

# Latvian Language Resources and Tools in CLARIN Research Infrastructure

Inguna Skadiņa

National coordinator of CLARIN-LV

05.03.2020.



NATIONAL  
DEVELOPMENT  
PLAN 2020



EUROPEAN UNION

European Regional  
Development Fund

INVESTING IN YOUR FUTURE

# CLARIN - Common Language Resources and Technology Infrastructure

CLARIN 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**.

# CLARIN

VIENOTA VALODAS RESURSU UN TEHNOLOĢIJU INFRASTRUKTŪRA



- Preparatory phase (2008-2010) :  
Put everything in place
- Construction phase: (2011-2015):  
Build and populate with tools and resources
- Exploitation phase (2016 - )  
CLARIN in full service



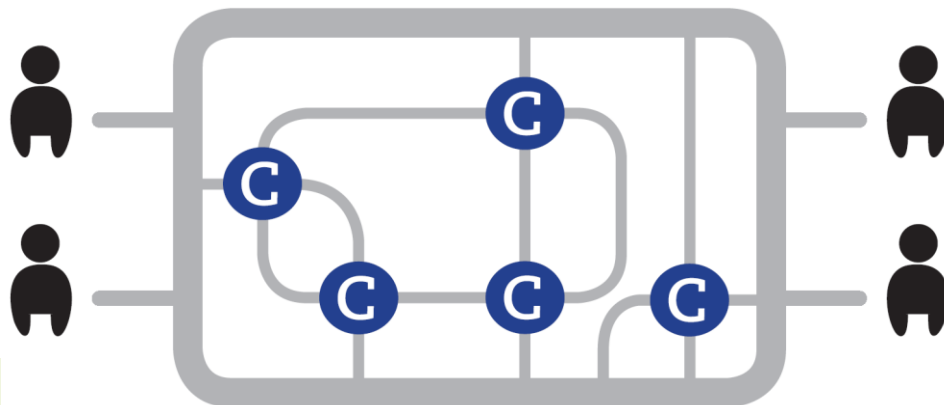
# CLARIN ERIC in members and centres

A consortium of:

- 20 members: AT, BG, CY, CZ, DE, DK, EE, FI, GR, HR, HU, IT, LT, LV, NL, NO, PL, PT, SE, SI
- 4 observers: IS, FR, UK, ZA
- >50 centres

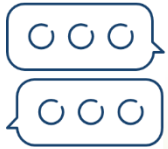


services to researcher



A distributed architecture  
spread all over Europe

# CLARIN - Common Language Resources and Technology Infrastructure



- Easy and sustainable access to **digital language data** (in written, spoken, or multimodal form) for scholars in the **social sciences** and **humanities**, and beyond.



- **Advanced tools** to discover, explore, exploit, annotate, analyse or combine data sets, wherever they are located.

# Latvia in CLARIN

2008-2011: Preparatory Phase. MoU signed.

2016: Latvia joins CLARIN ERIC

2018-2022: Construction phase (funding for national activities)

NATIONAL  
DEVELOPMENT  
PLAN 2020



EUROPEAN UNION  
European Regional  
Development Fund

INVESTING IN YOUR FUTURE

# CLARIN Latvia: language dimension

## CLARIN RESOURCE FAMILIES

Parliamentary corpora are a very important multidisciplinary language resource that can be approached from many research perspectives, including not only political science, but also sociology, history, psychology, and applicative approaches to linguistics, for instance, critical discourse analysis. The good availability of parliamentary proceedings in digitized form and granted access rights to public information in the EU countries have motivated a number of national as well as international initiatives to compile, process and analyse parliamentary corpora.

[www.clarin.eu/resource-families](http://www.clarin.eu/resource-families)

## PARLIAMENTARY CORPORA

18

### parliamentary corpora

1 MULTILINGUAL:  
*Europarl* (21 languages)

17 MONOLINGUAL IN  
14 LANGUAGES:

1 Czech	1 Finnish
1 Danish	3 German
2 English	1 Greek
1 French	1 Lithuanian
1 Estonian	2 Norwegian
	1 Polish
	1 Portuguese
	1 Slovenian
	1 Swedish

### AVAILABILITY

- 3 through a concordancer
- 9 for download
- 4 both

### SIZE

- 7 small (<10 million tokens)
- 7 medium (10–100 million tokens)
- 3 large (>100 million tokens)

### ANNOTATION

- 8 PoS-tagged
- 7 lemmatised

Latvian language resources and tools are represented in CLARIN language resource families:

- Historical corpora
- L2 learner corpora
- Parliamentary corpora
- Spoken corpora
- Lexical resources

## Par CLARIN

CLARIN (Common Language Resources and Technology Infrastructure –Vienota valodas resursu un tehnoloģiju infrastruktūra) ir daļa Eiropas pētniecības infrastruktūra (ERIC), kas integrē darbu, zināšanas, resursus, rīkus, repositorijus un servišus, ko piedāvā daudzie CLARIN centri un nacionālie konsorģiji.

