

# CLARIN-LV: Many Steps till Operation

Inguna Skadiņa, Ilze Auziņa, Roberts Dargis,  
Eduards Lasmanis and Arnis Voitekāns



NATIONAL  
DEVELOPMENT  
PLAN 2020



EUROPEAN UNION  
European Regional  
Development Fund

INVESTING IN YOUR FUTURE

## CLARIN ERIC

- Preparatory phase (2008-2010): putting everything in place
- Construction phase: (2011-2015): building and populating with tools and resources
- Exploitation phase (2016 - ): CLARIN in full service



## CLARIN-LV

2008-2011: Preparatory Phase.  
MoU signed.

2016: Latvia joins CLARIN ERIC

2018-2022: Construction phase

# Outline: Many Steps till Operation

- Centre and Repository
  - building
  - filling
  - certification
- Consortium
  - building
  - collaboration
  - strengthening collaboration with digital humanities community
- Teaching and training

# CLARIN-LV

[About](#)[Participants](#)[Repository](#)[News](#)[Publications](#)[Contacts](#)

## About

CLARIN stands for "Common Language Resources and Technology Infrastructure". It is a research infrastructure that was initiated from the vision that all digital language resources and tools from all over Europe and beyond are accessible through a single sign-on online environment for the support of researchers in the humanities and social sciences.

In 2012 CLARIN ERIC was established and took up the mission to create and maintain an infrastructure to support the sharing, use and sustainability of language data and tools for research in the humanities and social sciences.

Currently CLARIN provides easy and sustainable access to digital language data (in written, spoken, or multimodal form) for scholars in the social sciences and humanities, and beyond. CLARIN also offers advanced tools to discover, explore, exploit, annotate, analyse or combine such data sets, wherever they are located. This is enabled through a networked federation of centres: language data repositories, service centres and knowledge centres, with single sign-on access for all members of the academic community in all participating countries. Tools and data from different centres are interoperable, so that data collections can be combined and tools from different sources can be chained to perform complex operations to support researchers in their work.

The CLARIN infrastructure is fully operational in many countries, and a large number of participating centres are offering access services to data, tools and expertise. At the same time, CLARIN continues to be constructed in some countries that joined more recently, and CLARIN's datasets and services are constantly updated and improved. On the services page we show the services accessible at this moment and we explain how and by whom the various services can be accessed.



### CLARIN-LV organizes the 10th Conference Human Language Technologies - the Baltic Perspective

The Tenth International Conference "Human Language Technologies - the Baltic Perspective" will take place on October 6-7 at the University of Latvia. The aim of the conference is to provide a forum for the sharing of new ideas and recent advances in human language processing and to promote cooperation between the research communities of computer science, linguistics and digital humanities from the Baltic states and the rest of the world.

[Conference program.](#)

NACIONĀLAIS  
ATTĪSTĪBAS  
PLĀNS 2020



EIROPAS SAVIENĪBA  
Eiropas Reģionālās  
attīstības fonds

Activities of CLARIN Latvia are supported by project of the European Regional Development Fund Nr. 1.1.1.5/18/I/016 University of Latvia and institutes in the European Research Area - Excellency, activity, mobility, capacity

IEGULDĪJUMS TAVĀ NĀKOTNĒ

[www.clarin.lv](http://www.clarin.lv)  
[info@clarin.lv](mailto:info@clarin.lv)

# Repository

# CLARIN DSpace

- After analysis of different options used by other CLARIN centres, CLARIN Dspace (formerly known as LINDAT Dspace) repository system was chosen.
- The main reasons for this choice were:
  - availability of installation resources and documentation
  - rather easy installation and documented
  - approval by many CLARIN centers
  - reliability
  - technical support

# Core Trust Seal and B-centre certification



[Source: [CoreTrustSeal Requirements 2020-22](#)]

## CoreTrustSeal self-assessment template

The CoreTrustSeal self-assessment includes 16 requirements (R1–R16) plus R0 providing contextual information on the repository. This template contains all requirements along with guidance for formulating self-assessment statements and gathering related evidence. Along with the self-assessment statements, the applicant must indicate a compliance level for R1–R16:

- 0. Not applicable
- 1. The repository has not considered this yet
- 2. The repository has a theoretical concept
- 3. The repository is in the implementation phase
- 4. The guideline has been fully implemented in the repository

The [CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020–2022](#) document contains more information on e.g. missing evidence, understandability, non-English language documentation, sensitive/internal documentation, structure and length of responses, and further guidance for certain requirements. Please read it before starting!



