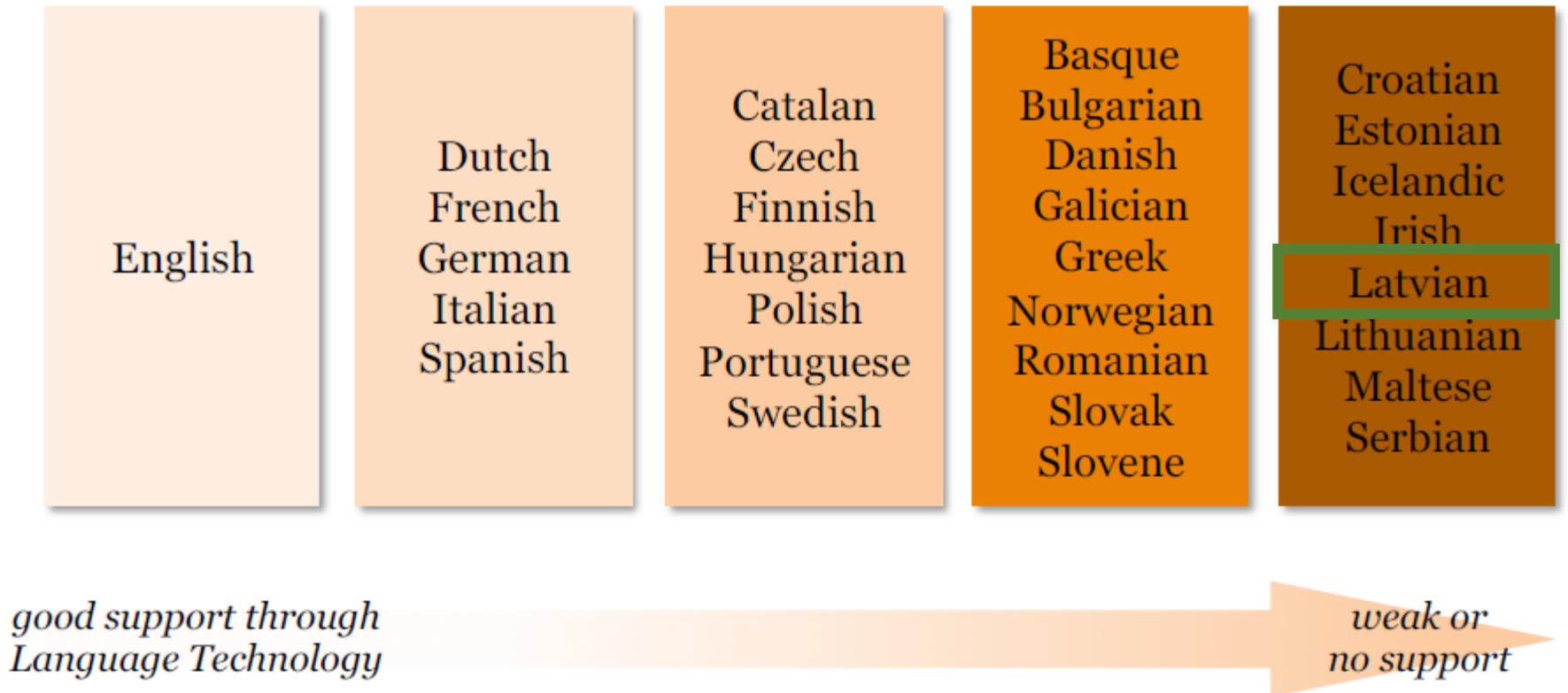
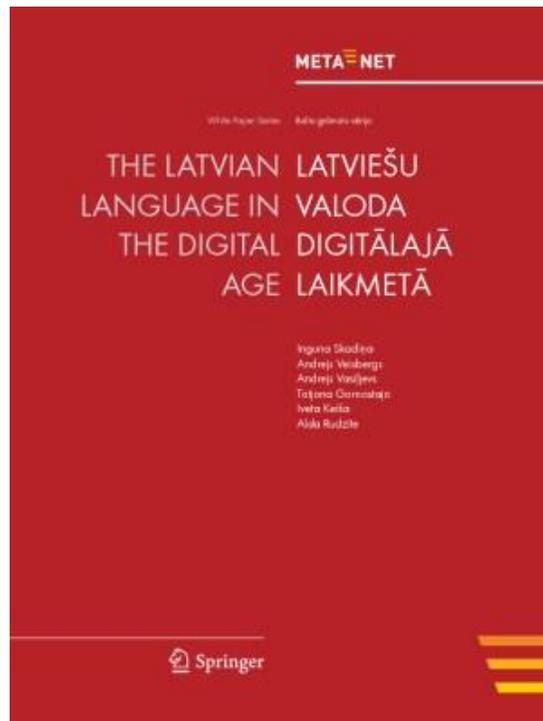
*Krievija Barona
Daiņu skapis*

Katalogs

- Par dziesmām un dziedāšanu.(8512)
- Cilvēka mūža ritums, ģimenes un radu sadzīve.(

to laike / kad Kyrenius Semmesoge exkan Syrien
by. Vnd Ietwens nogana / ka the šōw mheslotes
lichte / nectwens exkan šouwe Pille.
Ead nogan arridtezan Joseph no Galilea / aran
tās Pilles Nazareth / exkan to Judde semme / vs to
Pille David / kattrā tur dhēweey Bethlehem / Ea
pectez ka tas no to Namme vnd Kadde David by /
vs to / ka tās šōw mheslote lichte / ar Maria šouwe
šaloulata Gaspasche / kattrā tur apgrutenata by.
Vnd kad the tur path by / nātee tas laix ka thay pe
ezimpt by. Vnd tha peezimme šouwe pirmone Dhe
le / vnd tinne to exkan autims / vnd lichte to exkan we
ne Szille. Aesto themš nhe by ezittur neewena
meta exkan to Maiawete.

Language resources and tools for European languages



European Parliament resolution of 11 September 2018 on language equality in the digital age

Voting (11 Sept. 2018): 592 yes – 45 no



European Parliament

[Index](#) [Previous](#) [Next](#) [Full text](#)

Procedure : [2018/2028\(INI\)](#)

Document selected : [A8-0228/2018](#)

Texts tabled :
[A8-0228/2018](#)

Debates :
[PV 10/09/2018 - 27](#)
[CRE 10/09/2018 - 27](#)

Votes :
[PV 11/09/2018 - 6.15](#)
[Explanations of votes](#)

Texts adopted

Tuesday, 11 September 2018 - Strasbourg

[Language equality in the digital age](#)

[European Parliament resolution of 11 September 2018 on language equality in the digital age \(2018/2028\(INI\)\)](#)

The European Parliament,

- having regard to Articles 2 and 3(3) of the Treaty on the Functioning of the European Union (TFEU),
- having regard to Articles 21(1) and 22 of the Charter of Fundamental Rights of the European Union,
- having regard to the 2003 UNESCO Convention for the Safeguarding of the Intangible Cultural Heritage,
- having regard to Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information⁽¹⁾,
- having regard to Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information⁽²⁾,
- having regard to Decision (EU) 2015/2240 of the European Parliament and of the Council of 25 November 2015 establishing a programme on interoperability solutions and common frameworks as a means for modernising the public sector⁽³⁾,
- having regard to the Council resolution of 21 November 2008 on a European strategy for multilingualism (2008/C 320/01)⁽⁴⁾,



European Parliament
2014-2019

TEXTS ADOPTED
Provisional edition

P8_TA-PROV(2018)0332

Language equality in the digital age

European Parliament resolution of 11 September 2018 on language equality in the digital age (2018/2028(INI))

The European Parliament,

- having regard to Articles 2 and 3(3) of the Treaty on the Functioning of the European Union (TFEU),
- having regard to Articles 21(1) and 22 of the Charter of Fundamental Rights of the European Union,
- having regard to the 2003 UNESCO Convention for the Safeguarding of the Intangible Cultural Heritage,
- having regard to Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information¹,
- having regard to Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information²,
- having regard to Decision (EU) 2015/2240 of the European Parliament and of the Council of 25 November 2015 establishing a programme on interoperability solutions and common frameworks for European public administrations, businesses and citizens (ISA2 programme) as a means for modernising the public sector³,
- having regard to the Council resolution of 21 November 2008 on a European strategy for multilingualism (2008/C 320/01)⁴,
- having regard to the Council decision of 3 December 2013 establishing the specific programme implementing Horizon 2020 – the Framework Programme for Research and

¹ OJ L 345, 31.12.2003, p. 90.

² OJ L 175, 27.6.2013, p. 1.

³ OJ L 318, 4.12.2015, p. 1.

⁴ OJ C 320, 16.12.2008, p. 1.

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

**ELE**

The European Language Equality (ELE) project develops a strategic research, innovation and implementation agenda as well as a roadmap for achieving full digital language equality in Europe by 2030.

