

CLARIN-LV

solis pretī valodu digitālai vienlīdzībai

Inguna Skadiņa

CLARIN-LV nacionālā koordinatore

12.09.2023.



NATIONAL
DEVELOPMENT
PLAN 2020



EUROPEAN UNION
European Regional
Development Fund

INVESTING IN YOUR FUTURE

Vienota valodas resursu un tehnoloģiju infrastruktūra

CLARIN (Common Language Resources and Technology Infrastructure) ir digitāla **Eiropas pētniecības infrastruktūra** (ERIC – European Research Infrastructure Consortium)

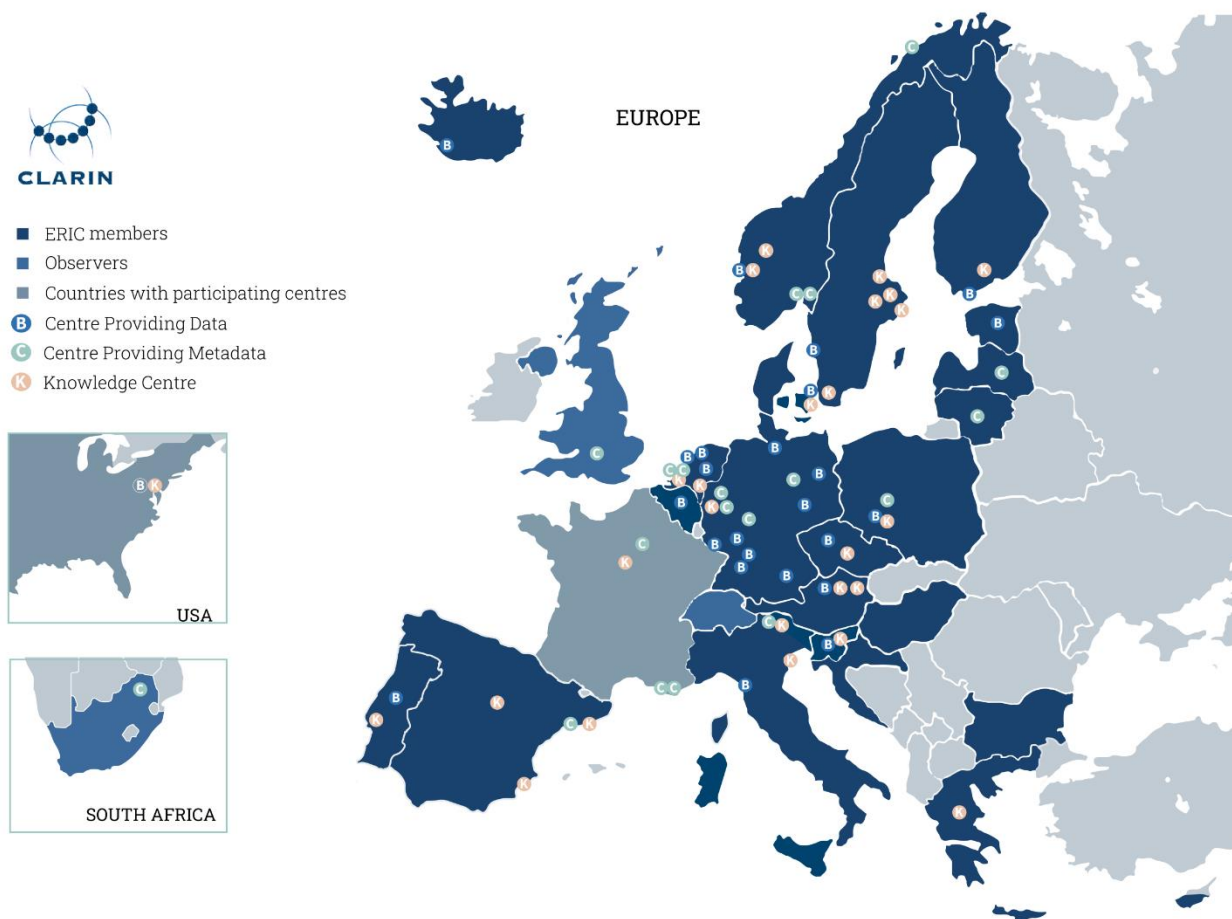
Tās pamatā ir ideja, ka

visi **digitālie valodas resursi un rīki** Eiropā un citur pasaulē ir pieejami

visu jomu, it īpaši **humanitāro un sociālo zinātņu** zinātniekiem un studentiem, izmantojot **vienotu pierakstīšanos**