CLARIN mērķis ir novērst sadrumstalotību valodas resursu jomā un padarīt valodas resursus pieejamus visu jomu, it īpaši humanitāro un sociālo zinātņu zinātniekiem un studentiem, izmantojot vienotu pierakstīšanas. CLARIN piedāvā ilgtermiņa risinājumus un tehnoloģiju servišus digitālo valodas datu un rīku izvietošanai, savienošanai, analīzei un atbalstīšanai. CLARIN atbalsta pētniekus, kas vēlas iesaistīties vadošos datos balstītos pētījumos, lai veicinātu patiesi daudzvalodīgu Eiropas pētniecības telpu.

CLARIN infrastruktūra daudzās valstīs pilnībā darbojas un daudzi centri piedāvā piekļuvi datiem, rīkiem un zināšanām. CLARIN datu kopas un serviši tiek nepārtraukti atjaunināti un pilnveidoti. Dažās valstīs, kas CLARIN pievienojās nesen, CLARIN infrastruktūra tiek veidota. Latvija pievienojās CLARIN ERIC 2016. gada 1. jūnijā.



### Konference "Latviešu valodas digitālie resursi un rīki vienotā pētniecības infrastruktūrā"

Aicinām piedalīties Latvijas Universitātes 78. starptautiskās zinātniskās konferences datorlingvistikas sekcijas sēdē "Latviešu valodas digitālie resursi un rīki vienotā pētniecības infrastruktūrā", kas notiks **2020. gada 5. martā pulksten 12.00 LU Bibliotēkas 2. stāva zālē (Microsoft Inovāciju centrā) Kalpaka bulvārī 4.**

#### Konferences programma

11.30–12.00 Reģistrācija

12.00–12.05 Atklāšana

12.05–12.20 *Aleksandrs Mārtiņš Blūms* (Latvijas Republikas Izglītības un zinātnes ministrija) Valodas tehnoloģijas digitālo humanitāro zinātņu ekosistēmā



# repository.clarin.lv



Repository

Corpus Search

About



Login

## Deposit Free and Safe

License of your Choice (Open licenses encouraged)

Easy to Find

Easy to Cite



Search

[Advanced Search](#)

### Author

Levāne-Petrova, Kristīne (4)

Darģis, Roberts (3)

Grūzītis, Normunds (3)

Pretkalniņa, Lauma (3)

Saulīte, Baiba (3)

[... View More](#)

### Subject

Abstract Meaning Rep ... (1)

child language (1)

Coreferences (1)

corpus (1)

dependencies (1)

[... View More](#)

### Language (ISO)

Latvian (7)

Latgalian (1)

Lithuanian (1)

# CLARIN Latvia: language dimension

## What's New

### Corpus

CLARIN LATVIA

#### Latvian Treebank v2.5



**Author(s):**

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

**Description:**

Latvian Treebank (LVTB) is in development since 2010. Corpus is annotated manually according to SemTi-Kamols hybrid dependency-constituency grammar model. This version of the treebank contains the data used for deriving ...

 This item contains 1 file (16.44 MB).

Publicly Available

### LexicalConceptualResource

CLARIN LATVIA

#### Tēzauris.lv (Winter 2020)



**Author(s):**

Spektors, Andrejs ; Pretkalniņa, Lauma ; Grūzītis, Normunds

**Description:**

Tezaurs is a machine-readable lexicon and an online dictionary (<https://tezaurs.lv/>) for Latvian. The initial human-oriented version of this resource was made publicly available in 2009, comprising more than 125,000 entries ...

 This item contains 1 file (14.36 MB).