D1.17 Report on Greek

D1.18 Report on Hungarian

D1.19 Report on Icelandic

D1.20 Report on Irish

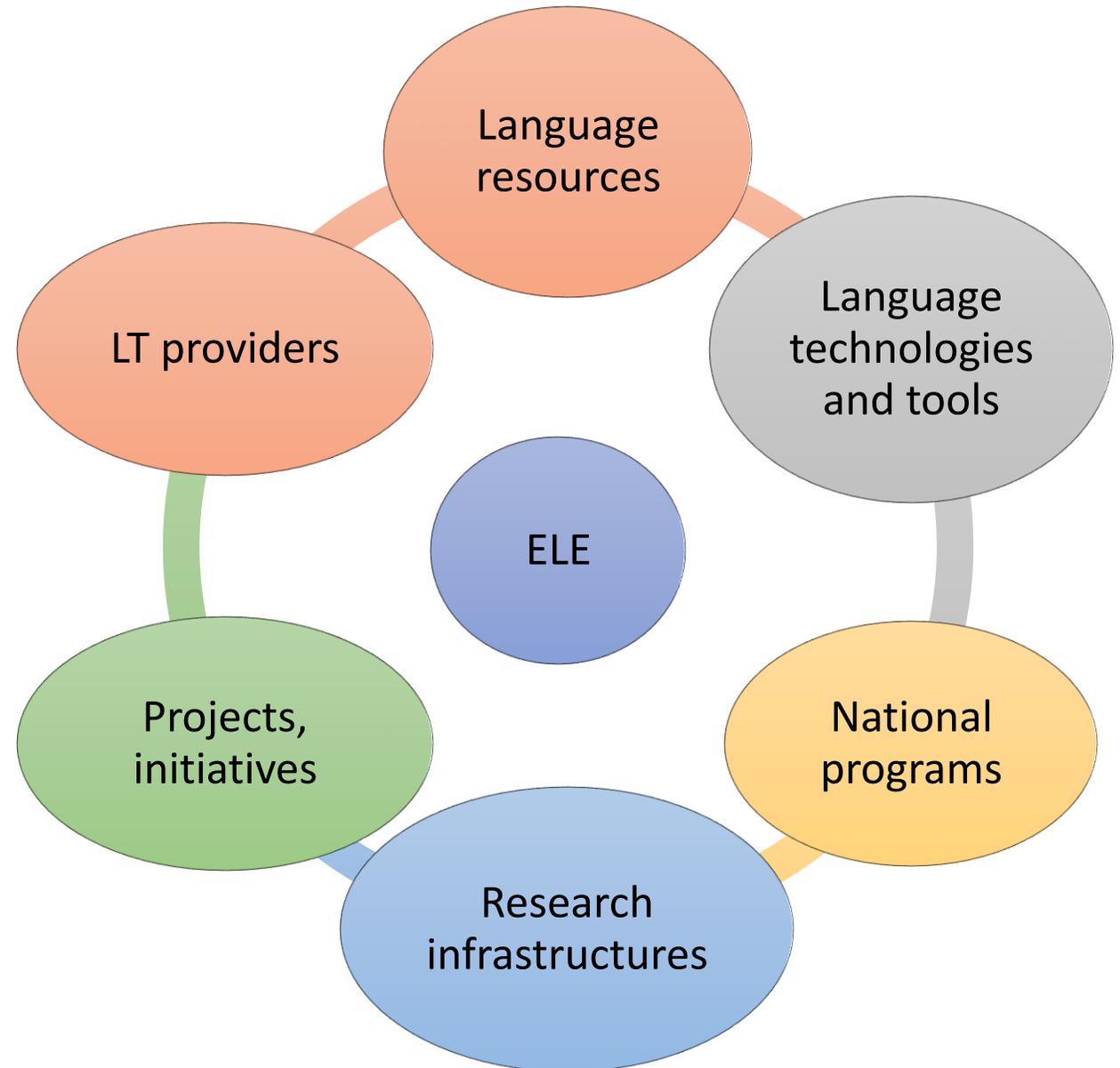
D1.21 Report on Italian

D1.22 Report on Latvian

D1.23 Report on Lithuanian

D1.24 Report on Luxembourgish

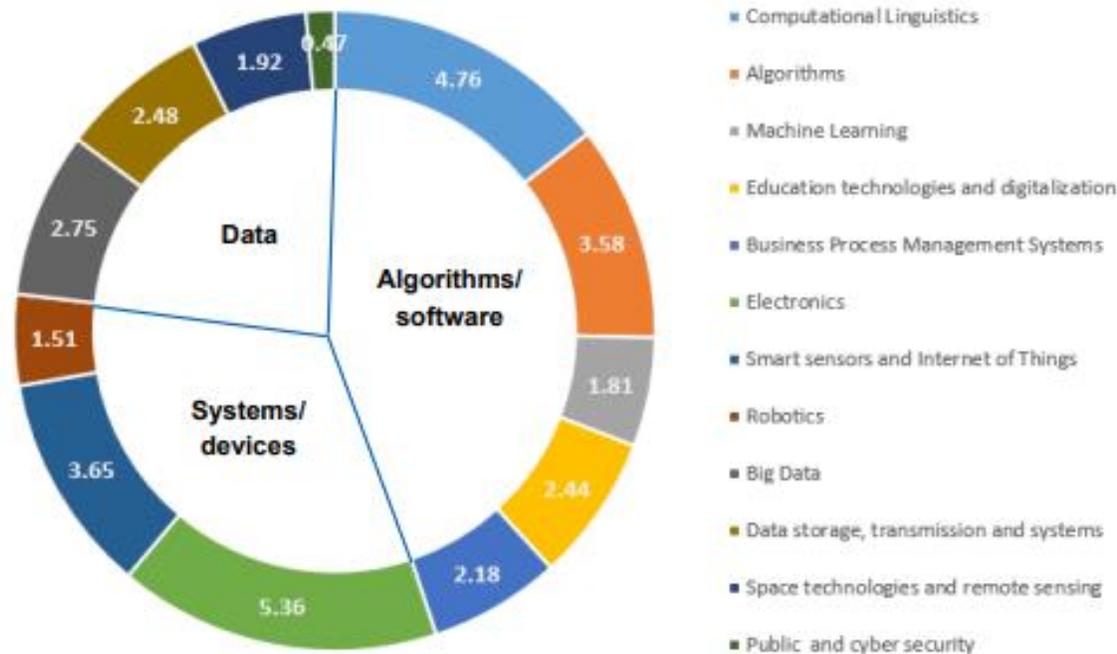
<https://european-language-equality.eu/deliverables/>



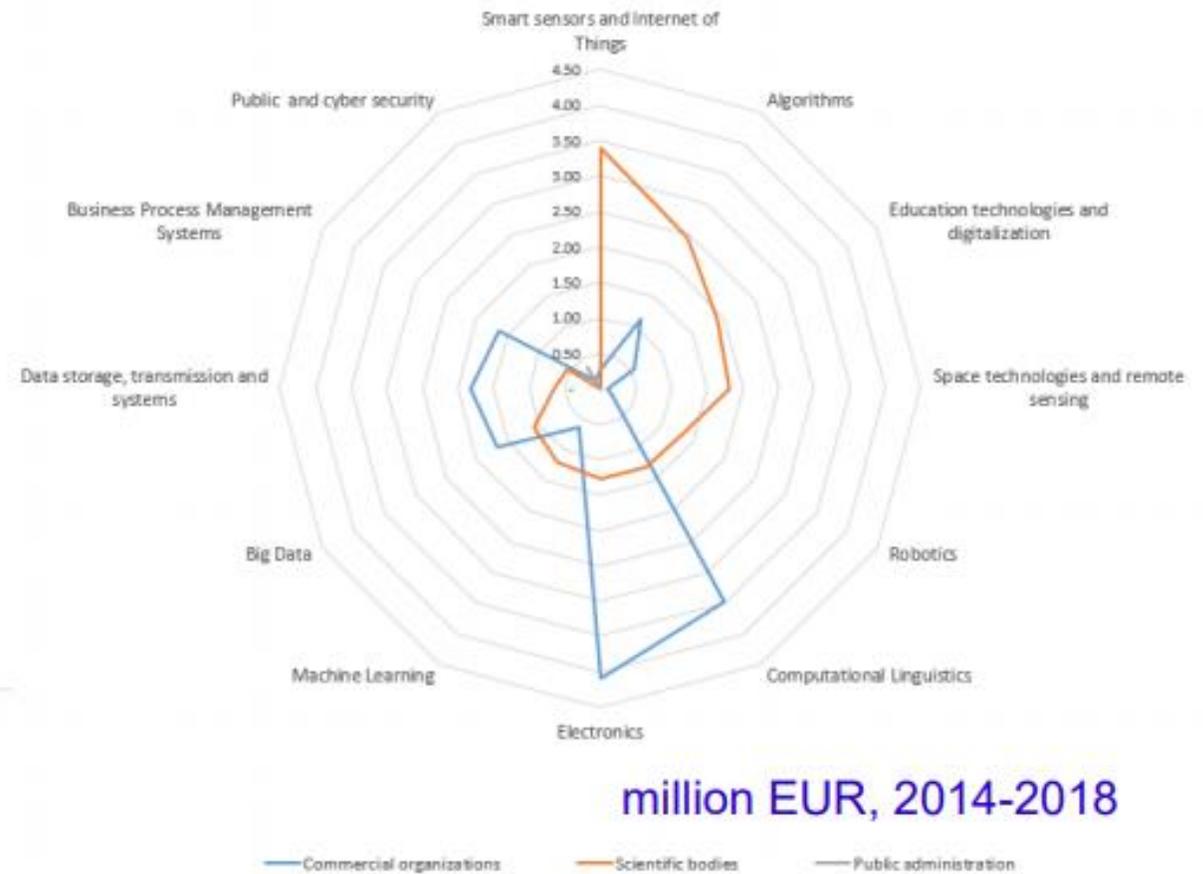
National programmes and initiatives

LANGUAGE TECHNOLOGIES IN ICT

Funding for R&D projects in RIS3 ICT thematic areas, scientific and commercial institutions



million EUR, 2014-2018



million EUR, 2014-2018

Actual policy planning documents

- **State Language Policy Guidelines for 2021-2027** lists several activities related to the creation and further development of Latvian language resources and tools. Since 2022 the guidelines are being implemented through the three-year State Research Programme “Letonica – Fostering a Latvian and European Society”.
- **Digital Transformation Guidelines for 2021-2027** include actions to enable Latvian citizens to access European Digital Space in their native language and to support development of the most important language resources for sustainable and wide use in digital services.
- **Report on National AI strategy on Developing Artificial Intelligence Solutions** lists several directions of action related to the development of AI-based language technologies, such as machine translation, speech technologies, inclusive technologies and terminology databases.

Language technologies in the Recovery and sustainability plan

- Development and implementation of high-level skills in three areas: **language technology** (5M EUR), quantum computing, and HPC
- Establishment of **Excellence Centre for Language Technology** to prepare curriculum for LT teaching, to advance language resources and create platforms and tools for studying and experimentation, and to conduct research involving young researchers
- The main research activities:
 - development of language resources for speech and text processing,
 - advancement of large pre-trained language models,
 - advancement of state-of-the-art speech technologies and machine translation,
 - development of software platforms and a shared technical infrastructure for education and research.

National initiatives and projects

Competence Centre Programmes

- In 2011, major IT companies and research institutions have established the IT Competence Centre (ITCC) **to create innovative language technology solutions**
- More than **10 projects** have been supported through the ITCC programme:
 - Creation of the first orthographically and phonetically transcribed Latvian speech corpus, followed by several projects on speech recognition
 - Several projects addressed machine translation
 - Others were devoted to intelligent human-computer interaction
 - As well as semantic parsing for media monitoring



Industry-Driven Research Programme

Five large-scale research and development projects have been supported through an industry-driven research programme (European Regional Development Fund):

- development of language resources and tools for natural language understanding and generation
- development of neural network solutions for less resourced languages, multilingual affective human-computer interaction
- application of speech technologies for multilingual meeting management, and the transcription of medical speech
- development of Latvian speech recognition and synthesis for medical applications

NATIONAL
DEVELOPMENT
PLAN 2020



EUROPEAN UNION
European Regional
Development Fund

INVESTING IN YOUR FUTURE

Latvian Council of Science

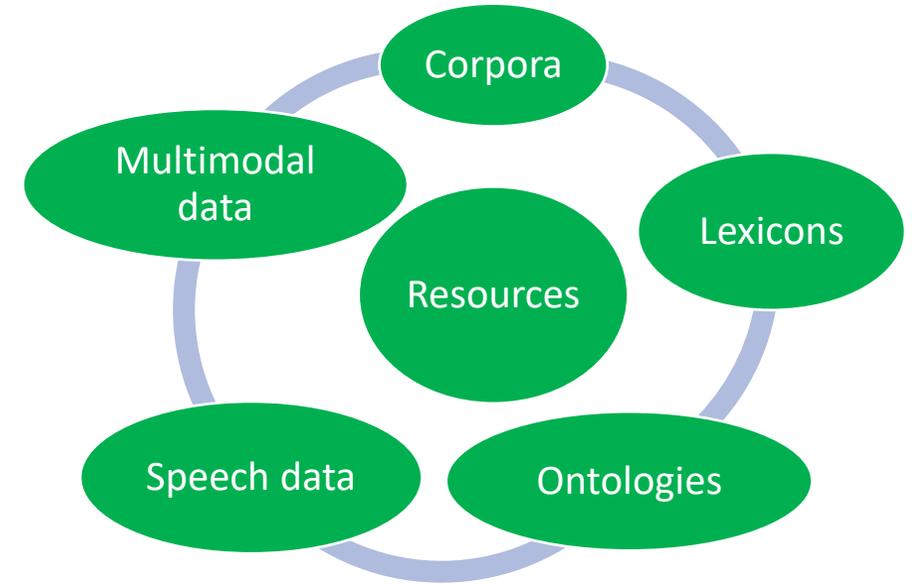


- State Research Programme, project **Digital Resources for Humanities** (2020-2022) has developed integrated language resources and tools for Digital Humanities
- State Research Programme, project **Research on Modern Latvian Language and Development of Language Technology** (2022-2024) aims to advance research on the grammatical, lexical-semantic, phonetic and phonological systems of modern Latvian and the Latvian sign language, using data-driven methods
- Since 2018, several projects have been also supported by the Latvian Council of Science, including:
 - the development of Latvian Learner Corpus
 - creation of a pilot Latvian WordNet and means for neural word-sense disambiguation
 - research on natural language understanding and generation for human-computer interaction

International projects and cooperation

During the last decade, the Latvian language technologies have been part of research, innovation and deployment actions in several FP7, Horizon 2020 and CEF projects on automated translation, speech technologies, human-centred AI, etc.