# CLARIN-LV repository

## 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

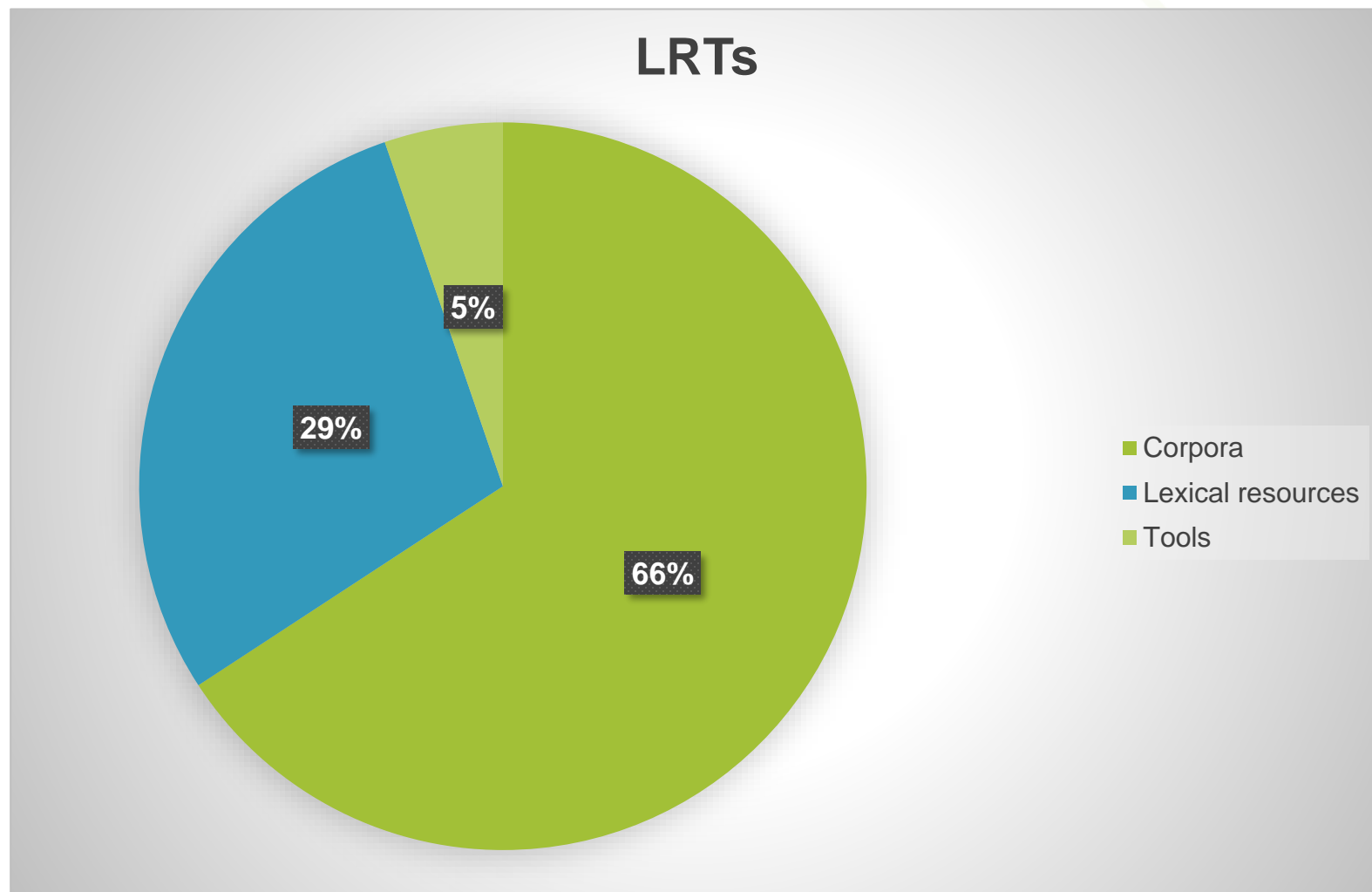
The repository was set up in March, 2020