Publicly Available

### Top Last Week

Corpus

CLARIN LATVIA

#### Full Stack of Latvian Language Resources for NLU



**Author(s):**

Grūzītis, Normunds ; Pretkalniņa, Lauma ; Saulīte, Baiba ; Rītuma, Laura ; Nešpore-Bērzkalne, Gunta ; Paikens, Pēteris ; Auziņa, Ilze ; Znotiņš, Artūrs ; Levāne-Petrova, Kristīne ; Darģis, Roberts

**Description:**

This repository contains a multilayer text corpus of Latvian. The multilayer corpus is anchored in cross-lingual state-of-the-art representations: Universal Dependencies (UD), FrameNet, PropBank and Abstract Meaning ...

This item contains no files.

Publicly Available

Corpus

CLARIN LATVIA

#### Lithuanian-Latvian-Lithuanian Parallel Corpus (LILA)



**Author(s):**

Utka, Andrius ; Levāne-Petrova, Kristīne ; Vēvere, Daira ; Rābante-Buša, Guna ; Kovalevskaitė, Jolanta ; Rimkutė, Erika

**Description:**

The Latvian-Lithuanian parallel corpus represents the language of the Independence period (starts in 1990), includes fictional and non-fictional texts, periodicals, and documents. The total size of corpus is 8 million ...

This item contains no files.

Corpus

CLARIN LATVIA

#### Annotated longitudinal corpus of Latvian children's language



**Author(s):**

Auziņa, Ilze ; Darģis, Roberts ; Rābante-Buša, Guna ; Levāne-Petrova, Kristīne ; Saulīte, Baiba

**Description:**

This collection contains three longitudinal corpora of monolingual Latvian-speaking children, and one longitudinal corpus of simultaneous Latvian-Russian bilingual child. Participants were recorded for 30 minutes each week ...

This item contains no files.

# Advanced Search

Corpus

CLARIN LATVIA

## Full Stack of Latvian Language Resources for NLU

(Ailab IMCS UL / 2019-11-30)



### Author(s):

Grūzītis, Normunds ; Pretkalniņa, Lauma ; Saulīte, Baiba ; Rituma, Laura ; Nešpore-Bērzkalne, Gunta ; Paikens, Pēteris ; Auziņa, Ilze ; Znotiņš, Artūrs ; Levāne-Petrova, Kristīne ; Dargis, Roberts

 This item contains no files.

Publicly Available

LexicalConceptualResource

CLARIN LATVIA

## Tēzauris.lv (Winter 2020)

(Ailab IMCS UL / 2019-12-20)



### Author(s):

Spektors, Andrejs ; Pretkalniņa, Lauma ; Grūzītis, Normunds

 This item contains 1 file (14.36 MB).

Publicly Available

ToolService

CLARIN LATVIA

## NLP-PIPE: Latvian NLP Tool Pipeline

(Ailab IMCS UL / 2015-05-05)



### Author(s):

Znotiņš, Artūrs

 This item contains no files.

Search

×

# Advanced Search

## Limit your search

### Author

Levāne-Petrova, Kristīne (4)

Darģis, Roberts (3)

Grūzītis, Normunds (3)

Pretkalniņa, Lauma (3)

Saulīte, Baiba (3)

Auziņa, Ilze (2)

Nešpore-Bērzkalne, Gunta (2)

Rituma, Laura (2)

Rābante-Buša, Guna (2)

Znotiņš, Artūrs (2)

Briška, Anna (1)

Kovalevskaitē, Jolanta (1)

Paikens, Pēteris (1)

Prokatniece, Kristīne (1)

Rimkutē, Erika (1)

Spektors, Andrejs (1)

Sperga, Ilze (1)

Utka, Andrius (1)

Vēvere, Daira (1)

### Subject

### Language (ISO)

Latvian (7)

Showing 1 through 8 out of 8 results

1



#### Corpus

CLARIN LATVIA

### Full Stack of Latvian Language Resources for NLU

(Ailab IMCS UL / 2019-11-30)



#### Author(s):

Grūzītis, Normunds ; Pretkalniņa, Lauma ; Saulīte, Baiba ; Rituma, Laura ; Nešpore-Bērzkalne, Gunta ; Paikens, Pēteris ; Auziņa, Ilze ; Znotiņš, Artūrs ; Levāne-Petrova, Kristīne ; Darģis, Roberts

This item contains no files.

Publicly Available

#### LexicalConceptualResource

CLARIN LATVIA

### Tēzauris.lv (Winter 2020)

(Ailab IMCS UL / 2019-12-20)



#### Author(s):

Spektors, Andrejs ; Pretkalniņa, Lauma ; Grūzītis, Normunds

This item contains 1 file (14.36 MB).