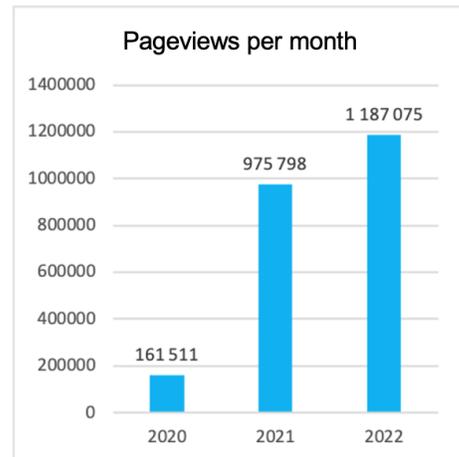
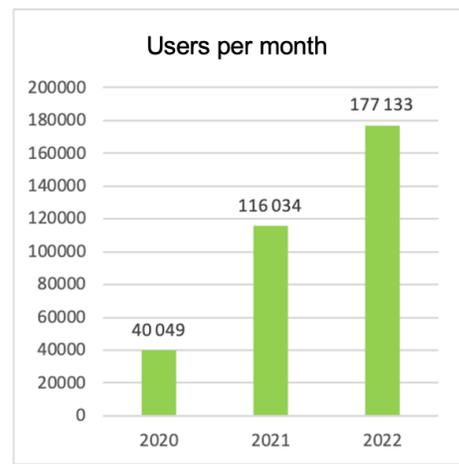
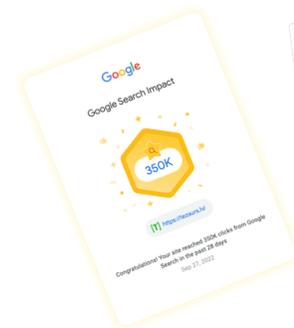
Lexical, text and speech resources

The Tēzaur.lv platform

The screenshot displays the Tēzaur.lv platform interface. At the top, there are three search bars: one for 'Latviešu literārās valodas vārdnīca' (64 265 šķirķji), one for 'Mūsdienu latviešu valodas vārdnīca' (45 939 šķirķji), and one for 'Rudens versija 2022' (387 576 šķirķji). The main content area shows search results for 'valsts' (state), including a list of homonyms, a list of related terms (e.g., valsts, valstis, valstīte), and a detailed definition of 'valsts' as a political organization. The interface is clean and user-friendly, with a sidebar on the left for navigation.

Public interface

Editing interface
Data model
Database
Open data
Search
API



Seko mums:

Kas ir Tēzauris Android lietotne Autori
 Apzīmējumi Kindle vārdnīca Pateicības
 Izmantotie avoti Atvērte dati Iestatījumi
 Citas vārdnīcas Tēzaura API BUJ

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

**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

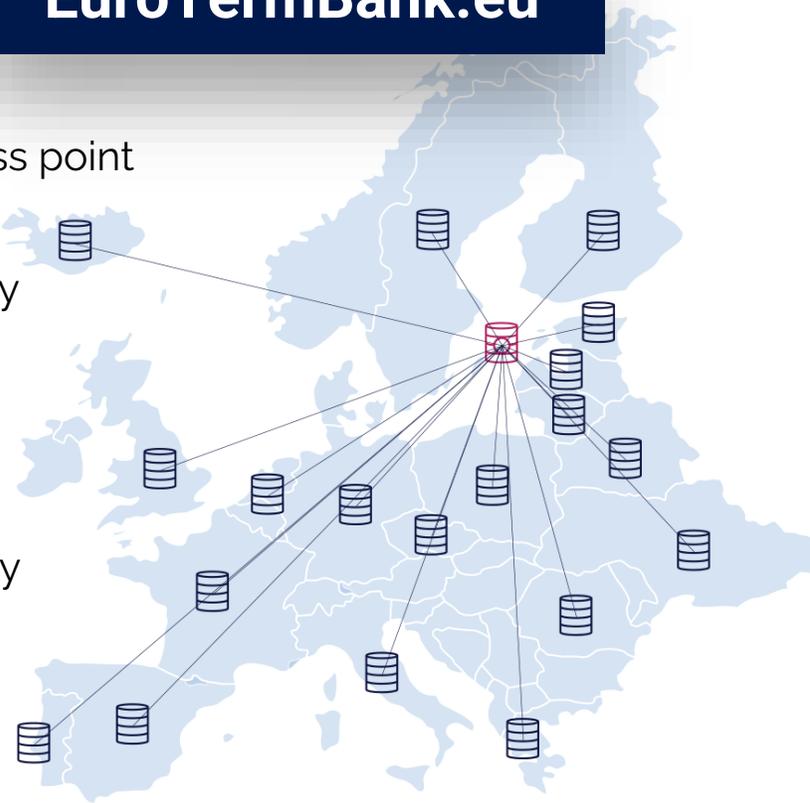
EN natural language generation

EN quasi-natural language

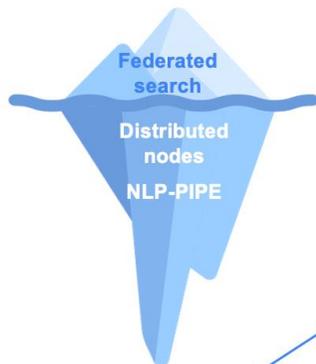
EuroTermBank

EuroTermBank.eu

- ✓ eurotermbank.com - 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



The Korpuss.lv platform



Developer: LU LFMI
Node: nosketch.lnb.lv

Developer: LU LaVI, LU MII, LU HZF
Node: nosketch.korpuss.lv

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

Nacionālā korpusu kolekcija **Meklēšana** Materiāli ▾ Par Korpuss.lv LV ▾

sirds*

Korpuss	Relatīvais biežums	Absolūtais biežums
uz 1 miljonu		
Pārspridumi Skolēnu pārspridumu korpus		
Izstrādātāji: LU MII, LiePU, RTA Mezglis: nosketch.korpuss.lv	2257	510
LatSenRom Latviešu senākie romāni		
Izstrādātāji: LNB Mezglis: nosketch.lnb.lv	1420	8234
Rainis Raiņa darbu korpus		
Izstrādātāji: LU MII Mezglis: nosketch.korpuss.lv	1196	2752
ĪsprozaS Stieviešu isproza		
Izstrādātāji: LU LFMI Mezglis: nosketch.lnb.lv	1083	1272
Senie Latviešu valodas seno tekstu korpus		
Izstrādātāji: LU LaVI, LU MII, LU HZF Mezglis: nosketch.korpuss.lv	765	2081
LVMED Radioloģisko izmeklējumu transkripciju korpus		
Izstrādātāji: LU MII, RAKUS Mezglis: nosketch.korpuss.lv	616	97
MuLa2012 Mūsdienu latgaliešu tekstu korpus 2012		
Izstrādātāji: LU MII, RTA Mezglis: nosketch.korpuss.lv	426	565
PanDi Pandēmijas dienasgrāmatas		
Izstrādātāji: LU LFMI Mezglis: nosketch.lnb.lv	408	289
MuLa2022 Mūsdienu latgaliešu tekstu korpus 2022		
Izstrādātāji: RTA, LU MII Mezglis: nosketch.korpuss.lv	394	1101
Timeklis2020 Latviešu valodas timekļa korpus		
Izstrādātāji: LU MII Mezglis: nosketch.korpuss.lv	334	164 504
Līla Lietuviešu-latviešu-lietuviešu paralēlo tekstu korpus		
Izstrādātāji: LU MII, VDU Mezglis: nosketch.korpuss.lv	305	1734
LVK2018 Lidzsvarotais mūsdienu latviešu valodas tekstu korpus		
Izstrādātāji: LU MII Mezglis: nosketch.korpuss.lv	270	3321
Emuāri Latviešu valodas emuāru korpus		
Izstrādātāji: LU MII Mezglis: nosketch.korpuss.lv	256	2128

CONCORDANCE

simple sirds* 1,272 (1,083.32 per milion)

Details	Left context	KWIC	Right context
11	doc#0 atļu? ta caur savu augstprātību un aku mīlestību tāpat samaitā – par viņas	sirdi	nomaz, bet tikai par ārēju izskatu rūpēdamās. Kad Latiņa ko iedomājusēs, tu
12	doc#0 mi vien sprēžot laiku pavadijām, un tu man nemaz tālāk nestāstīji, kas tavš	sirdsdraugs	un kādā vietā tas." "Viņš tagad Z. viesnicā par sulaini," Anna stāstīja, "bet v
13	doc#0 talga, kādā valodā vārds tiek nosaukts." "Tev gan taisnība, ka vārds cilvēka	sirdi	un dabu nepārvērš, un pa lielākai daļai tas jau no tēva mantots – jo ļotiak, ka
14	doc#0 bas varētu būt arī viltīgs – 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 jau nu neper
16	doc#0 zar ģeķiem." "Ne, ne, mīļā Trūlī!" sacīja Anna, Trūtes roku saķērusē un pie	sirds	splezdamā. "Nevaru par to neko vairāk sacīt, – es tikai zinu, ka viņš ir skais
17	doc#0 – es tikai zinu, ka viņš ir skais, laips un patīkams, un ka es viņu no visas	sirds	milu. Nekad neesmu ko dziļāki apdomājuse, kā tikai, cik jauki un laimīgi bij i
18	doc#1 i, kura veda uz ārpusi; te tas sev priekšā ieraudzīja kādu sievietes tēlu, viņa	sirds	sāka ātrāk pukstēt – vai tā varbūt nebija viņa? Šis slaiskais stāvs tam jau izli
19	doc#1 Jūsū abēju saruna, kuru neviļus dzirdēju, bija pilnens vairāk, kurš lika manai	sirdij	iet pāri. Es ļus negribēju apvainot, jo ļusu persona man sveša un vienaldzīg
20	doc#1 i, – nemandiniet mani, neļaujieties aizrautes no acumirkļa, nemandiniet manā	sirdi	ļūtas, kuras es vairs nespēju apslāpēt. Es esmu dzīve tik daudz cietusi, tik c

Rows per page: 10 11–20 of 1,272

CONCORDANCE

simple sirds* 2,081 (194.98 per milion)

Details	Left context	KWIC	Right context
31	doc#17 i thas paßudde priefchan them. Vnde Batoy ftarpan Bôw / nhe detctze muffa	Szyrdtzc	exkan mums / Khad thas ar mums trefley wûrBon tho Czele / Khad thas murr
32	doc#17 i aridctzan nofukmichenne / Beth es grib yums atkal redtzeeth / vnde yuufe	Syrdtzc	Bôw lyxmofes / vnd yuufe Lyxmbe bus newenam no yums yempt. Vnde py
33	doc#17 i / Kur tu no effe ? Beth peetzth tho / ka es to vs yums trefleys efme / gir yuufe	Syrdtzc	puelina nofukmichenne tappuffil. Beth es Backe yums tho trefchan. Thas yumt
34	doc#17 inne Mhere dome es yums / nhe ka tha Palfoulde dode / dome es yums. Juufe	Szyrdtzc	nhe iftruczenas / vnde nhe ybvytas Bôw. Juss effieth cyrdichfche / ka es yums
35	doc#17 tho loune / perakeses tam labbam. Ta Bralige myleibete ftarpan wen otre gyr	Szyrdtzc	duebeniga. Wens nake them otram ar Godennachenne priefchan. Nehe effe
36	doc#19 ru.abhängig.wipisch stahwu atkahrn,Er stehet abhängig. Tu ne dsihwo wis no	firds	, tu atkahrn wens dsihwo,du arbeitest nicht vom hertzen, du ar-beitest hoÿe,
37	doc#19 i,Morgenland. At=aust,licht werden.Metaph.Sich erholen. Vid. ante fol. sub F.	Sirds	ataustdas hertz erholet sich Akkotsdie Spitze von einer Korn Åhren Akkotiſc
38	doc#19 lem Ewigen leben trachten ringen. Duhkitbetribt, abgemattet seÿn.Sadukkufi	Sirds	-Ein betribt, untraunigesCur. hertz. Duhzamais,der hindern, hinterste Duhmp
39	doc#19 arsche, die im Korn schreyet Greestauff und überlegen.agpreest.umbkehren,	Sirds	agpreeschahs GraestSchneiden,tkehren,Tas Nasis atgreeschahs,i
40	doc#19 peem tinnahs er gehet mit Vieh um. Kluksteht,wie eine bruhthenne glucken	Sirds	ta Klukst. KrusfaHagel Krusfeene, ablu dshehronesApffel Trank Kuzzeneseli