CLARIN ERIC – dalīta Eiropas pētniecības infrastruktūra

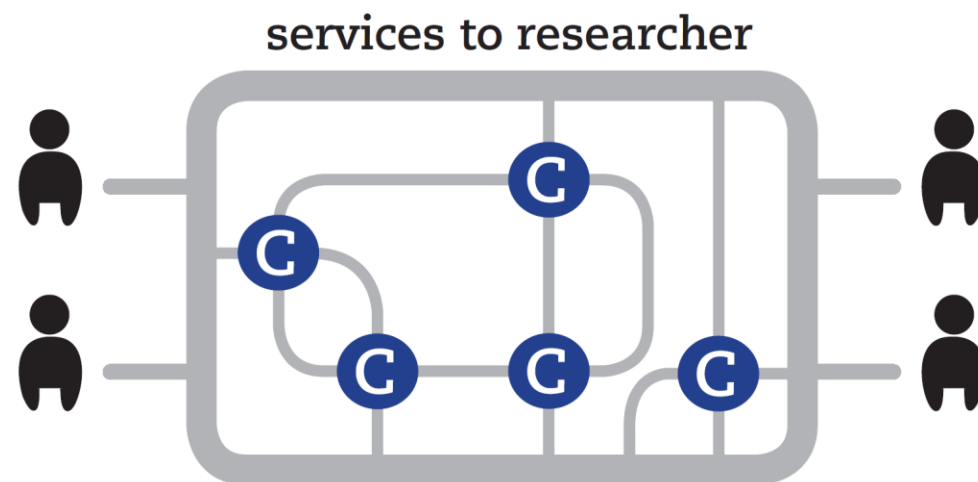
- 23 + 3 valstis
- Federētu servisu piedāvājums:
 - Vienkārša un **ilgtspējīga** piekļuve **digitālajiem valodas datiem** (rakstiskā, runātā vai multimodālā formā)
 - Mūsdienīgi **valodas apstrādes rīki**, valodas datu izpētei un analīzei neatkarīgi no to atrašanās vietas
- > 60 reģistrētu centru
- Zināšanu centru tīkls (K-centri)
- Cieša sasaiste ar Atvērtās zinātnes programmu, fokuss uz **sadarbspēju**



<https://www.clarin.eu/content/participating-consortia>

FAIR principles for open language resources

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



Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. [The FAIR Guiding Principles for scientific data management and stewardship](#). *Sci Data* 3, 160018, 2016

Par CLARIN

CLARIN (Common Language Resources and Technology Infrastructure –Vienota valodas resursu un tehnoloģiju infrastruktūra) ir dalīta Eiropas pētniecības infrastruktūra (ERIC), kurā integrētas zināšanas, valodas resursi un rīki, repozitoriji un servisi, 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. Lai to īstenotu, izveidota kopīga izklaidēta infrastruktūra, kurā integrēti valodas resursi un rīki no visas Eiropas. CLARIN piedāvā ilgtermiņa risinājumus un servisu digitālo valodas datu un rīku izvietojšanai, savienojšanai un analīzei.