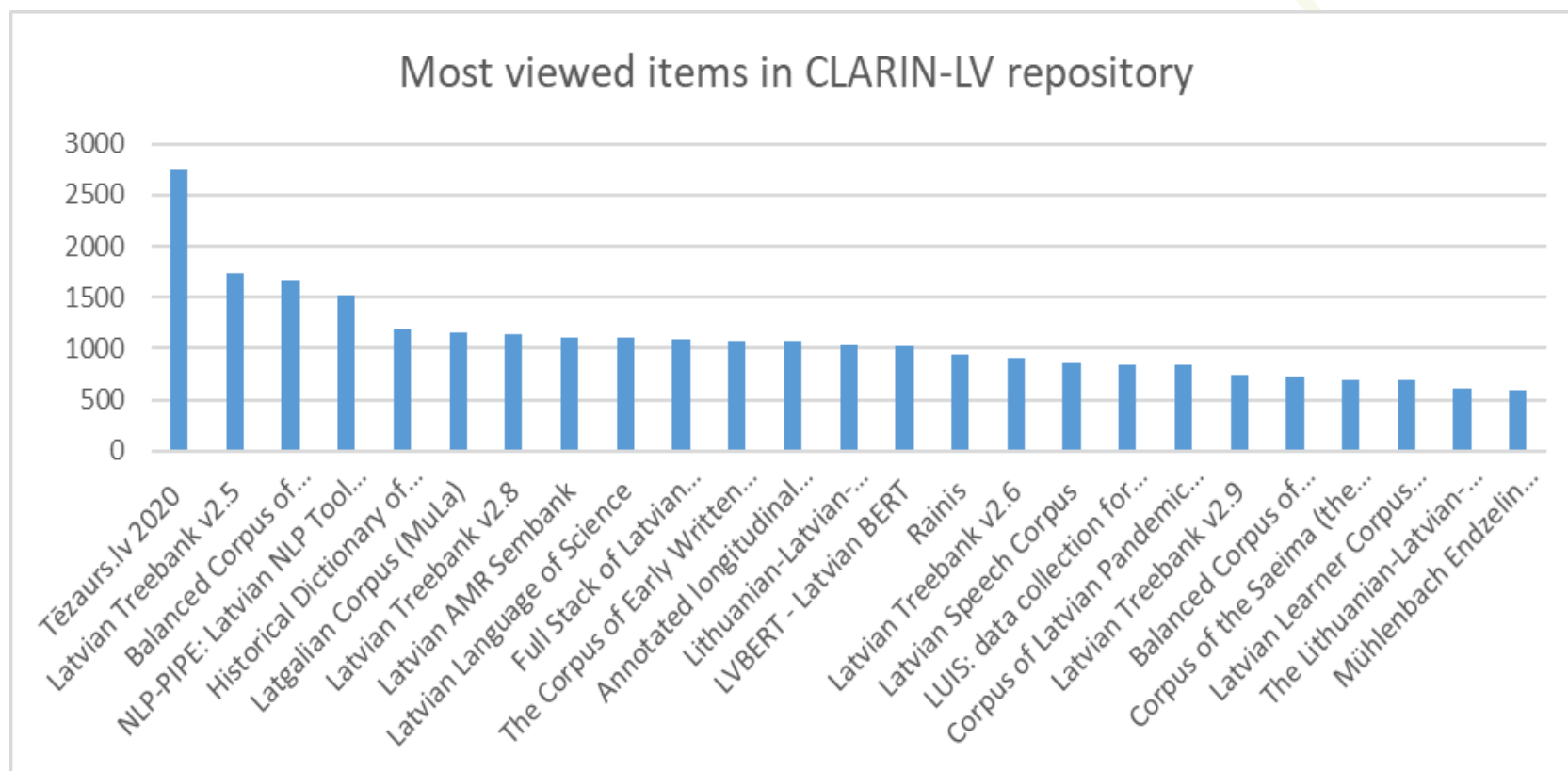
*repository.clarin.lv*



# Language Resources and Tools in CLARIN-LV



# Top 25 most viewed language resources and tools in CLARIN-LV repository



# Corpora

general (6) learner (2) literary (4) parliamentary (1) diachronic (1) specialised (12) representative (7) parallel (1) text (24)  
speech (5) morphology (22) syntax (3) semantics (1) error annotation (2) manually annotated (4) web (2)