Rows per page: 10 31–40 of 2,081

CONCORDANCE

simple sirds* 1,101 (393.93 per milion)

Details	Left context	KWIC	Right context
21	doc#44 is Kuorsovā, sevi skaitu arī kai varakļoniti un cylvāku no Aglyunys. Muna	sirds	Latgola ir Varakļonūs, Aglyuna n dzymtajā Kuorsovā. Sovim bārim kult
22	doc#50 Iņa i tai tys vyss ar golā sagojua. Sirsneiba, vinkuorsēiba, gaidēšona, nu	sirds	nastys duovonys — Latgola otkon ldeve muzikim jaunus breinumus, par k
23	doc#50 s dzimtyš iiktyrūs. Romans teik raksturtoots kal šokts na tik Latgolyš jauzū	sirds	i dīvēsēlis dzēivē, nu arī Latvejis vīesturis kryspunktūs — skaiteltojam bys
24	doc#56 ģmār pateik bārim rudoit, kai tei kasdīna var nūtkit struodojūt. Maņ suop	sirds	, ka mes drupeit par vālu aasm saprotūsē sovaiduokys skolyš vajadzēbu. M
25	doc#57 , kū stuosteit: „90. godūs populāri beja balderiāni, niule nīvīnam navajaga,	sirdsmāteri	padavai. Dažus augus poši padarejom populārus, pīrāmāru, lofantu — Kore
26	doc#70 sātu, a sastajam navārieju nūsacēit, kas vainis. Smird alkohols, dreizuoka	sirdsdarbelba	, a pats nī par kū nasasyudz. Es pasacēju, ka ļū vajāg nūvārrot. Vītu nav, i
27	doc#70 sīt kū cytu. Interesantuok raksteit i skaitēit, atkluojūt nazynomus faktus. Nu	sirds	vieleļu Osyuna pamatskolyš nūvodpīeni lceibys ustobā dabuot Janīnys atr
28	doc#70 busu atbraukušī, eipašī baļtinavīšī, saacējūsī: „Es pi jums atjuoju.“ Jai suop	sirds	, ka Osynā vaira nazynūs, ka pareizais, eisti latgaliskais ltuos vītyš nūsai
29	doc#75 vede juos izšyvumus. Par tū ļei lelu naudu sajēme. Pats Dymša nūmyra ar	sirdstriku	. Ar sunēiti izguoja pasastaļguot, izskrēja ļeli suni. Cikoms jis sovu kuceiti gi
30	doc#75 te atguojuse da babenis atsavadeit. A kungs šytymā teiruma nūmyra ar	sirdstriku	. Naspēja puordzeivout, ka juopamat Rušyuna.“ Rušyuns luocs Antoneja:

Rows per page: 10 21–30 of 1,101

20+ corpora, 10+ institutions, 3+ nodes

NLP-PIPE: Latvian NLP Pipeline as a Service

Used by: Korpuss.lv, National Library of Latvia, News agency LETA, IT students, etc.

Select a sample text

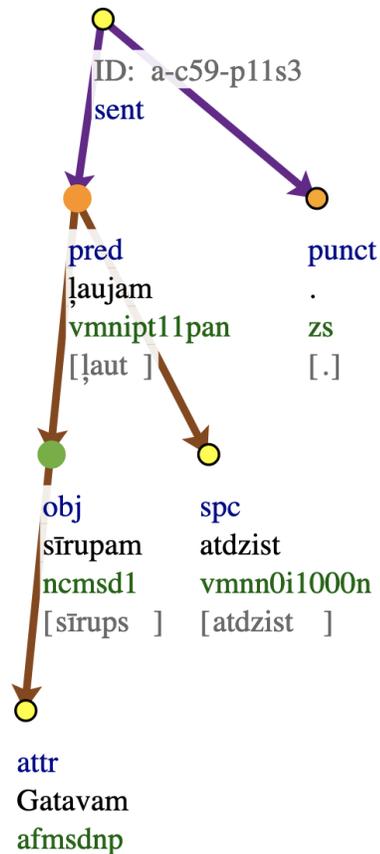
2022. gada 6.-7. oktobrī Rīgā norisinās Baltijas Valodu tehnoloģiju konference, ko rīko LU, LU MII un Tilde, pastāstīja programmkomitejas vadītāja Inguna Skadiņa.

Go tokenizer × ner × morpho × parser × Visual CoNLL JSON

2022. gada 6.- 7. oktobrī time Rīgā GPE norisinās Baltijas Valodu tehnoloģiju konference event , ko rīko LU organization , LU MII organization un Tilde organization , pastāstīja programmkomitejas vadītāja Inguna Skadiņa person .

INDEX	FORM	LEMMA	UPOSTAG	XPOSTAG	FEATS	HEAD	DEPREL
#text=2022. gada 6.- 7. oktobrī Rīgā norisinās Baltijas Valodu tehnoloģiju konference , ko rīko LU , LU MII un Tilde , pastāstīja programmkomitejas vadītāja Inguna Skadiņa .							
1	2022.	2022.	ADJ	xo	NumType=Ord	2	amod
2	gada	gads	NOUN	ncmsg1	Case=GenlGei	3	nmod
3	6.-	6.-	NUM	xn	NumType=Car	5	amod
4	7.	7.	ADJ	xo	NumType=Ord	5	amod
5	oktobrī	oktobris	NOUN	ncmsl2	Case=LoclGer	7	obl
6	Rīgā	Rīga	PROPN	npfsl4	Case=LoclGer	7	obl
7	norisinās	norisināties	VERB	vmyip_330an	Evident=FhlMc	22	ccomp
8	Baltijas	Baltija	PROPN	npfsg4	Case=GenlGei	9	nmod
9	Valodu	valoda	NOUN	ncfpg4	Case=GenlGei	10	nmod
10	tehnoloģiju	tehnoloģija	NOUN	ncfpg4	Case=GenlGei	11	nmod
11	konference	konference	NOUN	ncfsn5	Case=NomlGe	7	nsubj
12	,	,	PUNCT	zc	-	14	punct
13	ko	kas	PRON	pr000an	Case=AcclProi	14	obj
14	rīko	rīkot	VERB	vmnpt230an	Evident=FhlMc	11	acl
15	LU	LU	PROPN	yp	Abbr=Yes	14	nsubj
16	,	,	PUNCT	zc	-	17	punct
17	LU	LU	PROPN	yp	Abbr=Yes	15	conj
18	MII	mii	X	y	Abbr=Yes	17	iobj
19	un	un	CCONJ	cc	-	20	cc
20	Tilde	tilde	NOUN	ncfsn5	Case=NomlGe	18	conj
21	,	,	PUNCT	zc	-	7	punct
22	pastāstīja	pastāstīt	VERB	vmnist330an	Evident=FhlMc	0	root
23	programmkomitejas	programmkomiteja	NOUN	ncfsg4	Case=GenlGei	24	nmod
24	vadītāja	vadītāja	NOUN	ncfsn4	Case=NomlGe	26	nmod
25	Inguna	Inguns	PROPN	npmsg1	Case=GenlGei	26	nmod
26	Skadiņa	Skadiņa	NOUN	ncfsn4	Case=NomlGe	22	nsubj
27	.	.	PUNCT	zs	-	22	punct

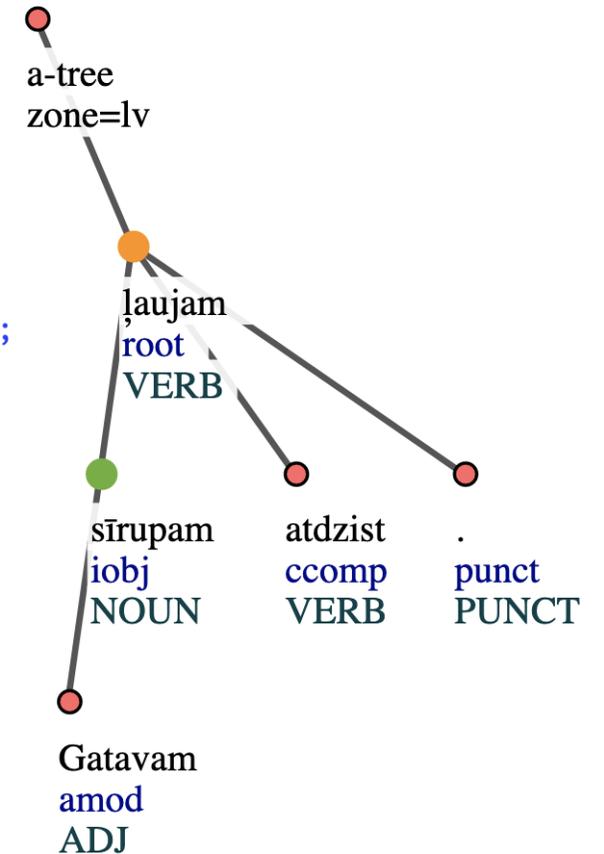
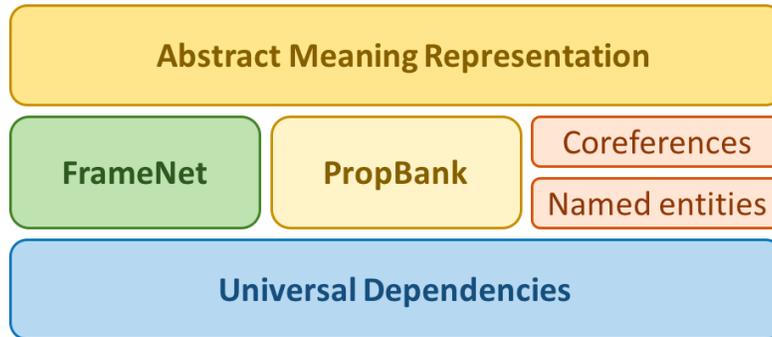
Latvian Treebank & Latvian UD Treebank



PML Tree Query: <https://lindat.mff.cuni.cz/services/pmltq/>
(incl. LVTB & UDLV; UDLT)

```
a-node [
  role="obj"
  and m/tag~"n...d.",
  parent
  a-node [m/tag~"vm.*"]];
```

```
a-node [
  deprel="iobj"
  and tag="NOUN",
  parent
  a-node [tag="VERB"]];
```



Grūzītis, N., Pretkalniņa, L., Saulīte, B., Rituma, L., Nešpore-Bērzkalne, G., Znotiņš, A., Paikens, P. [Creation of a balanced state-of-the-art multilayer corpus for NLU](#). Proceedings of the 11th LREC Conference, 2018

Latvian FrameNet

<http://framenet.korpus.lv>

570 BFN frames

2900 lexical units (5.1 per frame)

25897 annotation sets (8.9 per LU)

