

# CLARIN-LV Infrastructure for Research and Education

Inguna Skadiņa

National coordinator of CLARIN-LV

02.06.2021.



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-LV

[Par CLARIN](#)[Repozitorijs](#)[Konferences un semināri](#)[Latviešu valodas resursi](#)[Jaunami](#)[Kontakti](#)

## Par CLARIN

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

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 servissus digitālo valodas datu un rīku izvietojšanai, savienojšanai, 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 servisi tiek nepārtraukti atjaunināti un pilnveidoti. Dažās valstīs, kas CLARIN pievienojās nesen, CLARIN infrastruktūra tiek veidota. Latvija pievienojās CLARIN ERIC 2016. gada 1. jūnijā.



NACIONĀLAIS  
ATTĪSTĪBAS  
PLĀNS 2020



EIROPAS SAVIENĪBA  
Eiropas Reģionālās  
attīstības fonds

REGULĀCIJAS TAVĀ NĀKOTNĒ

CLARIN Latvija nacionālā mezglu punkta izveidi atbalsta Eiropas Reģionālās attīstības fonda projekts "Latvijas Universitāte un institūti Eiropas pētniecības telpā - ekselence, aktivitāte, mobilitāte, kapacitāte" (Nr. 1.1.1.5/18/I/016)

Reģistrācija konferencei "Latviešu valodas digitālie resursi un rīki izglītībai"

Aicinām piedalīties konferencē "Latviešu valodas digitālie resursi un rīki izglītībai", kas notiks **2021. gada 2. jūnijā plkst. 10.30-15.00 Zoom platformā**. Konferencē iespējams reģistrēties šeit.



# CLARIN-LV repository



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

|                              |
|------------------------------|
| Grūzītis, Normunds (7)       |
| Levāne-Petrova, Kristīne (7) |
| Dargis, Roberts (6)          |
| Auziņa, Ilze (5)             |
| Saulīte, Baiba (5)           |
| ... View More                |

### Subject

|                             |
|-----------------------------|
| corpus (2)                  |
| orthographic transcr ...(2) |
| speech corpus (2)           |
| Abstract Meaning Rep ...(1) |
| AMR (1)                     |
| ... View More               |

### Language (ISO)

|                |
|----------------|
| Latvian (16)   |
| German (2)     |
| English (1)    |
| Latgalian (1)  |
| Lithuanian (1) |

*repository.clarin.lv*

## ToolService

CLARIN Centre Of Latvian Language Resources And Tools

## LVBERT - Latvian BERT

## Author(s):

Znotiņš, Artūrs

## Description:

LVBERT is the first publicly available monolingual BERT language model pre-trained for Latvian. For training we used the original implementation of BERT on TensorFlow with the whole-word masking and the next sentence ...

 This item contains 2 files (1.51 GB).

Publicly Available

## Corpus

CLARIN Centre Of Latvian Language Resources And Tools

## Latvian Learner Corpus (LaVa)

## Author(s):

Auziņa, Ilze ; Kaija, Inga ; Levāne-Petrova, Kristīne ; Pokratiece, Kristīne ; Darģis, Roberts

## Description:

The corpus includes almost 1000 texts created by foreign students studying at a Latvian higher education institution who are learning Latvian as a foreign language in the first or second semester. The morphologically ...

 This item contains no files.

## Corpus

CLARIN Centre Of Latvian Language Resources And Tools

## Rainis

## Author(s):

Spektors, Andrejs ; Grūzītis, Normunds ; Darģis, Roberts ; Auziņa, Ilze ; Saulīte, Baiba ; Levāne-Petrova, Kristīne

## Description:

This specialised text corpus contains all of Rainis work: plays, poetry, prose, journalism, translations, letters, etc.

 This item contains 2 files (56.83 MB).

# 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 ; Dargis, Roberts

This item contains no files.

Publicly Available

LexicalConceptualResource

CLARIN LATVIA

Tēzaurs.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

CLARIN

8

# 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

#### Tēzaurs.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

# Citation

## Rainis





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

BIBTEX

CMDI

Spektors, Andrejs; Grūzītis, Normunds; Dargis, Roberts; et al., 2018, *Rainis*, CLARIN-LV digital library at IMCS, University of Latvia, <http://hdl.handle.net/20.500.12574/41>.