## LVK2018

The Balanced Corpus of Modern Latvian

2016–2018, 10M words (12M tokens)

Developers: IMCS UL

[i More](#)

[Q Search](#)

## MuLa2022

Corpus of Contemporary Latgalian Texts 2022

2020–2022, 2M words (2.8M tokens)

Developers: RAT, IMCS UL

[i More](#)

[Q Search](#)

## LaVA

Latvian Language Learner Corpus

2018–2021, 192k words (241k tokens)

Developers: IMCS UL

[i More](#)

[Q Search](#)

## LVTB

Latvian Treebank

2010–2022, 16803 sentences (282 167 tokens)

Developers: IMCS UL

[i More](#)

[Q Search](#)

## Barometrs

Corpus of News Portal Comments

2011–2022, 26M comments (642M tokens)

Developers: RSU, IMCS UL

[i More](#)

[Q Search](#)

## Disertācijas

Corpus of Latvian PhD Theses

2022, 16.7M words (23.4M tokens)

Developers: IMCS UL

[i More](#)

[Q Search](#)

# Parliament Data Corpus

Saeimas stenogrammas Meklēšana

## Par vietni

Šī vietne ir izveidota, lai atvieglotu darbu žurnālistiem, politologiem un citiem, kas interesējas par Saeimas darbu un deputātu (11. Saeimai ieskaitot) korpusu, kas ļauj meklēt gan vienlaicīgi visās Saeimas stenogrammās, gan arī konkrētās sēdēs, laikā.

Lai darbu atvieglotu, katrai konkordancei ir norādīts to teikušais deputāts, viņa dzimšanas gads, frakcija, kurā deputāts ir b, parādīts plašāks konteksts.

Stenogrammas iegūtas no [Saeima.lv](http://Saeima.lv)

## Daži meklējumu piemēri:

### „no zāles”

...Par - 77, pret un atturas - nav. Likumprojekts par steidzamu atzīti. (	<b>No zāles</b>	dep. A.Bērziņš: "Veiksmī
...u svētkiem piešķirt brīvdienu. Tāpēc komisijā tas netika atbalstīts. (	<b>No zāles</b>	dep. Dz.Ābiķis: "Lauku ei
...4, pret un atturas - nav. Likumprojekts pirmajā lasījumā atbalstīts. (	<b>No zāles</b>	dep. J.Ādamsons: "Uz G



Augusts Brigmanis (1952)  
ZZS  
2014/10/02 kārtējā sēde

...kos. Pēo šīs sēdes tālād... Nu, pareizi Čepānes kundze mani palabojā (	<b>No zāles</b>	dep. I.Čepāne: "Es neko netaul!"), ka tālād dalībniekiem paši Dziesm...
...im par atzīmējumu dienu. Tātad - vīrieši, atcerieties par sievietēm! (	<b>No zāles</b>	dep. Dz.Ābiķis: "Dziesm katru dienu!"...
...umprojektā, bet es šo priekšlikumu neatbalstīšu, godātie vīrieši, jo (	<b>No zāles</b>	deputāti vīrieši: "Āāā...!" Zālē smiekli.) 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 uzveru: šeit nav runa, es atvai...

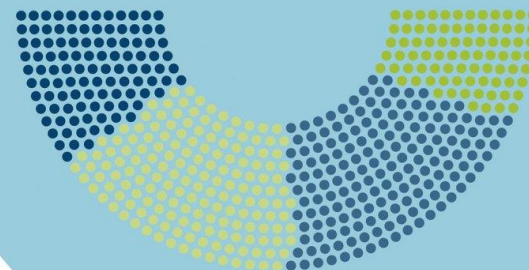
### aizver muti



Ingmārs Līdaka (1966)  
ZZS  
2008/10/15 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) ⓘ

Page 1 of 31 Go Next | Last

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



# Lexical Resources

LexicalConceptualResource

CLARIN Centre Of Latvian Language Resources And Tools

## Historical Dictionary of Latvian

(Latvian Language Institute of the University of Latvia / 2016-12-05)



### Author(s):

Andronova, Everita ; Frīdenberga, Anna ; Siliņa-Piņķe, Renāte ; Trumpa, Anta and Vanags, Pēteris

This item contains no files.