Semantiskie freimi

Leksiskās vienības

Korpusa piemēri

Atsauces

Latvian
FrameNet

palīdzēt. VERB (Assistance)

Visi leksiskās vienības lietojuma piemēri: [83](#)

Freima elementi, to biežums un gramatiskā realizācija:

Freima elementi	Piemēri	Realizācija
Benefited party	45	iobj NOUN (23) iobj PRON (17) iobj PROPN (2) iobj NUM (1) iobj VERB (1) ccomp VERB (1)
Focal entity	4	obl NOUN (3) iobj NOUN (1)
Goal	56	ccomp VERB (53) iobj VERB (1) acl VERB (1) dep VERB (1)
Helper	64	nsubj NOUN (21) nsubj PRON (33) nsubj PROPN (5) nsubj X (2) obj NOUN (1) nsubj:pass PRON (1) nsubj VERB (1)
Time	1	obl PRON (1)

- X [a-p1039-p3s5] Varbūt tas ir vairāk ārēji - vainadziņi , svecītes , bet **gaišums** noteikti **PALĪDZ** arī iekšēji **cilvēkiem** **justies** labāk .
- X [a-p159-p33s2] Kad **viņš** atgriezās uz dzīvi šeit , daudz ziedoja dažādiem projektiem - lāpīja jumbus , **PALĪDZĒJA** **skolām** **iegādāties** datorus un grāmatas .
- X [a-p1604-p2s2] Ģimenes , **kurām** **PALĪDZĒT** , tika izvēlētas , sadarbojoties ar Sociālo lietu pārvaldi un Izglītības pārvaldi .
- X [a-p2379-p9s2] Viņa informēja , ka Latvijā patlaban ir recepšu medikaments « Champix » , **kas** efektīgi **PALĪDZ** **smēķētājam** **atmest** kaitīgo ieradumu .
- X [a-p3129-p19s1] **PALĪDZIET** savam **mazulim** **izvēlēties** mīļāko grāmatu par dzīvniekiem .
- X [a-p3344-p6s5] Bet manā uztverē apkalpojošā sfērā **strādājošajiem** ir **JĀPALĪDZ** **cilvēkiem** !
- X [a-p3315-p3s12] Ar māsu mums ir labas attiecības , viņai ir divas lielas meitas , **mēs** **PALĪDZAM** viena **otrai** .
- X [a-p2098-p7s2] Valdošā Afrikāņu Nacionālā Kongresa vēlēšanu manifestā Zuma solīja darīt vairāk , lai **PALĪDZĒTU** **viņiem** .
- X [a-p14546-p4s1] Paredzēts , ka **viņš** **PALĪDZĒS** **pacientiem** , kas sirgst ar Alcheimera slimību un citām vecuma plānprātības formām .
- X [a-p1297-p4s10] **Jēkabam** **izdzīvot** **PALĪDZĒJIS** **tas** , ka vienmēr pratis ievērot mērenību , arī ēšanā .

Assistance

Definition:

A **Helper** benefits a **Benefited_party** by enabling the culmination of a **Goal** that the **Benefited_party** has. A **Focal_entity** that is involved in reaching the **Goal** may stand in for it.

FEs:

Core:

Benefited_party The **Benefited_party** receives a benefit from the action of the **Helper**.

Focal_entity This FE identifies a **Focal_entity** involved in achieving the **Goal**.

Goal [Goal] The desirable state of affairs that the **Benefited_party** is involved in and which is enabled by the **Helper**.

Helper The **Helper** performs some action that benefits the **Benefited_party**.

Frame-frame Relations:

Inherits from: [Intentionally_act](#)

Is Inherited by: [Supporting](#)

Parallel corpora

Filter by:

45 Language Resources (Page 1 of 3)

« Previous | Next »

Order by: Resource Name A-Z ▾

Language

- Latvian (45)
- English (45)
- Bulgarian (16)
- French (16)
- Modern Greek (1453-) (16)

more

Multilinguality Type

- Parallel (45)

Resource Type

Media Type

Licence

Conditions of Use

Linguality Type

Data Format

Domain

Appropriateness For DSI

Funding Project

Resource Type:

Corpus: 

 Anonymised ParaCrawl release 7 Latvian-English	↓ 0 👁 1
English Latvian	CC0-1.0
 Bilingual corpus from the European Vaccination Portal (LV-EN)	↓ 8 👁 8
English Latvian	CC-BY-4.0
 Bilingual corpus from the Publications Office of the EU on the medical domain (EN-LV)	↓ 8 👁 7
English Latvian	CC-BY-4.0
 Bilingual corpus from the Publications Office of the EU on the medical domain v.2 (EN-LV)	↓ 0 👁 6
English Latvian	CC-BY-4.0
 Bilingual corpus made out of PDF documents from the European Medicines Agency, (EMA), https://www.ema.europa.eu, (February 2020) (EN-LV).	↓ 6 👁 6
English Latvian	CC-BY-4.0
 Corpus of State-related content from the Latvian Web	↓ 17 👁 118
English Latvian	CC-BY-SA-4.0

ParlaMint: UD-annotated comparable corpora



ParlaMint-LV 2.1 (Latvian parliament)

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

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

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

CQL query:

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

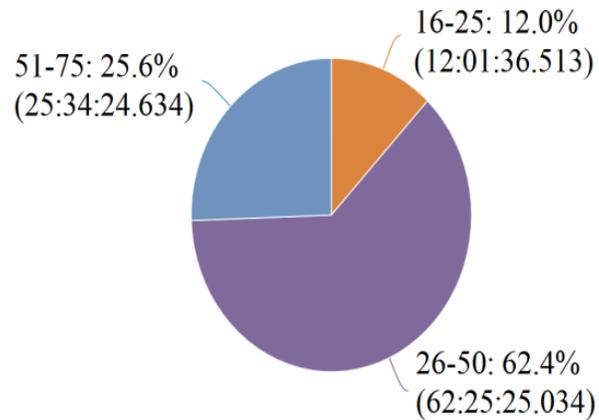
Erjavec, T., Ogradniczuk, M., ..., Dargis, R., ..., Krilavičius, T., et al. [ParlaMint: Comparable Corpora of European Parliamentary Data](#). Proceedings of the CLARIN Annual Conference, 2021

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

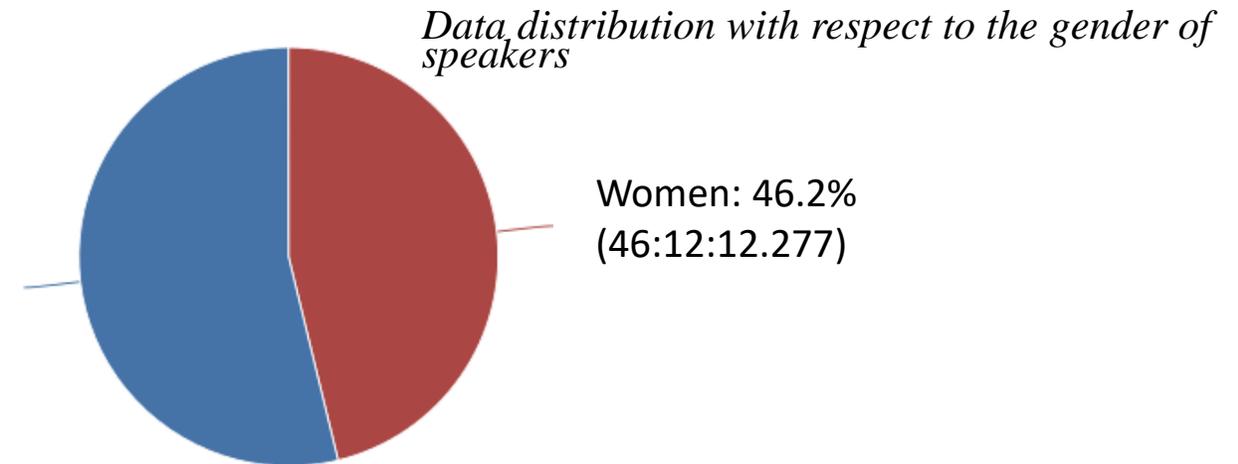


Latvian speech corpus

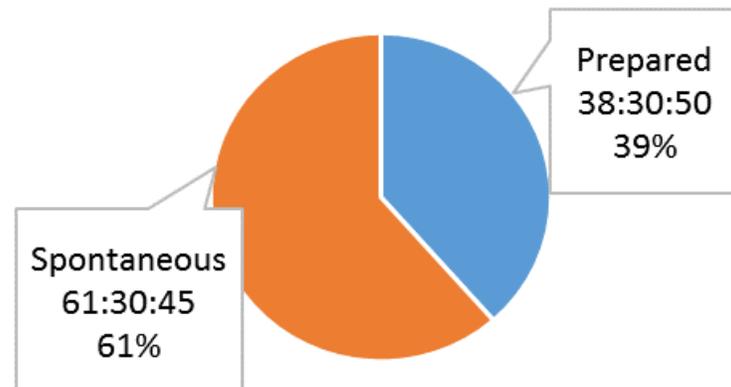
The **first** Latvian speech corpus (2013) is designed specifically for speech recognition purposes. Contains an orthographically (100h) and phonetically (4h) annotated data.



Men: 53.8%
(53:49:13.904)



Data distribution with respect to the age of speakers



Duration of segmented recordings (hh:mm)

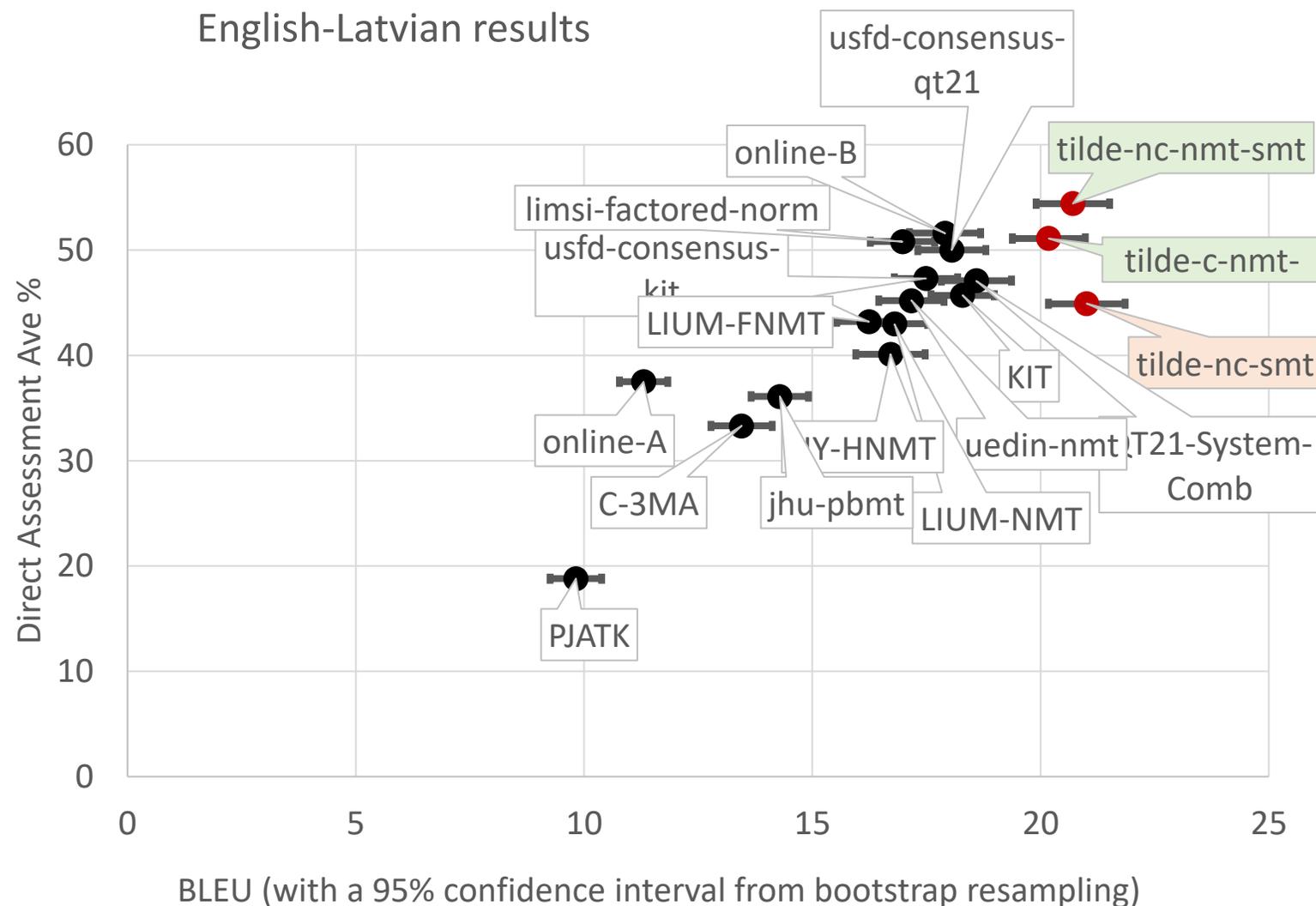
	Total dur of recordings	Word- and segment
Dial.	76:41	74:13
Monol.	12:51	9:05
Total	89:32	83:19