Share:  



# Usage scenario: Corpus search

Repository

Corpus Search

About



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           | stāvēt pie | stūres . Pūta spirts dienvidrietenis , un « Speķa Lode             |
| <input type="checkbox"/> | <a href="#">doc#15350</a> | lati diennaktī , jo ir jāēd . " Es nevaru  | stāvēt pie | operāciju galda , stundām skatoties dažādu centimetru laukumā , un |
| <input type="checkbox"/> | <a href="#">doc#15802</a> | , pārcilāt it visas lietas istabā , atvērt logu un                                   | stāvēt pie | tā , līdz sirds piesalst pie ribām un pārstāj pukstēt              |
| <input type="checkbox"/> | <a href="#">doc#19048</a> | gadiem darīja Hilarija Klintonē . Jau viņai toreiz pārmeta nevēlēšanos               | stāvēt pie | plīts . Mišele Obama tagad mierīgi atzīst , ka neprot              |
| <input type="checkbox"/> | <a href="#">doc#19107</a> | smagi strādāt , lai nopelnītu iztiku , uzkopt telpas ,                               | stāvēt pie | plīts . Psihologi izpētījuši , ka tikai 30 % cilvēku               |
| <input type="checkbox"/> | <a href="#">doc#20485</a> | smaržo gatavais pamatēdiens . Vai leģendārajai multiplikatorei vienmēr atļicis laika | stāvēt pie | katliem ? " Atziņos , 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 pārspriedumi | 226k |  | <a href="#">Details</a> |

# Usage scenarios: access to another CLARIN node

LVTB | AiLab

LVTB - Latvian dependency-constituency treebank

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 sēdējies zem tējas koka un dzēris vārītu ūdeni, sacēlies vējš un imperatora krūzē iepūtis pāris tējas koka lapas.

```
graph TD
    Root["ID: a-c2-p4s2<br/>sent"] --> Adv1["adv<br/>brīdī<br/>ncmsl2<br/>[brīdis]"]
    Root --> Pred["pred<br/>crdParts"]
    Root --> Punct["punct<br/>zs<br/>[.]"]
    Adv1 --> Attr1["attr<br/>Kādā<br/>pq0msln<br/>[kāds]"]
    Adv1 --> Adv2["adv<br/>brīdī<br/>ncmsl2<br/>[brīdis]"]
    Adv1 --> Attr2["attr<br/>atpūtas<br/>ncfsg4<br/>[atpūta]"]
    Adv1 --> Attr3["attrCl<br/>subrCl"]
    Pred --> CrdPart1["crdPart<br/>xPred<br/>vm0rp000san[actpmn]"]
    Pred --> Subj["subj<br/>vējš<br/>ncmsn1<br/>[vējš]"]
    Pred --> Conj["conj<br/>un<br/>cc<br/>[un]"]
    Pred --> CrdPart2["crdPart<br/>xPred<br/>vm0rp000san[actpmn]"]
    CrdPart1 --> BasElem1["basElem<br/>sacēlies<br/>vmypdmsnasnnpn<br/>[sacēlies]"]
    CrdPart1 --> AuxVerb1["auxVerb<br/>reduction:<br/>vcnrpii00an(esot)"]
    CrdPart2 --> Adv3["adv<br/>krūzē<br/>ncfsl5<br/>[krūze]"]
    CrdPart2 --> BasElem2["basElem<br/>iepūtis<br/>vmnpdmsnasnnpn<br/>[iepūst]"]
    CrdPart2 --> AuxVerb2["auxVerb<br/>reduction:<br/>vcnrpii00an(esot)"]
    CrdPart2 --> Obj["obj<br/>lapas<br/>ncfpa4<br/>[lapa]"]
```

# Language resources from another CLARIN node

Corpus

CLARIN.SI Data & Tools

## Multilingual comparable corpora of parliamentary debates ParlaMint 2.0

(CLARIN ERIC / 2021-05-10)