LexicalConceptualResource

CLARIN Centre Of Latvian Language Resources And Tools

## Dictionary of Contemporary Latvian Language (MLVV)

(Latvian Language Institute of the University of Latvia; AiLab IMCS UL / 2019)



### Author(s):

Jērāne, Santa ; et al.

► show everyone

This item contains 1 file (2.68 MB).

Publicly Available

LexicalConceptualResource

CLARIN Centre Of Latvian Language Resources And Tools

## Language consultations: electronic database

(Latvian Language Agency / 2017)



### Author(s):

---

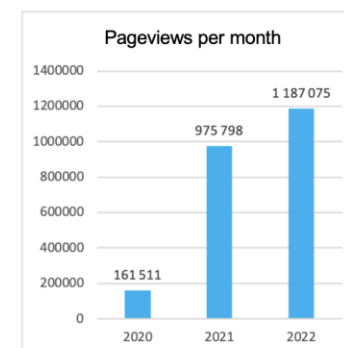
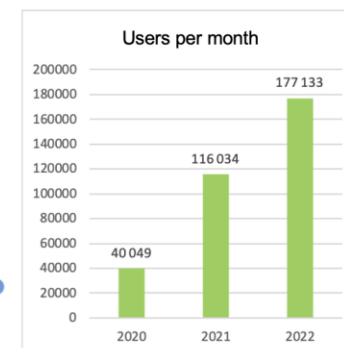
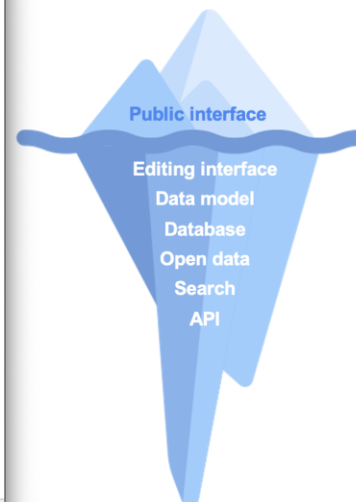
This item contains no files.

LexicalConceptualResource

CLARIN Centre Of Latvian Language Resources And Tools

# The Tēzaurs.lv platform

The screenshot displays the Tēzaurs.lv platform interface. At the top, there are three search bars: [LLVV]: vārds vai frāze, [MLVV]: vārds vai frāze, and [T]: vārds vai frāze. Below the search bars, a sidebar on the left lists various categories like 'Apkaimvalsts', 'valsts', 'valstība', etc. The main content area shows search results for 'valsts'. It includes a list of synonyms and definitions, such as '1. Politiska organizācija, ar kuru tiek realizēta sabiedrības vispārobligāta vadīšana, pastāvošās iekārtas saglabāšana; zeme, teritorija, kurā pastāv šāda organizācija.' and '2. novads, pagasts'. The interface also features a 'Zināt' (Know) button and a 'Dalīties' (Share) button.




<https://tezaurs.lv>



# Tools

Showing 1 through 2 out of 2 results


1 

---

ToolService CLARIN Centre Of Latvian Language Resources And Tools

**NLP-PIPE: Latvian NLP Tool Pipeline**  
(AiLab IMCS UL / 2015-05-05)

**Author(s):**  
Znotiņš, Artūrs


 This item contains no files.

---

ToolService CLARIN Centre Of Latvian Language Resources And Tools


**LVBERT - Latvian BERT**  
(AiLab IMCS UL / 2020)

**Author(s):**  
Znotiņš, Artūrs

 This item contains 3 files (1.51 GB).

Publicly Available

---

1 

# Toolchain for Latvian Language Processing: NLP-PIPE

## NLP-PIPE: Latvian NLP Pipeline as a Service

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 .

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

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	MIl	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