Latvija CLARIN iniciatīvā piedalās kopš tās pirmsākumiem. 2008. gadā LU MII kļūst par CLARIN sagatavošanas posma projekta konsorcijs dalībnieku, būtisku finansiālu atbalstu dalībai projektā sniedz Izglītības un zinātnes ministrija. 2016. gada 1. jūnijā Latvija pievienojas CLARIN ERIC. Izglītības un zinātnes ministrija uzticējusi LU Matemātikas un informātikas institūtam pārstāvēt Latviju CLARIN ERIC un vadīt Latvijas CLARIN konsorcijs. 2019. gadā CLARIN-LV pievienojas CLARIN zināšanu centram morfoloģiski bagātām valodām SAFMORIL. 2020. gada martā tiek izveidots CLARIN-LV repozitorijs, kurā tiek reģistrēti latviešu valodas resursi un rīki. Esot CLARIN ERIC konsorcijs, CLARIN-LV ir apņēmusies ilgstoši rūpēties un saglabāt repozitorijā deponētos digitālos valodas resursus un rīkus. CLARIN-LV īsteno CLARIN ERIC misiju Latvijā, vācot, dokumentējot, kurējot un nodrošinot ilgtermiņa piekļuvi digitālajiem latviešu valodas resursiem un rīkiem. 2022. gada janvārī, konsorcijs apvienojoties septiņiem partneriem, tiek noslēgts CLARIN nacionālā konsorcijs līgums, kurā CLARIN-LV konsorcijs vienojas par CLARIN-LV darbību un turpmāku attīstību. 2023.gada janvārī CLARIN-LV repozitorijs iegūst B-centra statusu un saņem CoreTrustSeal sertifikātu kā uzticams datu repozitorijs.

CLARIN-LV regulāri rīko konferences un seminārus, kuros iepazīstina ar dažādiem valodas resursiem un rīkiem un to lietojumu pētniecībā un valodu tehnoloģiju izveidē. Latvijā paveiktais apkopots vairākās publikācijās.

Kopš 2018. gada CLARIN Latvija darbību atbalsta ERAF projekts "Latvijas Universitāte un institūti Eiropas pētniecības telpā - ekselence, aktivitāte, mobilitāte, kapacitāte", kā arī projekti "Humanitāro zinātņu digitālie resursi: integrācija un attīstība" (5.10.2020-5.10.2022), "Atvērtas un FAIR principi atbilstīgas digitālo humanitāro zinātņu ekosistēmas attīstība Latvijā" (kopš 2022. gada decembra), "Mūsdienu latviešu valodas izpēte un valodas tehnoloģiju attīstība" (kopš 2022. gada) un "Valodu tehnoloģiju iniciatīva" (kopš 2023. gada).



Reģistrācija konferencei "Jaunākie valodas resursi un rīki digitālajām humanitārajām zinātnēm"

Aicinām piedalīties konferencē "Jaunākie valodas resursi un rīki digitālajām humanitārajām zinātnēm", kas notiks **2023. gada 12. septembrī plkst. 13.00-17.00**

Latvijas Universitātes Matemātikas un informātikas institūtā, Raiņa bulvārī 29.

Reģistrācija dalībai konferencē: <https://ej.uz/CLARINLV-registracija>.



CLARIN-LV valodas resursu repozitorijs



Atrast

Lingvistiskie dati un valodas apstrādes rīki
Citēšanas atbalsts (ar pastāvīgajiem ID)



[Detalizētā meklēšana](#)

Autors

Grūzītis, Normunds (21)

Darģis, Roberts (20)

Saulīte, Baiba (15)

Pretkalniņa, Lauma (13)

Rituma, Laura (12)

... Skatīt vairāk

Temats

text (17)

morphology (14)

dictionary (12)

manual annotation (8)

constituency (7)

... Skatīt vairāk

Valoda (ISO)

Latvian (58)

English (5)

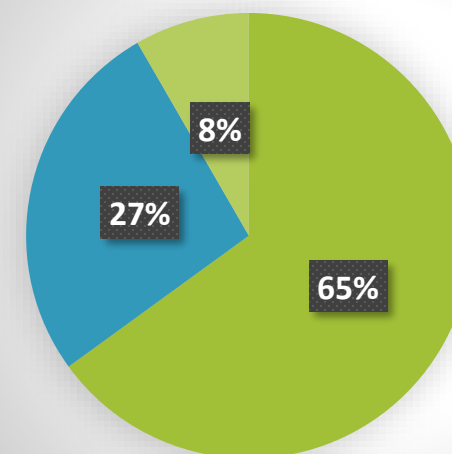
German (4)

Lithuanian (4)

Latgalian (3)

... Skatīt vairāk

Valodas resursi un rīki



- Korpusi
- Leksikoni
- Rīki

repository.clarin.lv

CLARIN-LV valodas resursu repozitorijs

repository.clarin.lv

Kas jauns

ToolService

CLARIN Centre Of Latvian Language Resources And Tools

Ilvars - Latvian Male VITS Text-to-Speech Model (vers. 2023)



Autors (-i):

Darģis, Roberts

Apraksts:

A neural model for text-to-speech (TTS) synthesis in Latvian. Trained using VITS on a 25-hour speech corpus of audiobooks read in a male voice. Currently released for research purposes only.