Publicly Available

#### ToolService

CLARIN LATVIA

### NLP-PIPE: Latvian NLP Tool Pipeline

(Ailab IMCS UL / 2015-05-05)



#### Author(s):

Znotiņš, Artūrs

# Advanced Search

Limit your search

Author ▾


Subject ▾

Language (ISO) ▾  
Latvian (7)  
Latgalian (1)  
Lithuanian (1)



Type ▾  
corpus (6)  
lexicalConceptualResource (1)  
toolService (1)

Contain Files ▾



Showing 1 through 8 out of 8 results

1 

**Corpus** CLARIN LATVIA

**Full Stack of Latvian Language Resources for NLU**   
(Ailab IMCS UL / 2019-11-30)  
**Author(s):**  
Grūzītis, Normunds ; Pretkalniņa, Lauma ; Saulīte, Baiba ; Rituma, Laura ; Nešpore-Bērzkalne, Gunta ; Paikens, Pēteris ; Auziņa, Ilze ; Znotiņš, Artūrs ; Levāne-Petrova, Kristīne ; Darģis, Roberts  
 This item contains no files. Publicly Available

**LexicalConceptualResource** CLARIN LATVIA

**Tēzauris.lv (Winter 2020)**   
(Ailab IMCS UL / 2019-12-20)  
**Author(s):**  
Spektors, Andrejs ; Pretkalniņa, Lauma ; Grūzītis, Normunds  
 This item contains 1 file (14.36 MB). Publicly Available

# Usage scenarios: On-line access

## NLP-PIPE: Latvian NLP Tool Pipeline

### ToolService

### NLP-PIPE: Latvian NLP Tool Pipeline

(Ailab IMCS UL / 2015-05-05)

#### Author(s):

Znotiņš, Artūrs

This item contains no files.

Samples ...

5. martā tiksīmies konferencē LU Bibliotēkā.

Go

tokenizer ×

morpho ×

parser ×

ner ×

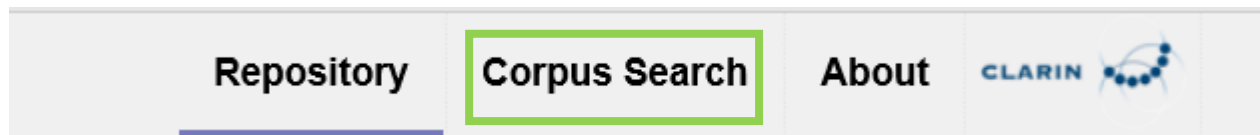
NER

CONLL

JSON

INDEX	FORM	LEMMA	UPOSTAG	XPOSTAG	FEATS	HEAD	DEPREL
#text=5. martā tiksīmies konferencē LU Bibliotēkā .							
1	5.	5.	ADJ	xo	NumType=Ord < >	2	amod
2	martā	marts	NOUN	ncmsl1	Case=Loc Gen < >	3	obl
3	tiksīmies	tikties	VERB	vmyif_11pan	Evident=Fh Mo < >	0	root
4	konferencē	konference	NOUN	ncfsl5	Case=Loc Gen < >	3	obl
5	LU	LU	PROPN	yp	Abbr=Yes < >	6	nmod
6	Bibliotēkā	bibliotēka	NOUN	ncfsl4	Case=Loc Gen < >	3	obl
7	.	.	PUNCT	zs	< >	3	punct

# Usage scenarios: Corpus search



Query Corpora Save Concordance Filter Frequency Collocations View Help

Corpus: LVK2018 (beta) | Query: stāvēt, pie (6 hits)

Hits: 6 | i.p.m.: 0.49 (related to the whole corpus) | ARF: 3.97 | Result is sorted