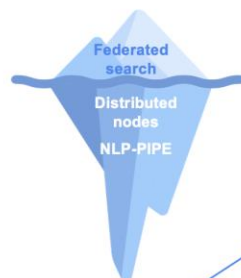
<http://nlp.ailab.lv>

# Consortium and Collaboration with Digital Humanities



# Collaboration with digital humanities

## The Korpus.lv platform

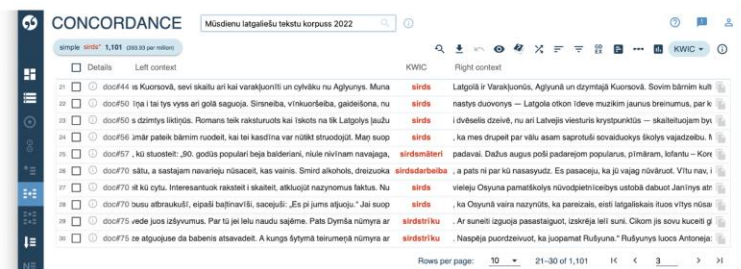
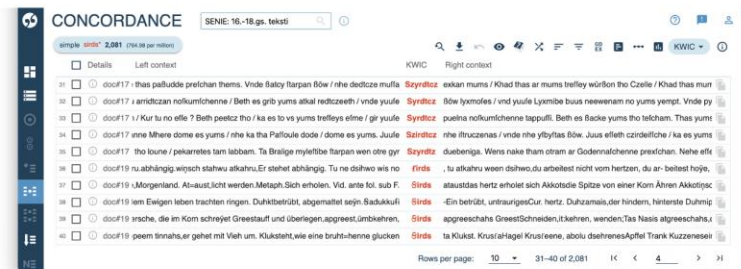


Developer: LU LFM  
Node: nosketch.lnb.lv

Developer: LU LaVi, LU MII, LU HZF  
Node: nosketch.korpus.lv

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

Nacionālā korpusu kolekcija Meklēšana Materiāli Par Korpus.lv LV		
sirds*		
Meklēt		
Korpus	Relatīvais biežums	Absolūtais biežums
uz 1 miljonu		
<b>Pārspridumi</b> Skolēnu pārspridumu korpus	2257	510
Izstrādājis: LU MI, LNB, RTA Meklē: nosketch.korpus.lv		
<b>LatSenRom</b> Latviešu senākie romāni	1420	8234
Izstrādājis: LNB Meklē: nosketch.lnb.lv		
<b>Rainis</b> Raiņa darbu korpus	1196	2752
Izstrādājis: LU MI Meklē: nosketch.korpus.lv		
<b>Isproza8</b> Sieviešu isproza	1083	1272
Izstrādājis: LU LFM Meklē: nosketch.lnb.lv		
<b>Senie</b> Latviešu valodas seno tekstu korpus	765	2081
Izstrādājis: LU LaVi, LU MII, LU HZF Meklē: nosketch.korpus.lv		
<b>LV MED</b> Radioloģisko izmeklējumu transkripciju korpus	616	97
Izstrādājis: LU MI, RAKUS Meklē: nosketch.korpus.lv		
<b>MuLa2012</b> Mūsdienu latgaļu tekstu korpus 2012	426	565
Izstrādājis: LU MI, RTA Meklē: nosketch.korpus.lv		
<b>PanDi</b> Pandēmijas dienasgrāmatas	408	289
Izstrādājis: LU LFM Meklē: nosketch.lnb.lv		
<b>MuLa2022</b> Mūsdienu latgaļu tekstu korpus 2022	394	1101
Izstrādājis: RTA, LU MII Meklē: nosketch.korpus.lv		
<b>Timekls2020</b> Latviešu valodas timekļa korpus	334	164 504
Izstrādājis: LU MI Meklē: nosketch.korpus.lv		
<b>LiLa</b> Lietuvišu-latviešu-lietuvišu parafēzu tekstu korpus	305	1734
Izstrādājis: LU MI, YDU Meklē: nosketch.korpus.lv		
<b>LVK2018</b> Līdzsvarotais mūsdienu latviešu valodas tekstu korpus	270	3321
Izstrādājis: LU MI Meklē: nosketch.korpus.lv		
<b>Emuāri</b> Latviešu valodas emuāru korpus	256	2128
Izstrādājis: LU MI Meklē: nosketch.korpus.lv		



20+ corpora, 10+ institutions, 3+ nodes

The project “Digital Resources for Humanities: Integration and Development” (VPP-IZM-DH-2020/1-0001)