Šajā vienumā ir 1 fails (376.17 MB).

Publicly Available

LexicalConceptualResource

CLARIN Centre Of Latvian Language Resources And Tools

Dictionary of Contemporary Latvian Language (MLVV) (2023-07-07)



Autors (-i):

Jērāne, Santa ; et al.

► show everyone

Apraksts:

"Contemporary dictionary of Latvian language" (MLVV), which is developed by the UL Latvian Language institute, is a new explanatory dictionary based on Latvian language materials obtained during the last decade. The analysis ...

Šajā vienumā ir 1 fails (4.39 MB).

Publicly Available

LexicalConceptualResource

CLARIN Centre Of Latvian Language Resources And Tools

Tēzaurs.lv 2023 (Summer Edition)



Autors (-i):

Spektors, Andrejs ; et al.

► show everyone

Apraksts:

Tezaurs.lv is the largest open machine-readable dictionary for Latvian. This version contains more than 393,500 entries based on 345 sources. The dictionary is enriched with phonetic, morphological, semantic and other ...

Šajā vienumā ir 2 faili (26.16 MB).

Publicly Available

Kas jauns

Corpus

CLARIN Centre Of Latvian Language Resources And Tools

Latvian Treebank v2.8

Autors (-i):

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

Apraksts:

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 data used for deriving LVTB ...

📎 Šajā vienumā ir 1 fails (19.28 MB).



Publicly Available

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

Name	LVTB-v2.8-CLARIN.zip
Size	19.28 MB
Format	application/zip
Description	Latvian treebank manually annotated according hybrid dependency-constituency grammar model.
MD5	66cec920a8572ad1a326792ccdda1a2f

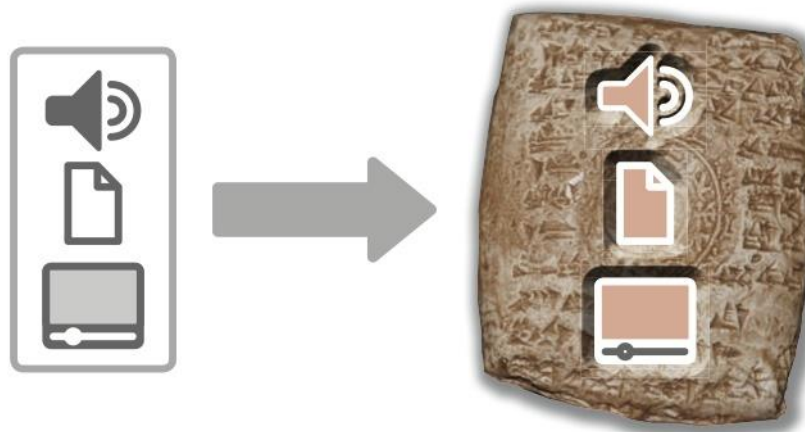


📄 Download file

👁 Preview

Datu uzglabāšanas servisi: deponēšana

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



https://www.clarin.lv/attachments/CLARIN%202020_resursu_iesniegsana.pdf

Citēšana

Linguistically annotated multilingual comparable corpora of parliamentary debates ParlaMint.ana 3.0



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

BIBTEX

CMDI

Erjavec, Tomaž; et al., 2023, *Linguistically annotated multilingual comparable corpora of parliamentary debates ParlaMint.ana 3.0*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1488>.