1 / 1

Line selection: simple ▾

<input type="checkbox"/>	<a href="#">doc#101</a>	pārāk ilgas pēdējās aizkavēšanās ruma mucas tuvumā Puspīrāgam atkal iznāca	<b>stāvēt pie</b>	stūres . Pūta spirts dienvidrietenis , un « Speķa Lode
<input type="checkbox"/>	<a href="#">doc#15350</a>		<b>stāvēt pie</b>	operāciju galda , stundām skatoties dažu centimetru laukumā , un
<input type="checkbox"/>	<a href="#">doc#15802</a>	, pārcilāt it visas lietas istabā , atvērt logu un	<b>stāvēt pie</b>	tā , līdz sirds piesalst pie ribām un pārstāj pukstēt
<input type="checkbox"/>	<a href="#">doc#19048</a>	gadiem darīja Hilarija Klintone . Jau viņai toreiz pārmeta nevēlēšanos	<b>stāvēt pie</b>	plīts . Mišeļa Obama tagad mierīgi atzīst , ka neprot
<input type="checkbox"/>	<a href="#">doc#19107</a>	smagi strādāt , lai nopelnītu iztiku , uzkopt telpas ,	<b>stāvēt pie</b>	plīts . Psihologi izpētījuši , ka tikai 30 % cilvēku
<input type="checkbox"/>	<a href="#">doc#20485</a>	smaržo gatavais pamatēdiens . Vai leģendārajai multiplikatorei vienmēr atlicis laika	<b>stāvēt pie</b>	katliem ? " Atzišos , apmēram divdesmit savus dzīves gadus

1 / 1

LVK2018 (beta)	12M		<a href="#">Details</a>
morphocorpus	220k		<a href="#">Details</a>
MuLa	1M		<a href="#">Details</a>
Rainis	2M		<a href="#">Details</a>
Saeima	22M		<a href="#">Details</a>
Vidusskolēnu pārspridumi	226k		<a href="#">Details</a>



# Usage scenarios: Access from another CLARIN node

LVTB | AiLab

LVTB - Latvian dependency-constituency treebank v 2.5

lindat.mff.cuni.cz/services/pmltq/#!/treebank/lvtb25/query/IYWgdg9gJgpgBAKANpwLYHoAuWdmcB+cA5AHPHEA0icNtoks1tcKG2ehpADkXALr8A3AiA/result/s/

LINDAT Repository Corpus Search TreeQuery Treex More Apps About CLARIN

Home > LVTB - Latvian dependency-constituency treebank v 2.5 My queries Public queries Query Help Login

a-node [ m/tag ~ '^n', a-node [ m/tag ~ '^p' ] ];

Execute query w/o Filters Suggest (2)

Result: 22 / 100 1 a-node 2 a-node Context: 2 / 5

Kādā atpūtas brīdī, kad imperators sēdējis zem tējas koka un dzēris vārītu ūdeni, sacēlies vējš un imperatora krūzē iepūtis pāris tējas koka lapas.

```
graph TD
    sent["ID: a-c2-p4s2  
sent"] --> punct["punct  
.zs  
[.]"]
    sent --> pred["pred  
crdParts"]
    pred --> crdPart1["crdPart  
xPred  
vm0rp000san[actpmn]"]
    pred --> subj["subj  
vējš  
ncmsn1  
[vējš]"]
    pred --> conj["conj  
un  
cc  
[un]"]
    pred --> crdPart2["crdPart  
xPred  
vm0rp000san[actpmn]"]
    crdPart1 --> basElem1["basElem  
sacēlies  
vmypdmsnasnpn  
[sacelties]"]
    crdPart1 --> auxVerb1["auxVerb  
reduction:  
vcnrpii00an(esot)"]
    crdPart2 --> adv["adv  
krūzē  
ncfsl5  
[krūze]"]
    crdPart2 --> basElem2["basElem  
iepūtis  
vmnpdmsnasnpn  
[iepūst]"]
    crdPart2 --> auxVerb2["auxVerb  
reduction:  
vcnrpii00an(esot)"]
    crdPart2 --> obj["obj  
lapas  
ncfpa4  
[lapa]"]
    crdPart2 --> obj2["obj  
lapas  
ncfpa4  
[lapa]"]
    crdPart1 --> adv2["adv  
brīdī  
ncmsl2  
[brīdis]"]
    adv2 --> attr["attr  
Kādā  
pq0msln  
[kāds]"]
    adv2 --> attr2["attr  
atpūtas  
ncfsg4  
[atpūta]"]
    adv2 --> attrCl["attrCl  
subrCl"]
```

# Usage scenarios: Download

LexicalConceptualResource

CLARIN LATVIA

[Tēzauris.lv \(Winter 2020\)](#)

(Ailab IMCS UL / 2019-12-20)

**Author(s):**

Spektors, Andrejs ; Pretkalniņa, Lauma ; Grūzītis, Normunds

 This item contains 1 file (14.36 MB).