# Consortium (January, 2022)



Institute of Mathematics and Computer Science  
University of Latvia



LATVIJAS  
UNIVERSITĀTE

LAMI

LU LITERATŪRAS,  
FOLKLORAS  
UN MĀKSLAS  
INSTITŪTS



L N B

LATVIJAS  
NACIONĀLĀ  
BIBLIOTĒKA



RĪGAS STRADIŅA  
UNIVERSITĀTE



LIEPĀJAS  
UNIVERSITĀTE

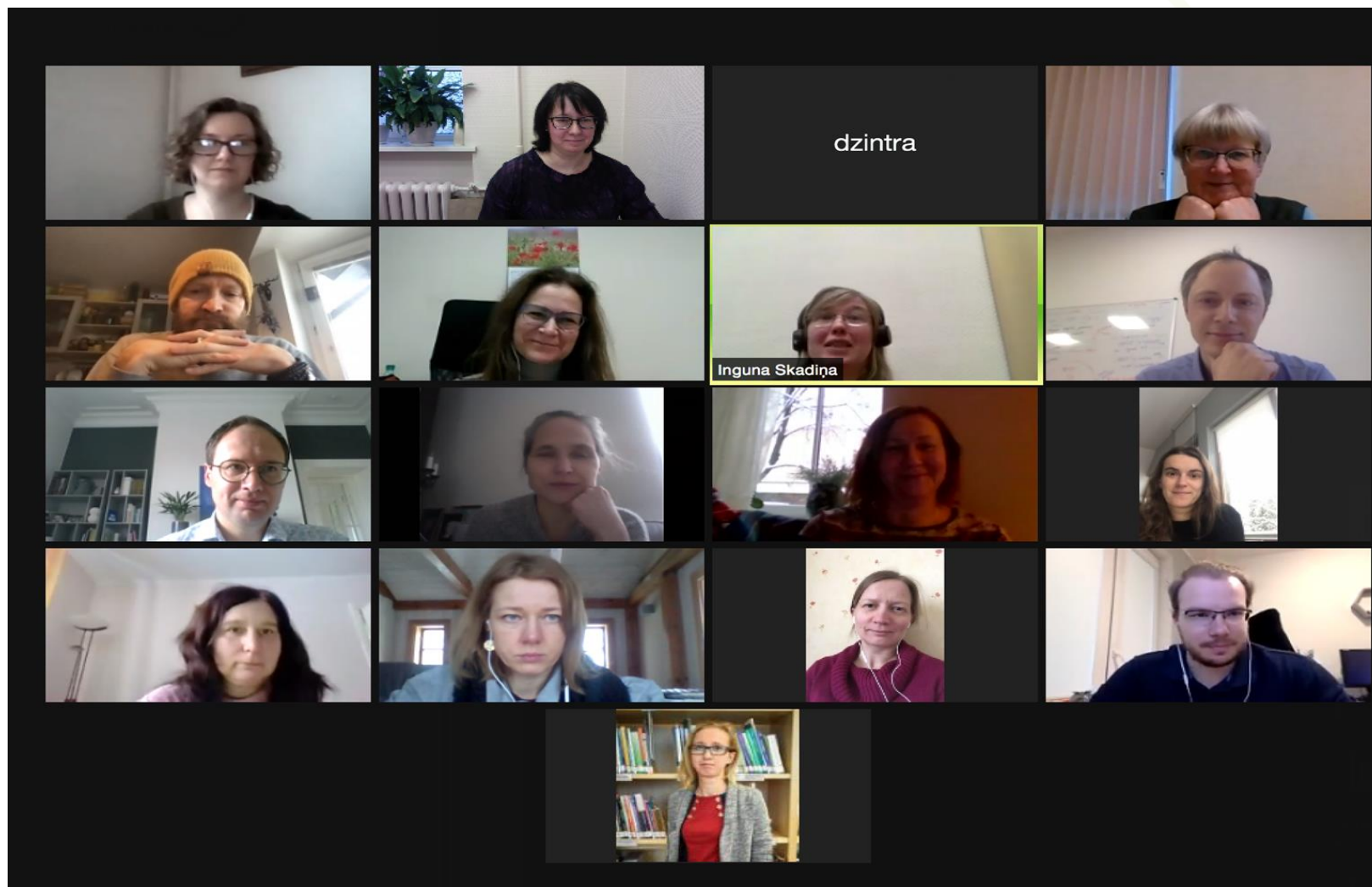
# User involvement



- The conference "Digital Latvian Language Resources and Tools in Common Research Infrastructure"
- The seminar "Tools and Resources for Digital Humanities Research"
- The practical seminar "Balanced Corpus of Modern Latvian "
- The practical seminar "Balanced Corpus of Modern Latvian "
- The practical seminar " How to use Latvian treebank in linguistic studies"
- The conference «Latvian language resources and tools for education»
- The conference «Human Language Technologies - the Baltic Perspective»