**Author(s):**  
Erjavec, Tomaž ; Ogrodniczuk, Maciej ; Osenova, Petya ; Ljubešić, Nikola ; Simov, Kiril ; Grigorova, Vladislava ; Rudolf, Michał ; Pančur, Andrej ; Kopp, Matyáš ; Barkarson, Starkaður ; Steingrímsson, Steinþór ; van der Pol, Henk ; Depoorter, Griet ; de Does, Jesse ; Jongejan, Bart ; Haltrup Hansen, Dorte ; Navarretta, Costanza ; Calzada Pérez, María ; de Macedo, Luciana D. ; van Heusden, Ruben ; Marx, Maarten ; Çöltekin, Çağrı ; Coole, Matthew ; Agnoloni, Tommaso ; Frontini, Francesca ; Montemagni, Simonetta ; Quochi, Valeria ; Venturi, Giulia ; Ruisi, Manuela ; Marchetti, Carlo ; Battistoni, Roberto ; Sebők, Miklós ; Ring, Orsolya ; Dargis, Roberts ; Utkā, Andrius ; Petkevičius, Mindaugas ; Briedienė, Monika ; Krilavičius, Tomas ; Morkevičius, Vaidas



 This item contains 17 files (2.03 GB).

Publicly Available 


# ParlaMint-LV



[Repository](#)[About](#)[Contact](#)[Query](#) [Corpora](#) [Save](#) [Concordance](#) [Filter](#) [Frequency](#) [Collocations](#) [View](#) [Help](#)

Corpus: ParlaMint-LV 2.0 (Latvian parliament) | Query: valoda (4,527 hits)

Hits: 4,527 | i.p.m. <sup>0</sup>: 545.74 (related to the whole "parlamint20\_lv") | ARF <sup>0</sup>: 739.42 | Result is sorted

1 / 114 ►►►

Line selection: simple ▼ |  [Display options](#)

|                          |  |   |         |  |
|--------------------------|--|---|---------|--|
| <input type="checkbox"/> | <a href="#">TretjakaZenta,2016-06-16</a>   | solu būt uzticīgs Latvijai, stiprināt tās suverenitāti un latviešu    | valodu  | kā vienīgo valsts valodu, aizstāvēt Latviju kā neatkarīgu un   |
| <input type="checkbox"/> | <a href="#">TretjakaZenta,2016-06-16</a>   | , stiprināt tās suverenitāti un latviešu valodu kā vienīgo valsts     | valodu  | , aizstāvēt Latviju kā neatkarīgu un demokrātisku valsti, savus  |
| <input type="checkbox"/> | <a href="#">BalodisRingolds,2017-02-02</a> | latviski runājot, jeb bestselleru, kā to sauc angļu                   | valodā  | . Tātad kādā provinces ciematīnā, piejūras ciematīnā meklē šerifu  |
| <input type="checkbox"/> | <a href="#">VijumsJuris,2017-02-02</a>     | tāpēc... es atvainojos tiem, kas seko manai latviešu                  | valodas | pareizībai, - nevis "mēdiji", bet "  |
| <input type="checkbox"/> | <a href="#">KrauzeArmands,2016-12-01</a>   | , ko es esmu iesniedzis, ir paredzēts, ka                             | valoda  | ir ļoti nozīmīgs instruments, jo lauksaimniecība ir augsta riska   |
| <input type="checkbox"/> | <a href="#">KrauzeArmands,2016-12-01</a>   | instruments, jo lauksaimniecība ir augsta riska nozare, kur           | valodas | zināšanas ir nepieciešamas kaut vai tādēļ, lai strādātu ar   |
| <input type="checkbox"/> | <a href="#">KrauzeArmands,2016-12-01</a>   | apritē un pārzinātu mūsu noteikumus. Un vēl jo vairāk                 | valodas | zināšanas lauksaimniecībā ir nepieciešamas saistībā ar darba drošību. Mēs  |
| <input type="checkbox"/> | <a href="#">KrauzeArmands,2016-12-01</a>   | darba drošības noteikumi ir pārkāpti, jo īpašnieki nepārzina latviešu | valodu  | un nevar iepazīties ar noteikumiem. Tādēļ es, balstoties   |
| <input type="checkbox"/> | <a href="#">KrauzeArmands,2016-12-01</a>   | saimniecībās paši arī uz vietas darbojas, ir jāzina latviešu          | valoda  | attiecīgā līmenī. Paldies.  [Sēdes vadītāja.] Vārds deputātam Ingmāram Līdakam          |
| <input type="checkbox"/> | <a href="#">SmiltēnsEdvards,2016-12-01</a> | dažāda veida ierobežojumi, to skaitā arī šis - par                    | valodu  | . Šādu priekšlikumu mēs esam skatījuši jau pirms vairākiem gadiem  |
| <input type="checkbox"/> | <a href="#">SmiltēnsEdvards,2016-12-01</a> | , un rūpīgi jāizvērtē tādi apgrūtinājumi kā, piemēram,                | valoda  | - vai tie līdzēs. Es aicinu balsot par šo  |
| <input type="checkbox"/> | <a href="#">DombravaJānis,2016-12-01</a>   | kungs - , ka viens no ierobežojumiem varētu būt valsts                | valodas | prasmes prasības tām personām, kuras vēlas iegādāties lauksaimniecības zemi  |
| <input type="checkbox"/> | <a href="#">DombravaJānis,2016-12-01</a>   | ka valsts zeme var piederēt tiem, kuri spēj valsts                    | valodā  | pamatot zemes turpmāko izmantošanu.  [Sēdes vadītāja.] Vārds deputātam Edgaram Putram |
| <input type="checkbox"/> | <a href="#">PutraEdgars,2016-12-01</a>     | taču neminēja to, ka Francijā šāda prasība par valsts                 | valodas | ... par prezentēšanu komisijā - kā tad to zemi izmantos  |