Publicly Available

## Files in this item



Download instructions for command line

This item is **Publicly Available** and licensed under:  
Creative Commons - Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

<b>Name</b>	tezaurs.zip
<b>Size</b>	14.36 MB
<b>Format</b>	application/zip
<b>Description</b>	A machine-readable lexicon of Latvian.
<b>MD5</b>	febee0e95bf9e11148a376e2ab868b05



 Download file

 Preview

# Usage: Citation



## NLP-PIPE: Latvian NLP Tool Pipeline

“ Please use the following text to cite this item or export to a predefined format:

BIBTEX CMDI

Znotiņš, Artūrs, 2015, *NLP-PIPE: Latvian NLP Tool Pipeline*, CLARIN-LV digital library at MII, University of Latvia, <http://hdl.handle.net/20.500.12574/4>.



Share:  

## How to Deposit

Only authenticated users can deposit items. If you cannot find your home organisation in the Login dialog list of organisations then register at [clarin.eu](http://clarin.eu) and authenticate using "clarin.eu website account". In case you cannot use any authentication method above or if you encounter a problem, do not hesitate to contact our [Help Desk](#) and we can create a local account for you.

### Step 1: Login

To start a new submission you have to login first. Click Login under My Account in the right menu panel.

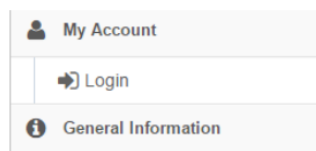


Fig1. Menu Login

### Step 2: Starting a new submission

Now you have a new menu item 'Submissions' under My Account. Click on Submissions to go to the Submissions screen.

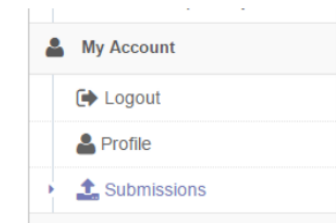


Fig2. Menu Submissions

Now you should be on the main Submission and Workflow tasks page where you can view your incomplete/archive submissions. Click on the 'Start another submission' link to start a new Submission.

#### Submissions & workflow tasks

##### Unfinished submissions

These are incomplete item submissions. You may also [start another submission](#).

	Title	Collection	Submitter
<input type="checkbox"/>	This is a test	External Data	email: Amir Kamran

Remove selected submissions



What can you do?



Browse

> All of the Repository

My Account

Login

General Information

Deposit

Cite

Submission Lifecycle

FAQ

About

Help Desk

Select your home organisation below. This is usually the organisation where you work or study. Signing in here will allow you to access certain CLARIN resources and services which are only available to users who have logged in. If you cannot find your organisation in the list below, please select the clarin.eu website account and use your CLARIN website credentials. If you don't have such credentials you can register an account [here](#). For questions please contact [spf@clarin.eu](mailto:spf@clarin.eu).

Previously chosen home organisation

University of Latvia

 Latvia

Home organisation list

All countries ▼

⌵ ⌵

clarin.eu website account

 European Union

# Deposit and access to licensed resources

Home organisation list

All countries ▼

⌵ ⌵

Art Academy of Latvia

 Latvia

Institute of Mathematics and Computer Science, University of Latvia

 Latvia

Latvia University of Agriculture

 Latvia

Latvian Academy of Culture

 Latvia

Latvian Academy of Music

 Latvia

University of Latvia

 Latvia

---



---



---



---



---



---



---




---

# VLO: Virtual Language Observatory

Search through 9 records



Showing 8 results within selection for  

Results per page:

Use the categories below to limit the search results to those matching the selected value(s).

Language



Collection



Resource type



Format



Availability



Search options

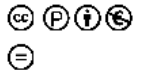


## Full Stack of Latvian Language Resources for NLU

(Part of Language resources and tools of AiLab IMCS UL)

This repository contains a multilayer text corpus of Latvian. The multilayer corpus is anchored in cross-lingual state-of-the-art representations: Universal Dependencies (UD), FrameNet, PropBank and Abstract Meaning Representation (AMR).

[Landing page for this record](#)



## Tēzauris.lv (Winter 2020)

(Part of Language resources and tools of AiLab IMCS UL)

Tezauris is a machine-readable lexicon and an online dictionary (<https://tezauris.lv/>) for Latvian. The initial human-oriented version of this resource was made publicly available in 2009, comprising more than 125,000 entries that were consolidated from around 40 sources: modern and historical dictionaries, mostly available...