# Workshop “How to Register Language Resources in the CLARIN Repository” (January, 2021)





## Item submission

### The Lithuanian-Latvian-Latgalian Dictionary

---

 **Project URL** [http://hipilatlit.ru.lv/eng/products/dictionary\\_info.html](http://hipilatlit.ru.lv/eng/products/dictionary_info.html)

 **Demo URL** <http://hipilatlit.ru.lv/dictionary/eng/dictionary.html>

 **Date issued** 2013

 **Type** [lexicalConceptualResource](#)

 **Publisher** [RTA](#)

Show full item record

Resume

Cancel

Share submission

# SAFMORIL Knowledge Centre

- SAFMORIL - CLARIN Knowledge Centre for Systems and Frameworks for Morphologically Rich Languages
- Collaboration between researchers and developers from the University of Helsinki, University of Tromsø, Vytautas Magnus University and the Institute of Mathematics and Computer Science (University of Latvia) in the area of computational morphology and its application to language processing.



# Teaching activities

# Teaching

Today most of teaching activities concentrate in **the University of Latvia** (Faculty of Humanities, Faculty of Education, Psychology and Art, and Faculty of Social Science)

- In 2017 new *Computational Linguistics* course for Master Students in English philology was started
- In Autumn, 2020 *Computational Linguistics* course started for Master Students in Baltic Philology
- In Autumn, 2020 *Digital Humanities and Social Sciences* for Master students in several study programmes
- In Autumn, 2022 *Language Corpora in the Educational Process* for the professional program for teachers
- Several corpus linguistics courses for Master and Doctoral students in English philology

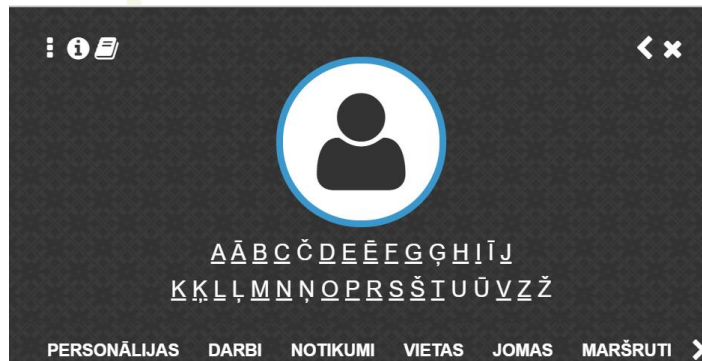
# CLARIN in the Classroom

- In lecture we introduce students with CLARIN research infrastructure, Virtual Language Observatory and resource families.
- We also highlight some LRTs that are more relevant to the research interests of students for particular year of studies.
- In the seminar students present their findings – tool or resource they found in CLARIN VLO and that seemed interesting for her/him.
- We ask students explore other resources than simple corpus

# Student's findings (Autumn, 2022)



**ASV Online  
Toolbox**



**ELAN - Linguistic Annotator**



# Conclusion and next steps

Being among the smallest countries of CLARIN infrastructure and joining CLARIN only in 2016, has some positive and some negative consequences:

- On one hand, many administrative and technical issues were solved when we joined, thus we can use reliable and stable solutions developed by the consortium.
- On the other hand, being among the newest members of CLARIN, clearly revealed a gap in needs, understanding and readiness to serve the SSH community.



# ConcluSSION and next steps

While the targeted project for setting up CLARIN-LV node will end in December, 2022, CLARIN-LV plans to continue activities through several language technology related projects:

- Recently started project “Research on Modern Latvian Language and Development of Language Technology” of the State Research Programme “Letonika – Fostering a Latvian and European Society” will continue supporting cooperation with digital humanities researchers, the development of new resources and their subsequent inclusion in the CLARIN-LV repository.
- Additional funding is also expected from the Latvia's Recovery and sustainability plan, which envisions activities for the development and implementation of high-level skills in language technology.
- New project proposal submitted for the the National Research Programme "Digital Resources of the Humanities"

# Thank you!



NATIONAL  
DEVELOPMENT  
PLAN 2020



**EUROPEAN UNION**  
European Regional  
Development Fund

INVESTING IN YOUR FUTURE