Machine Translation

Latvian in WMT 2017

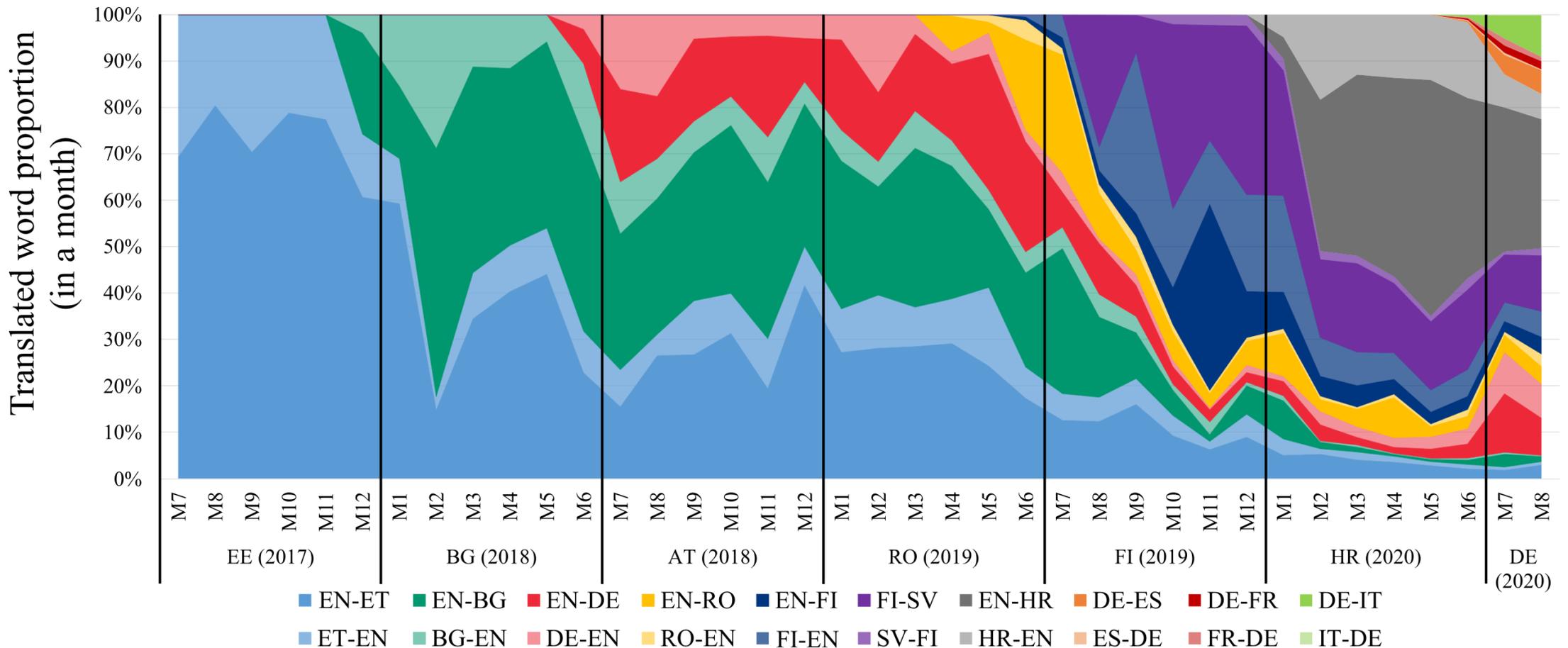
- In 2017, Latvian was included in the WMT shared task on news translation
- Tilde's NMT systems were the best-ranked systems that year
- Tilde's systems were also best-ranked for Estonian in 2018 and Lithuanian in 2019



EU Council Presidency Translator



- Tilde's EU Council Presidency Translator has served officials, translators, delegates, and visitors of 8 EU Council Presidencies since 2015.



What are the recent and current research topics of NMT for Latvian?

Some examples are:

- Terminology integration in NMT (Bergmanis and Pinnis, 2021a, 2021b)
- Analysis of biases in NMT systems (Stafanovičs et al., 2020)
- Robustness of NMT systems (Bergmanis et al., 2020)
- Speech translation (Alves et al., 2020)



Speech processing

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.



Today speech recognition technologies can recognize about 80% of spoken text, if the quality of the audio recording is good enough, and could reach 94% accuracy in specialized applications.

RUTA: MED - automatic transcription of medical records

RUTA^{med}

* Saglabāt 5s Pabeigt B I U x² x₂ doc json

MR prostatai ar k/v.
Prostata 3,9 x 3,3 x 3,8 cm. Tilpums 26,15 cm³.
PSA līmenis nav zināms.
Stāvoklis pēc TUR-P operācijas ar tipisku defektu.
Dziedzera pārejas daļā redzami labdabīgas hiperplāzijas mezgli.
Audus ar difūzijas ierobežojumu dotajā izmeklējumā nesaskatu.
Sēklas pūslīši simetriski, bez strukturālām izmaiņām.
Urīnpūslis gludām sienīņām, bez intralūmenāliem ieslēgumiem.
Patoloģiska lieluma l/m nesaskatu.

Slēdziens
Labdabīga prostatas hiperplāzija.
Stāvoklis pēc TUR-P operācijas.
Pašreiz nav datu par augstas malignitātes tumora audiem prostatā.

00:01:21 / 00:01:32

A prototype developed by IMCS at University of Latvia in cooperation with Riga East University Hospital

Being tested already by several hospitals and clinics

Znotiņš, A., Dargis, R., Grūzītis, N., Bārzdiņš, G., Goško, D. [RUTA: MED – Dual Workflow Medical Speech Transcription Pipeline and Editor](#). In Proceedings of the 27th NLDB Conference, 2022

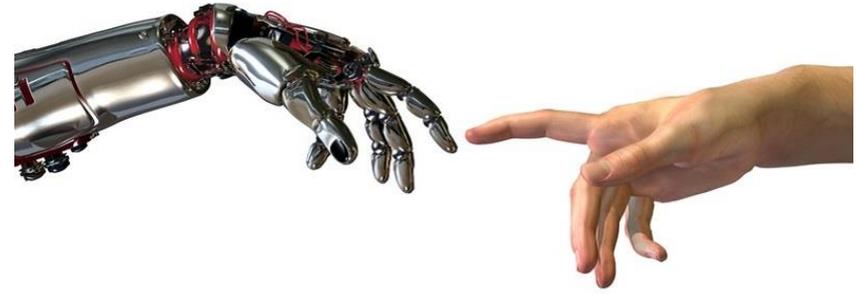
1.1.1.1/18/A/153

NATIONAL
DEVELOPMENT
PLAN 2020



EUROPEAN UNION
European Regional
Development Fund

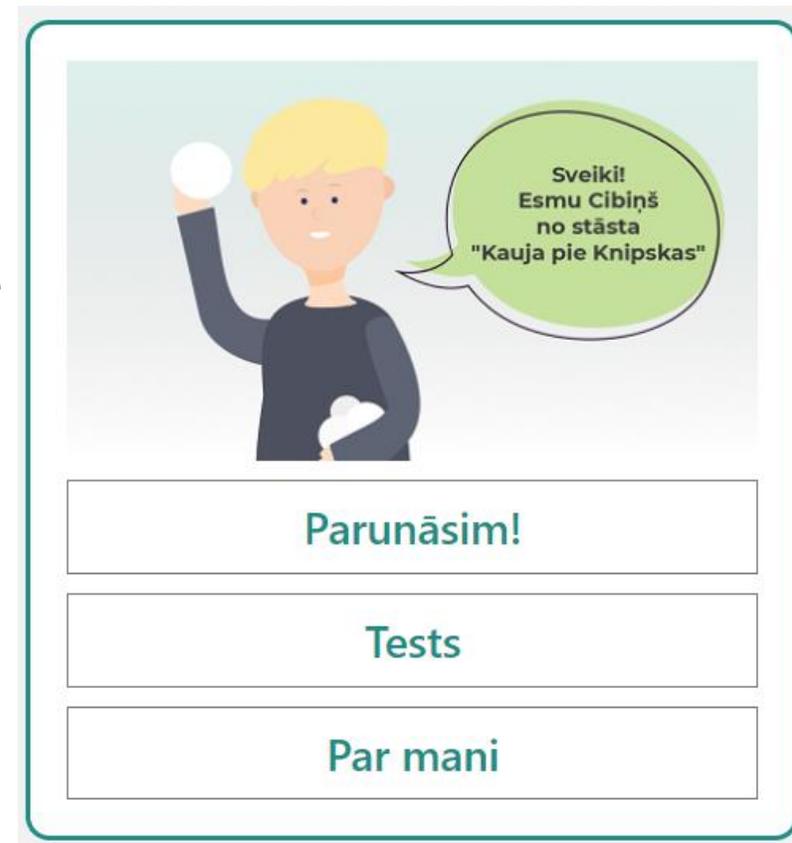
INVESTING IN YOUR FUTURE



Human-computer interaction

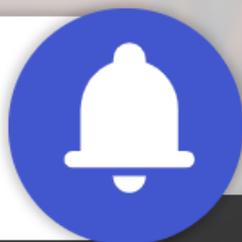
Human-computer interaction

- Latvian-speaking virtual assistants “work” in several public administration institutions, including the Bank of Latvia, the State Revenue Service, the Latvian State Radio and Television Centre and many others.