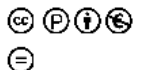
[Landing page for this record](#)



## The Balanced Corpus of Modern Latvian

(Part of Language resources and tools of AiLab IMCS UL)

LVK2018 is the 10 million representative corpus of contemporary Latvian. LVK2018 is enlarged version of LVK2013. LVK2018 is general language, representative and balanced corpus that aims to cover the variety of existing texts in some estimated proportions. The corpus contains five different domains: journalism (60%), f...



# VLO: Virtual Language Observatory



Search through 9 records



Showing 8 results within selection for CLARIN Centre of Latvian language resources and tools

Results per page: 10

Use the categories below to limit the search results to those matching the selected value(s).

Language



Collection

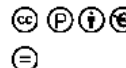


## Full Stack of Latvian Language Resources for NLU

(Part of Language resources and tools of AiLab IMCS UL)

This repository contains a multilayer text corpus of Latvian. The multilayer corpus is anchored in cross-lingual state-of-the-art representations: Universal Dependencies (UD), FrameNet, PropBank and Abstract Meaning Representation (AMR).

[Landing page for this record](#)



[Baltischer historisch-geographischer Kalender - 1907-1913](#) (4)

[CEDIFOR.Corpus.TITUS](#) (1)

[Center of Estonian Language Resources](#) (4)

[CLARIN.SI data & tools](#) (4)

[Clarino Bergen Centre - INESS](#) (1)

[European Language Resources Association](#) (9)

[Jahrbuch für Genealogie, Heraldik und Sphragistik - 1894-1930](#) (1)

[Language resources and tools of AiLab IMCS UL](#) (6)

[LINDAT / CLARIAH-CZ Data & Tools](#) (21)

)

MCS UL)

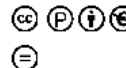
and an online dictionary (<https://tezaurs.lv/>) for Latvian. The source was made publicly available in 2009, comprising more than 10 million words extracted from around 40 sources: modern and historical dictionaries,



## Modern Latvian

MCS UL)

LVK2018 is the 10 million representative corpus of contemporary Latvian. LVK2018 is enlarged version of LVK2013. LVK2018 is general language, representative and balanced corpus that aims to cover the variety of existing texts in some estimated proportions. The corpus contains five different domains: journalism (60%), f...



# Virtual Language Observatory: EUROPEANA

VLO / Faceted search / Search results



Search through 1,859 records



Showing 11 to 20 of 1,783 results within selection for Europeana Latvian i

Results per page: 10

Use the categories below to limit the search results to those matching the selected value(s).

Language

Latvian

Collection

Resource type

Format

Availability

Search options

<< < 1 2 3 4 5 6 7 8 9 10 > >>

## Jauna isskaidrota Dzeesmu grahmata, lihdz ar swehtahm Luhgschanahm, teem Latweescheem par swehtischanu

(Part of 6\_RoL\_NLLatvia\_Collection57760)

Īpašuma ieraksts: J. Zivtiņ.

Landing page for this record at [www.europeana.eu](http://www.europeana.eu)

## Bildu-Ahbize

(Part of 6\_RoL\_NLLatvia\_Collection57760)

Autors grāmatā nav uzrādīts.; Ilustrācijas ielīmētas tekstā.; Ieraksti: Bahovskis, Jānis, īpašuma ieraksts un ieraksts par iegādi no 1860. gada Durbē.

Landing page for this record at [www.europeana.eu](http://www.europeana.eu)

## Jauna Gadda Wehleschanas pehz ikkatra gribbeschanas

(Part of 6\_RoL\_NLLatvia\_Collection57760)

Apsveikuma dzejoļi.

Landing page for this record at [www.europeana.eu](http://www.europeana.eu)

## Jauna ABC un Lassischanas Mahziba

(Part of 6\_RoL\_NLLatvia\_Collection57760)

Zīmogs: Rīgas pilsētas bibliotēka.

Landing page for this record at [www.europeana.eu](http://www.europeana.eu)

The search results include 1 record with the same title.



# CLARIN Latvia: User Dimension



- CLARIN Knowledge Centre for Systems and Frameworks for Morphologically Rich Languages SAFMORIL
- Workshops:
  - hands on corpus usage
  - treebanks
  - **Next: on VLO and CLARIN-LV repository**
- Contribution to DH course registry

# Looking forward to cooperation and new language resources

Thank you!



NATIONAL  
DEVELOPMENT  
PLAN 2020



**EUROPEAN UNION**  
European Regional  
Development Fund

INVESTING IN YOUR FUTURE