# Linguistically annotated multilingual comparable corpora of parliamentary debates ParlaMint.ana 2.0

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

BIBTEX

CMDI

Erjavec, Tomaž; et al., 2021, *Linguistically annotated multilingual comparable corpora of parliamentary debates ParlaMint.ana 2.0*, Slovenian language resource repository CLARIN.SI, <http://hdl.handle.net/11356/1405>.



This resource is also integrated in following services:



Share:



KonText



noSketch



CLARIN.SI Data &amp; Tools

## Authors

Erjavec, Tomaž ; Ogrodniczuk, Maciej ; Osenova, Petya ; Ljubešić, Nikola ; Simov, Kiril ; Grigorova, Vladislava ; Rudolf, Michał ; Pančur, Andrej ; Kopp, Matyáš ; Barkarson, Starkaður ; Steingrímsson, Steinþór ; van der Pol, Henk ; Depoorter, Griet ; de Does, Jesse ; Jongejan, Bart ; Haltrup Hansen, Dorte ; Navarretta, Costanza ; Calzada Pérez, María ; de Macedo, Luciana D. ; van Heusden, Ruben ; Marx, Maarten ; Çöltekin, Çağrı ; Coole, Matthew ; Agnoloni, Tommaso ; Frontini, Francesca ; Montemagni, Simonetta ; Quochi, Valeria ; Venturi, Giulia ; Ruisi, Manuela ; Marchetti, Carlo ; Battistoni, Roberto ; Sebök, Miklós ; Ring, Orsolya ; Darğis, Roberts ; Utká, Andrius ; Petkevičius, Mindaugas ; Briedienė, Monika ; Krilavičius, Tomas ; Morkevičius, Vaidas ; Bartolini, Roberto ; Cimino, Andrea

# Toolchain for Latvian Language Processing: NLP-PIPE

Ādams Panks vēl aizvien mitinās koka dobumā Vērmānes dārzā.