- Virtual assistants are used in education, call centers and user support services

**Virtuālais asistents uzziņām un
jautājumiem saistībā ar Covid-19**



Virtuālais asistents informācijai par palīdzības
saņemšanu un sniegšanu Ukrainai



SHOWCASE

Virtual Assistant UNA

Solution: Automated Customer Support

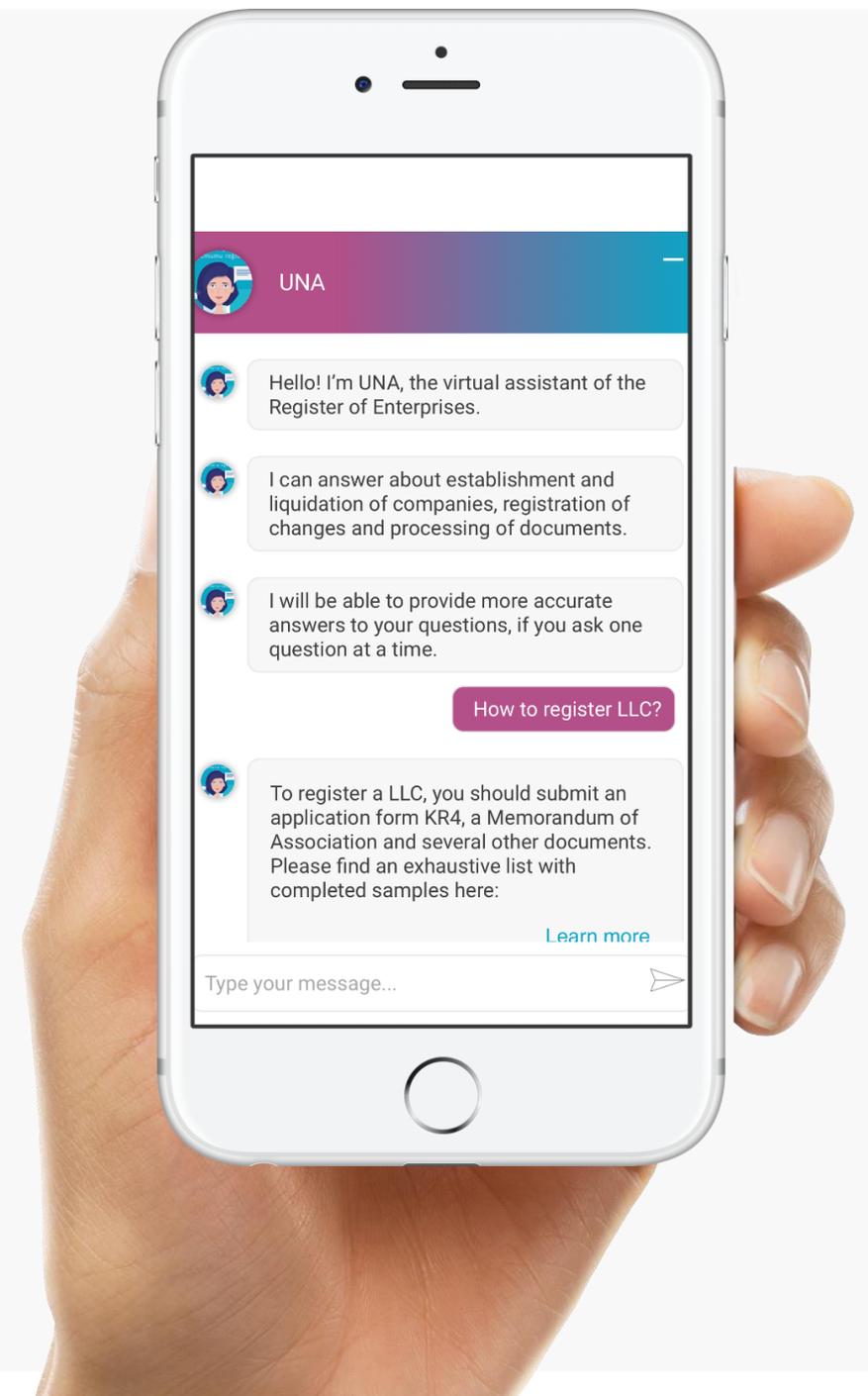
- Helps by providing replies to frequently asked questions in **Latvian** and **English**
- Customer-oriented tool - available at any time and communicates in a clear business-like manner
- Has been trained to understand questions and respond in natural language

30%

OF ALL CLIENT INQUIRIES
ARE SOLVED BY UNA

70 631

UNIQUE CLIENTS
SINCE JUNE, 2018

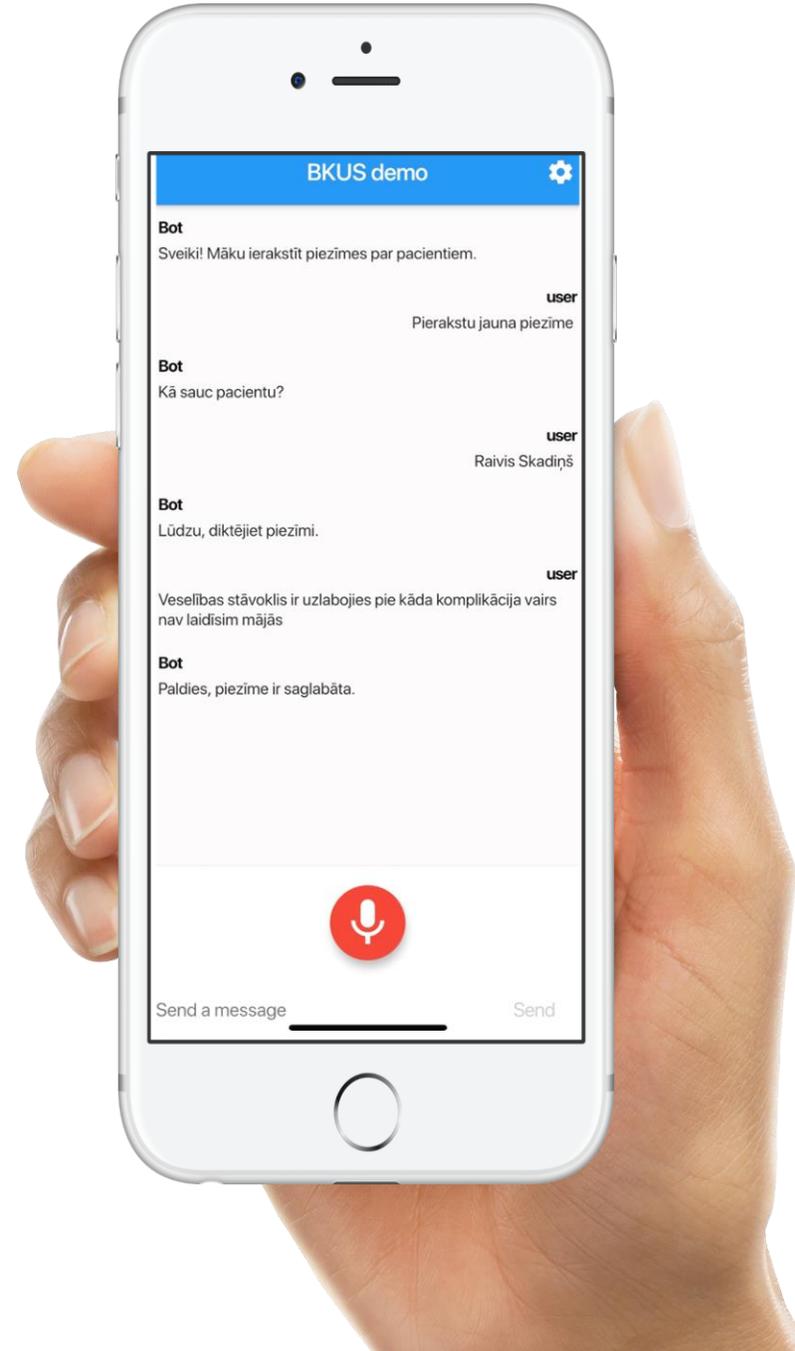


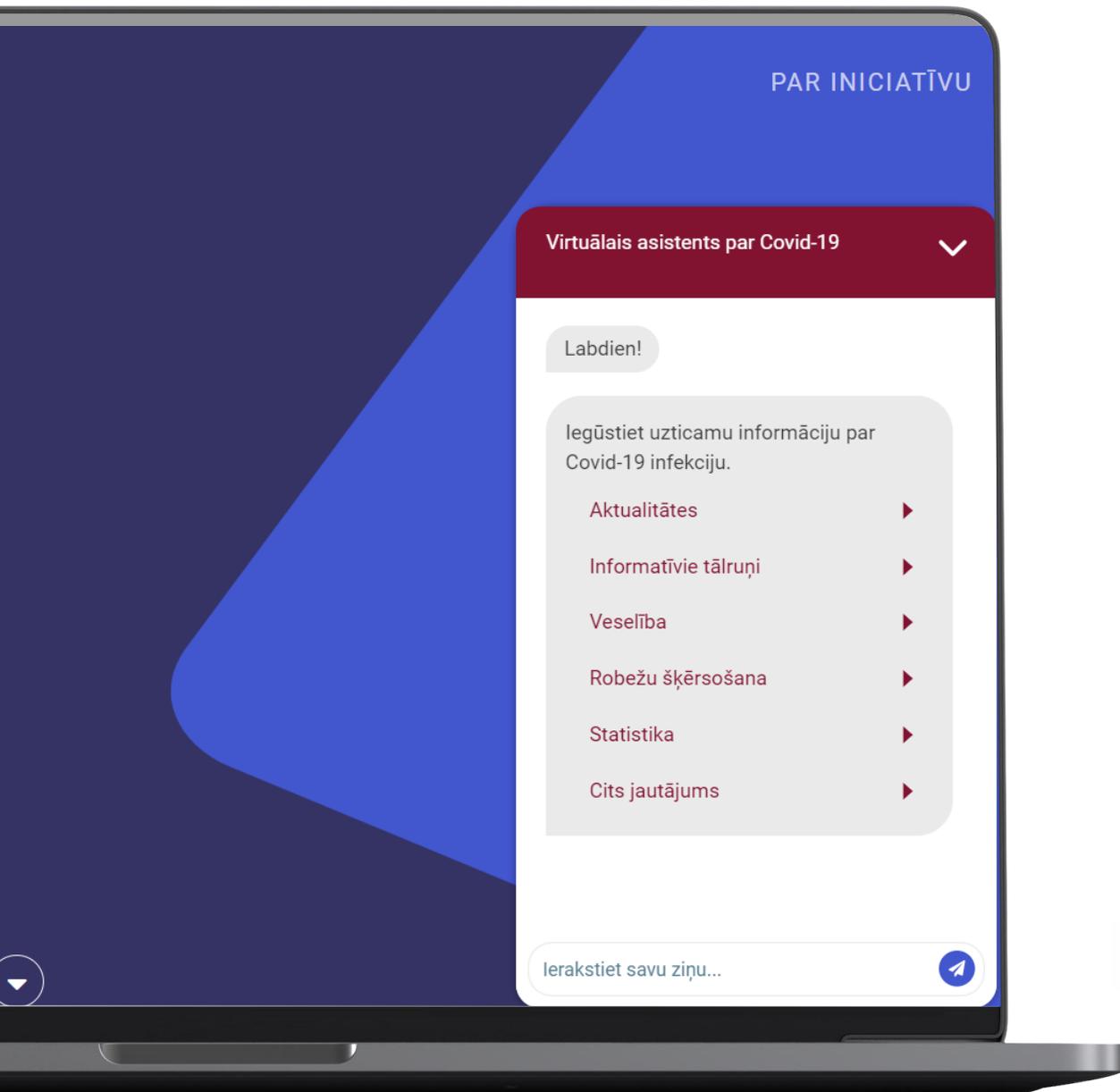
Use Case

Doctors' virtual assistant for Children Clinical University Hospital



Bērnu klīniskā
universitātes
slimnīca





Covidbots.lv

AI enabled language technologies can help citizens in a global pandemic

15+ institutions and organisations deployed solution



Covidbot is deployed on the official websites of 13 municipalities and several NGOs. We have received positive feedback regarding bot as it has reduced the workload of staff responsible for communication and support, providing instant and trustworthy information to citizens' 24/7.

25 000 + users (and counting)



This prior experience of using chatbots as a communication channel with state and local institutions has had an impact on our Covidbot's popularity

Infrastructures

What's New

Corpus

CLARIN Centre Of Latvian Language Resources And Tools

Corpus of Legal Acts of the Republic of Latvia (Likumi) **Author(s):**

Darģis, Roberts

Description:

The corpus contains all legal acts of the Republic of Latvia published on the website likumi. lv (until February 2022).

 This item contains 1 file (335.26 MB).

Publicly Available

Corpus

CLARIN Centre Of Latvian Language Resources And Tools

Latvian Wikipedia **Author(s):**

Darģis, Roberts

Description:

The corpus consists of all information published on Latvian Wikipedia until February 2022.

 This item contains 1 file (108.92 MB).

Publicly Available

Corpus

CLARIN Centre Of Latvian Language Resources And Tools

LVTB - Latvian Treebank v2.10 **Author(s):**

Rituma, Laura ; Pretkalniņa, Lauma ; Saulīte, Baiba ; Nešpore-Bērzkalne, Gunta and Grūzītis, Normunds

Description:

Latvian Treebank (LVTB) is being developed since 2010. It is manually annotated according to a hybrid dependency-constituency grammar model. This version of LVTB contains data used for deriving the corresponding version ...

 This item contains 2 files (38.34 KB).

Publicly Available

CLARIN-LV: most viewed items (total)

Items Viewed

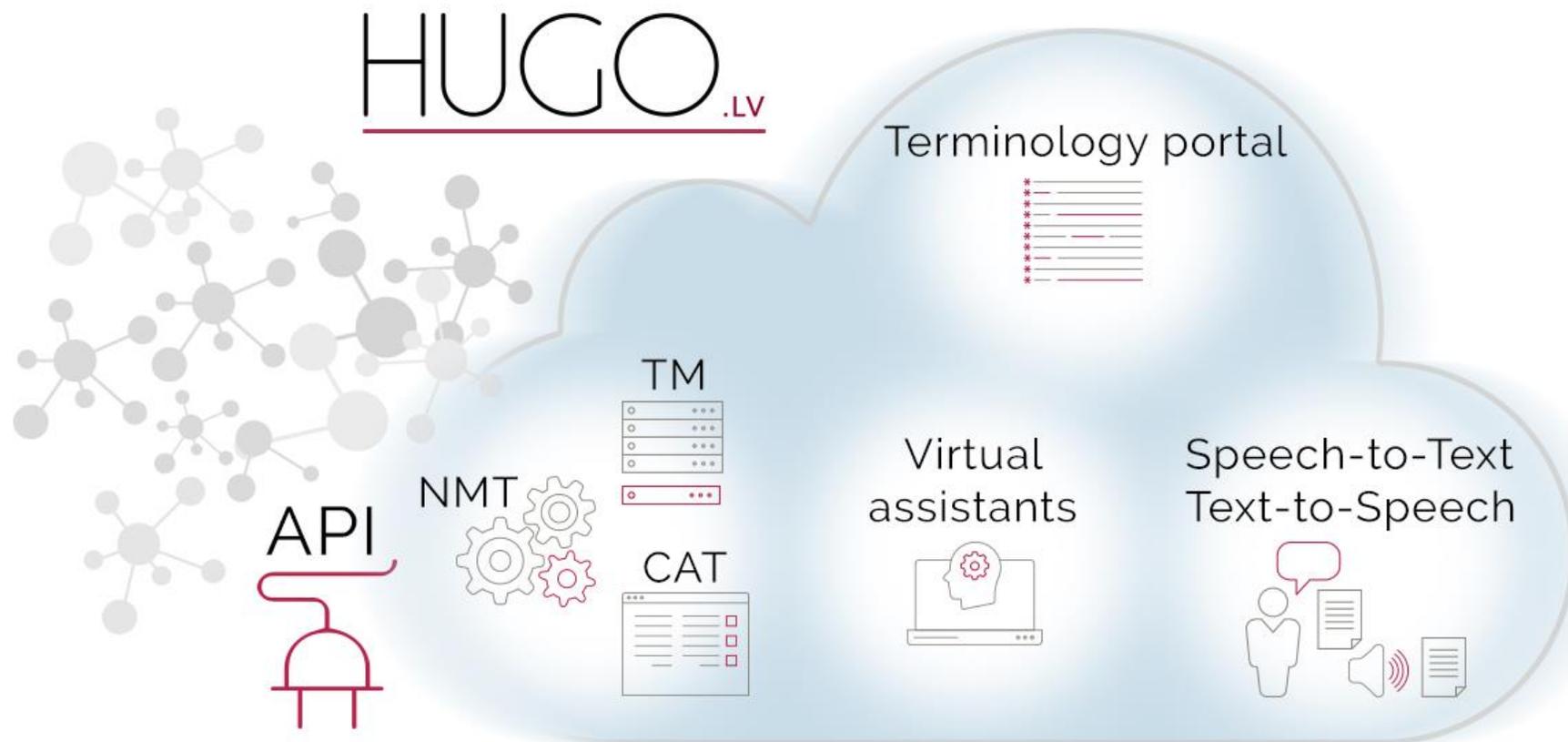
(more than 20 times)

Item/Handle	Number of views
Tēzauris.lv 2020 (Spektors, Andrejs et al) (20.500.12574/9)	2969 [+bots: 3151]
Latvian Treebank v2.5 (Rituma, Laura et al) (20.500.12574/10)	1894 [+bots: 2002]
Balanced Corpus of Modern Latvian (LVK2018) (Levāne-Petrova, Kristīne et al) (20.500.12574/11)	1759 [+bots: 1811]
NLP-PIPE: Latvian NLP Tool Pipeline (Znotiņš, Artūrs) (20.500.12574/4)	1597 [+bots: 1668]
Historical Dictionary of Latvian (Andronova, Everita et al) (20.500.12574/3)	1249 [+bots: 1293]
Latvian AMR Sembank (Znotiņš, Artūrs et al) (20.500.12574/40)	1244 [+bots: 1287]
Latgalian Corpus (MuLa) (Sperga, Ilze et al) (20.500.12574/8)	1226 [+bots: 1270]
Latvian Treebank v2.8 (Rituma, Laura et al) (20.500.12574/55)	1213 [+bots: 1238]
LVBERT - Latvian BERT (Znotiņš, Artūrs) (20.500.12574/43)	1195 [+bots: 1243]
Latvian Language of Science (Dubova, Agnese et al) (20.500.12574/30)	1188 [+bots: 1215]



Culture Information
Systems Centre
Republic of Latvia

HUGO.LV – LANGUAGE TECHNOLOGY PORTAL FOR LATVIAN E-GOV



Hugo.lv | Factsheet

Tilde partnered with the Latvian public administration to create a language technology platform with MT, speech, and terminology management functionalities

- Platform provides **integration** with governmental websites and information systems enabling **dynamic content translation**
- The next phase of the project will introduce a set of virtual assistants to **improve accessibility** to public services

149 M TRANSLATED WORDS USING MT
in 6 months

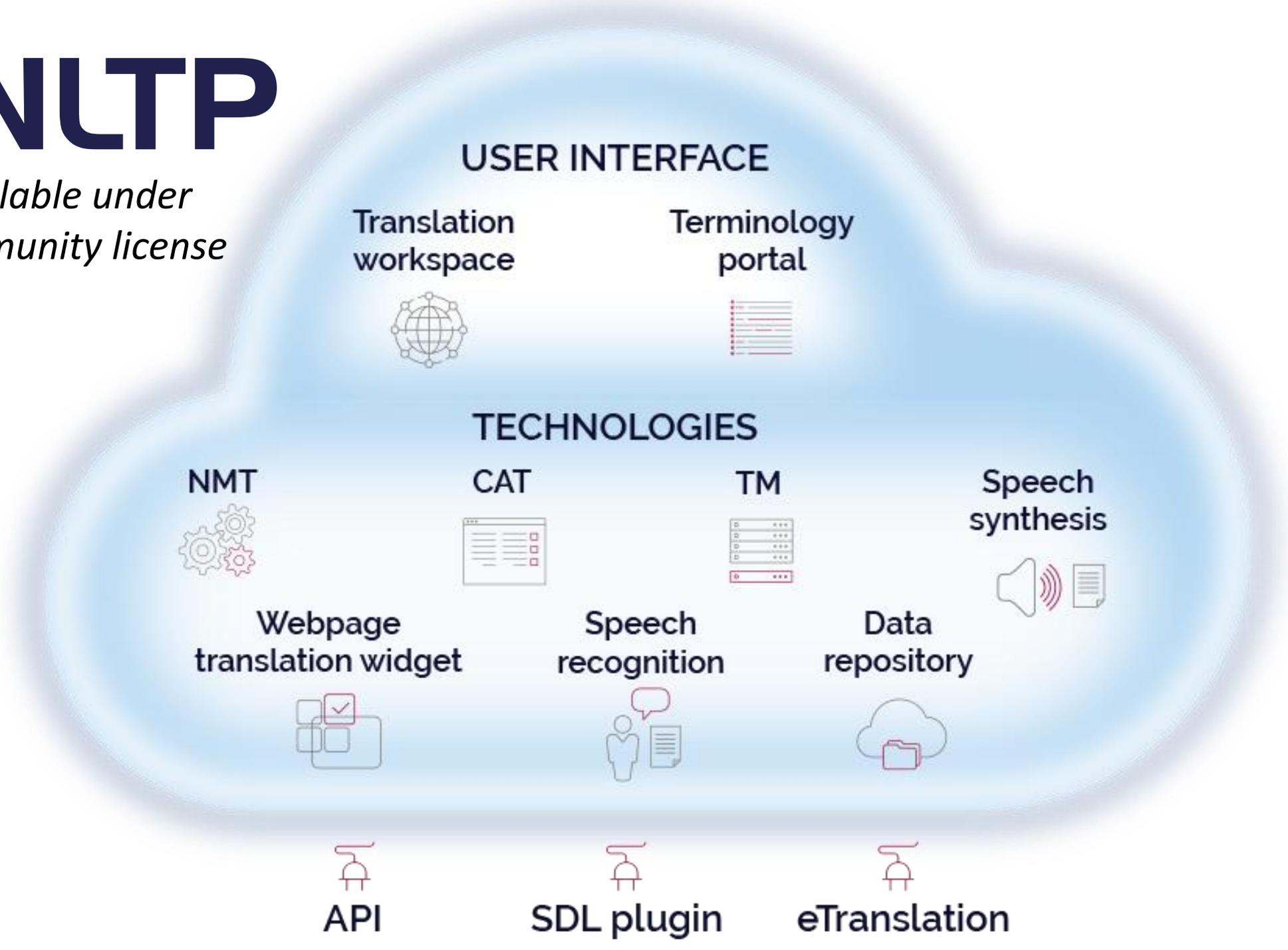
8.1 M RECOGNIZED WORDS
in 6 months

5.4 M SYNTHESISED WORDS
in 6 months



NLTP

*Freely available under
open community license*



Conclusions

- Although the Latvian language is used by a rather small number of speakers, and it is often categorised as less resourced, it is represented rather well not only by different language resources (digital libraries, text and speech corpora, lexicons, etc.) but also by core language technologies.
- Latvian has a reasonably good support for machine translation, speech recognition and synthesis, while solutions that involve deep state-of-the-art natural language understanding, like virtual assistants and text summarisers, are less developed.
- There are still significant gaps with respect to availability, size and technology readiness level (TRL) of language resources.
- There is insufficient amounts of quality corpora, including monolingual corpora, currently available for Latvian, as well as insufficient computational resources, for training large-scale SOTA language models.
- Limited availability of human resources leads to gaps and limitations also in language technology development.
- Gaps and fragmentation in research and development activities related to LT is a result of short, project-based research and development funding.
- Strong national and international support is necessary for further Latvian language research and development activities, including dedicated long-term LT programs that provide equal support for both research and industrial activities.

Thank you!



NATIONAL
DEVELOPMENT
PLAN 2020



EUROPEAN UNION
European Regional
Development Fund

INVESTING IN YOUR FUTURE