Go

tokenizer ×

morpho ×

parser ×

ner ×

☐ NER ☒ CONLL ☐ JSON

| INDEX  | FORM     | LEMMA      | UPOSTAG | XPOSTAG     | FEATS         | HEAD | DEPREL    |
|--|----------|------------|---------|-------------|---------------|------|-----------|
| #text=Ādams Panks vēl aizvien mitinās koka dobumā Vērmānes dārzā . |          |            |         |             |               |      |           |
| 1  | Ādams    | Ādams      | PROPN   | npmsn1      | Case=NomlGer  | 5    | nsubj     |
| 2  | Panks    | Panks      | NOUN    | ncmsn1      | Case=NomlGer  | 1    | flat:name |
| 3  | vēl      | vēl        | ADV     | r0_         | _             | 5    | advmod    |
| 4  | aizvien  | aizvien    | ADV     | r0_         | _             | 5    | advmod    |
| 5  | mitinās  | mitināties | VERB    | vmyip_330an | Evident=FhlMo | 0    | root      |
| 6  | koka     | koks       | NOUN    | ncmsg1      | Case=GenlGer  | 7    | nmod      |
| 7  | dobumā   | dobums     | NOUN    | ncmsl1      | Case=LoclGer  | 5    | obl       |
| 8  | Vērmānes | Vērmāne    | NOUN    | npfsg5      | Case=GenlGer  | 9    | nmod      |
| 9  | dārzā    | dārzs      | NOUN    | ncmsl1      | Case=LoclGer  | 7    | nmod      |
| 10   | .        | .          | PUNCT   | zs          | _             | 5    | punct     |

Ādams Panks **person** vēl aizvien mitinās koka dobumā Vērmānes dārzā **location** .



# VLO: Virtual Language Observatory

Language

Latvian

Latvian ✕

Latvian Sign Language (1)

more...

Collection

Resource type

Modality

Format

Keyword

▼

Availability

## The Latvian-Estonian Dictionary

(Part of Center of Estonian Language Resources)

⊞ A two-way Latvian-Estonian dictionary of nearly 47,000 entries.; Kahesuunaline läti-eesti sõnaraamat, mis sisaldab ligi 47 000 märksõnaartiklit.

Latvian Estonian

🏠 Landing page for this record

0

🔍 Ⓐ Ⓒ

## The Estonian-Latvian Dictionary

(Part of Center of Estonian Language Resources)

⊞ A two-way Estonian-Latvian dictionary of nearly 47,000 entries.; Kahesuunaline eesti-läti sõnaraamat, mis sisaldab ligi 47 000 märksõnaartiklit.

Estonian Latvian

🏠 Landing page for this record

0

🔍 Ⓐ Ⓒ

## JRC EU DGT Translation Memory Parsebank DGT-UD 1.0

(Part of CLARIN.SI data & tools)

⊞ DGT-UD is a 2 billion word 23-language parallel syntactically parsed corpus, which consists of the JRC DGT translation memory of European law, automatically annotated with UD-Pipe 1.2 (<http://ufal.mff.cuni.cz/udpipe>) using Universal Dependencies 2.0 models (<http://hdl.handle.net/11234/1-2364>). Note that the European ...

Bulgarian Czech Danish German Modern Greek ... (+19)

🏠 Landing page for this record


1 24


Ⓒ Ⓐ Ⓒ

# CLARIN-LV in 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




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](#)

 1





## Tēzaurs.lv (Winter 2020)

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

⊞ 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 that were consolidated from around 40 sources: modern and historical dictionaries, mostly availa...

 [Landing page for this record](#)

 1  1








## 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:

 1



# General Overview

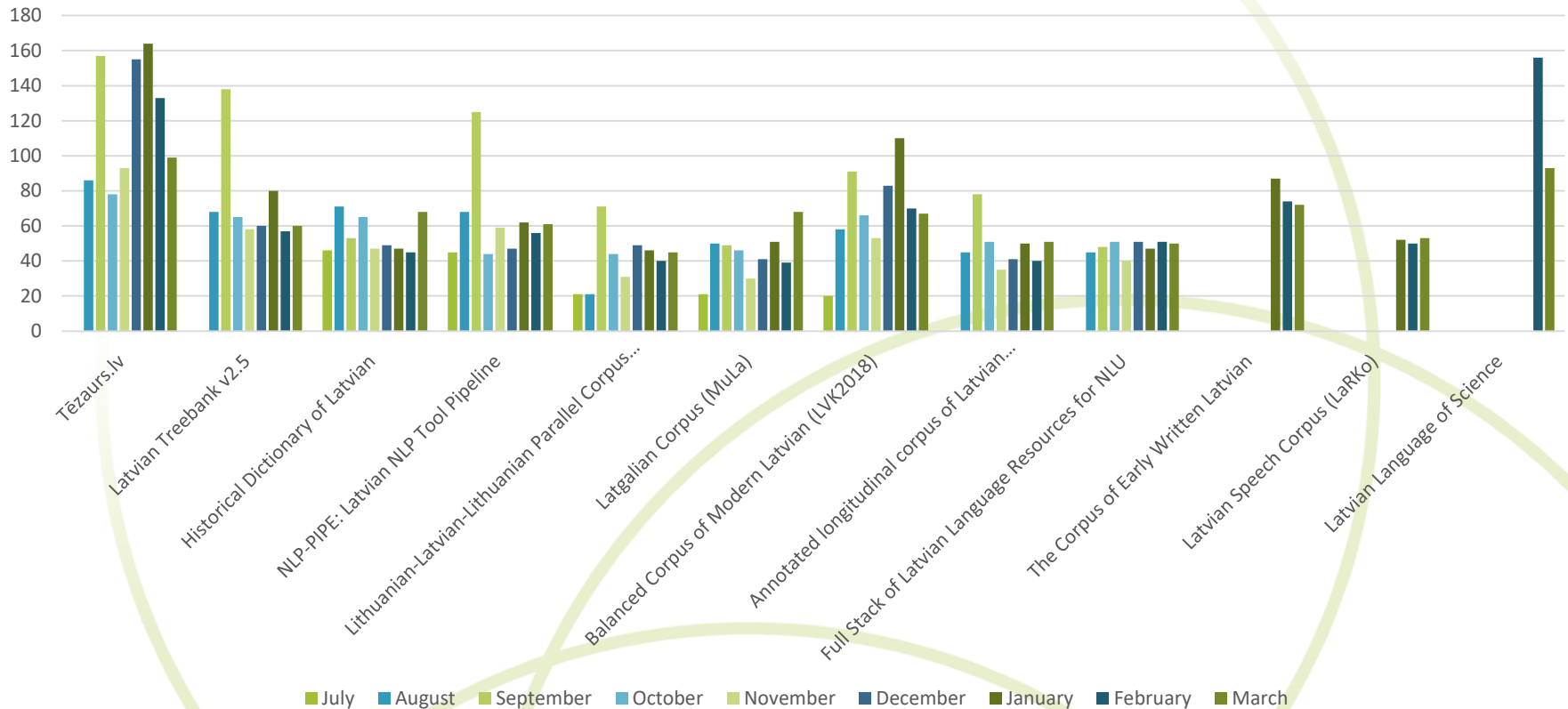
|                  |       |
|------------------|-------|
| Item Views       | 6,585 |
| Collection Views | 566   |
| Community Views  | 20    |
| User Logins      | 268   |

27 registered users, including users from Lithuania, Greece and Sweden

# Items viewed (July,2020-March, 2021)

| Item/Handle  | Number of views    |
|--|--------------------|
| <a href="#">Tēzaur.lv 2020 (Spektors, Andrejs et al) (20.500.12574/9)</a>  | 1052 [+bots: 1115] |
| <a href="#">Balanced Corpus of Modern Latvian (LVK2018) (Levāne-Petrova, Kristīne et al) (20.500.12574/11)</a>     | 665 [+bots: 696]   |
| <a href="#">Latvian Treebank v2.5 (Rituma, Laura et al) (20.500.12574/10)</a>                                      | 655 [+bots: 704]   |
| <a href="#">NLP-PIPE: Latvian NLP Tool Pipeline (Znotiņš, Artūrs) (20.500.12574/4)</a>                             | 629 [+bots: 685]   |
| <a href="#">Historical Dictionary of Latvian (Andronova, Everita et al) (20.500.12574/3)</a>                       | 544 [+bots: 579]   |
| <a href="#">Annotated longitudinal corpus of Latvian children's language (Auziņa, Ilze et al) (20.500.12574/7)</a> | 450 [+bots: 477]   |
| <a href="#">Full Stack of Latvian Language Resources for NLU (Grūzītis, Normunds et al) (20.500.12574/5)</a>       | 442 [+bots: 482]   |
| <a href="#">Latgalian Corpus (MuLa) (Sperga, Ilze et al) (20.500.12574/8)</a>                                      | 434 [+bots: 464]   |
| <a href="#">Lithuanian-Latvian-Lithuanian Parallel Corpus (LILA) (Utka, Andrius et al) (20.500.12574/6)</a>        | 431 [+bots: 461]   |
| <a href="#">The Corpus of Early Written Latvian (Andronova, Everita et al) (20.500.12574/12)</a>                   | 317 [+bots: 328]   |
| <a href="#">Latvian Language of Science (Dubova, Agnese et al) (20.500.12574/30)</a>                               | 314 [+bots: 322]   |
| <a href="#">Latvian Speech Corpus (LaRko) (Auziņa, Ilze et al) (20.500.12574/22)</a>                               | 187 [+bots: 191]   |
| <a href="#">Mühlenbach Endzelin Latvian Dictionary (MEV) (Andronova, Everita et al) (20.500.12574/38)</a>          | 53 [+bots: 54]     |

# Items viewed (July, 2020-March, 2021)



# Depositing Services

One of the fundamental services of the CLARIN infrastructure is making sure that language resources **can be archived** and **made available to the community in a reliable manner**. To help researchers to store their resources in a sustainable way, CLARIN-LV offers a **depositing service**.



# Workshop “How to Register Language Resources in the CLARIN Repository”



## 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

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



#### 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



Deposit and  
access to  
licensed  
resources

# Easy access to protected resources



eng | nob | Sign in: [CLARIN SPF](#) | [eduGAIN](#) | [OpenIdP](#)

## Corpuscle :: Home Page

Corpuscle Home  
Getting started  
Documentation  
FAQ  
Publications  
Links

Corpus list

Query  
Concordance  
Collocations  
Distribution  
Word List  
Text  
Metadata  
Overview  
Variables  
Corpus doc.

**Corpuscle** is a corpus management and analysis system for annotated corpora. The tool is being developed at [Clarino Centre Bergen](#), in collaboration with colleagues at the University of Bergen. The work has been supported by grants from [The Research Council of Norway](#), the [Meltzer foundation](#) and the [CLARINO](#) infrastructure.

This site hosts several types of corpus collections:

- Open-access corpora
- Restricted-access corpora
- Norwegian corpora
- Corpus in a variety of other languages
- General language corpora
- Specialised corpora (parallel corpora, dialect corpora, language learners' corpora, historical corpora and more)

All corpora are an integral part of the Norwegian [CLARINO](#) infrastructure.

To try out the system click on "Corpus list" to the left and choose a corpus to work with.

### Access

Access to many corpora is restricted to registered users after login.

See [Getting started](#) for more information.



**CLARINO**

**Corpuscle** :: ICAME – BROWN Family :: Concordance

## Localization


sister. Yet over the coming, tearstained months, she adopts her mother's so whether your idea of perfection is a blonde or brunette, male or female, anal legs, it is impossible for a human to keep up with, say, a three-legged they went out walking. George was sorry to hear that the pony is ill and the ft. The remaining 42 per cent were ambidextrous. Many owners think their en they were briefly joined in turn by their owner, an unfamiliar human, a re efficient layer of insulation. In hot weather, licking compensates for the y or unconsciously, their internal clocks had told them to expect food. The some were capable of having sex ten times in an hour. An average mother er litter, though a litter may contain as many as 13 kittens. In one case, a kittens. In one case, a cat was discovered to be carrying 18 foetuses. One '00 - almost four times as many as cats, which have approximately 470. A , takes to wearing her hand-knitted jumpers and learns to bake. At the sa or dog, by understanding how to clone you could be one step ahead in the or dog. Compare this with the risible image of an ill-fated ancestor attempt dirty'; Robert was anxious that his mother should learn baby to say Georg senses their arrival home in the car: The truth is more likely to be that its or an unfamiliar dominant dog, a Belgian Shepherd. The dogs' tails wagge 's lack of sweat glands and helps to cool down the fur. Historically, cats ha 's purr may be a self-healing mechanism. Cats purr at between 25 and 50 has between one and eight kittens per litter, though a litter may contain a was discovered to be carrying 18 foetuses. One cat was recorded as havin was recorded as having 420 kittens over a period of 17 years. The oldest l 's tastes are moulded when it is a kitten. During weaning its mother gives

# Sustainability

## Filter

- Subjects ⊕
- Content Types ⊕
- Countries ⊕
- Data access ⊕
- Data access restrictions ⊕
- Database access ⊕
- Database licenses ⊕
- Data licenses ⊕
- Data upload ⊕
- Data upload restrictions ⊕
- Enhanced publication ⊕
- Institution responsibility type ⊕
- Institution type ⊕
- Keywords ⊕
- Metadata standards ⊕
- PID systems ⊕
- Provider types ⊕
- Quality management ⊕
- Repository languages ⊕
- Repository types ⊕
- Versioning ⊕

clarin-lv

 Search

[Toggle short help](#)

← Previous 1 Next →

Sort by ▾

Found 1 result(s)

## CLARIN-LV repository

[CLARIN Latvia repository](#)



Subject(s)

Engineering Sciences Computer Science, Electrical and System Engineering

Artificial Intelligence, Image and Language Processing Computer Science Linguistics Humanities

Humanities and Social Sciences

Content type(s)

Standard office documents Plain text Software applications Audiovisual data Structured text Source code

Archived data

Country

Latvia European Union

CLARIN-LV is a national node of Clarin ERIC (Common Language Resources and Technology Infrastructure). The mission of the repository is to ensure the availability and long term preservation of language resources. The data stored in the repository are being actively used and cited in scientific publications.



# Sustainability



[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!

# 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.





- 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 tool "
- Contribution to DH course registry



[T]:

vārds vai frāze



Apkaime

Būšnieki

Bušnieku

bušprits

Buštehrada

Bušteni

Bušugrava

bušuruncis

**būt**

butadiēnkaučuki

butadiēnnitrilkaučuki

butadiēns

butadiēnstirolkaučuki

butadions

butaforija

butaforisks

**būt****būt** nekārtns darbības vārds; intransitīvs ➞

Īstenības izteiksme:

|          | Tagadne     |             | Pagātne     |              | Nākotne     |                      |
|----------|-------------|-------------|-------------|--------------|-------------|----------------------|
|          | Vsk.        | Dsk.        | Vsk.        | Dsk.         | Vsk.        | Dsk.                 |
| 1. pers. | <i>esmu</i> | <i>esam</i> | <i>biju</i> | <i>bijām</i> | <i>būšu</i> | <i>būsīm</i>         |
| 2. pers. | <i>esi</i>  | <i>esat</i> | <i>biji</i> | <i>bijāt</i> | <i>būsi</i> | <i>būsiet, būsīt</i> |
| 3. pers. |             | <i>ir</i>   |             | <i>bija</i>  |             | <i>būs</i>           |

Pavēles izteiksme: *esi* (vsk. 2. pers.), *esiet* (dsk. 2. pers.)Atstāstījuma izteiksme: *esot* (tag.), *būšot* (nāk.)Vēlējuma izteiksme: *būtu*Vajadzības izteiksme: *jābūt*



# Latvian Language Resources in CLARIN Language Resource Families

## 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

# Looking forward to cooperation and new language resources

## Thank you!



NATIONAL  
DEVELOPMENT  
PLAN 2020



**EUROPEAN UNION**  
European Regional  
Development Fund

INVESTING IN YOUR FUTURE