This resource is also integrated in following services:



Share:  

noSketch



KonText



CLARIN.SI Data & Tools

Citēšana

Linguistically annotated multilingual com debates ParlaMint.ana 3.0



Please use the following text to cite this item or export to

Erjavec, Tomaž; et al., 2023, *Linguistically annotated multilingual c ParlaMint.ana 3.0*, Slovenian language resource repository CLARIN



This resource is also integrated in following services:

noSketch



KonText



Multilingual comparable corpora of parliamentary debates ParlaMint 2.1

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

Publicēšanas datums 2021/6/18

Izdevējs

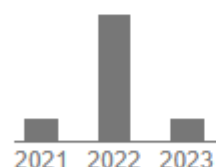
CLARIN ERIC

Apraksts

Description ParlaMint 2.1 is a multilingual set of 17 comparable corpora containing parliamentary debates mostly starting in 2015 and extending to mid-2020, with each corpus being about 20 million words in size. The sessions in the corpora are marked as belonging to the COVID-19 period (after November 1st 2019), or being "reference" (before that date). The corpora have extensive metadata, including aspects of the parliament; the speakers (name, gender, MP status, party affiliation, party coalition/opposition); are structured into time-stamped terms, sessions and meetings; with speeches being marked by the speaker and their role (eg chair, regular speaker). The speeches also contain marked-up transcriber comments, such as gaps in the transcription, interruptions, applause, etc. Note that some corpora have further information, eg the year of birth of the speakers, links to their Wikipedia articles, their membership ...

Atsauču kopskaits

Minēts 11



Ilgtspēja

Filter

[Reset all](#)

Subjects [▾](#)
Content Types [▾](#)
Countries [▾](#)
[European Union](#) (1)
[Latvia](#) (3)
AID systems [▾](#)
Certificates [▾](#)
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 [▾](#)
Software [▾](#)
Repository types [▾](#)
Versioning [▾](#)

Search...

[Search](#)

[Toggle short help](#)

[← Previous](#) **1** [Next →](#)

[Sort by ▾](#)

Found 3 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.

RTU research data repository

ORTUS



Subject(s)

[Humanities and Social Sciences](#) [Life Sciences](#) [Natural Sciences](#) [Engineering Sciences](#)

Content type(s)

[Images](#) [Raw data](#) [Databases](#) [Structured text](#) [Scientific and statistical data formats](#) [Audiovisual data](#) [Archived data](#)

Country

[Latvia](#)

Institutional research data repository of Riga Technical University (RTU). The aim is to collect and store scientific research data and other relevant information in all fields of knowledge of RTU, enabling free, easy and convenient access to it.

Riga Stradins University dataverse



Subject(s)

[Life Sciences](#) [Social Sciences](#) [Public Health, Health Services Research, Social Medicine](#) [Social and Behavioural Sciences](#) [Humanities and Social Sciences](#) [Medicine](#) [Medicine](#)

Content type(s)

[Raw data](#)

Country

[Latvia](#)

RSU data repository

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!

Populārākie valodas resursi (kopš repozitorija izveides)

Item/Handle	Number of views
Tēzauris.lv 2020 (Spektors, Andrejs et al) (20.500.12574/9)	4838 [+bots: 5510]
Latvian Treebank v2.5 (Rituma, Laura et al) (20.500.12574/10)	3087 [+bots: 3549]
Balanced Corpus of Modern Latvian (LVK2018) (Levāne-Petrova, Kristīne et al) (20.500.12574/11)	2699 [+bots: 2801]
LVBERT - Latvian BERT (Znotiņš, Artūrs) (20.500.12574/43)	2560 [+bots: 2949]
NLP-PIPE: Latvian NLP Tool Pipeline (Znotiņš, Artūrs et al) (20.500.12574/4)	2406 [+bots: 2554]
Tēzauris.lv 2022 (Spektors, Andrejs et al) (20.500.12574/66)	2270 [+bots: 2401]
The Corpus of Early Written Latvian (Andronova, Everita et al) (20.500.12574/12)	2072 [+bots: 2307]
Latvian AMR Sembank (Znotiņš, Artūrs et al) (20.500.12574/40)	2036 [+bots: 2243]
Corpus of Latvian Pandemic Diaries 2020–2021 (Reinsone, Sanita et al) (20.500.12574/48)	2012 [+bots: 2276]
Rainis (Spektors, Andrejs et al) (20.500.12574/41)	1992 [+bots: 2260]
Latvian Treebank v2.8 (Rituma, Laura et al) (20.500.12574/55)	1985 [+bots: 2162]
Latgalian Corpus (MuLa) (Sperga, Ilze et al) (20.500.12574/8)	1845 [+bots: 1924]
Latvian Treebank v2.6 (Rituma, Laura et al) (20.500.12574/54)	1835 [+bots: 2038]
Historical Dictionary of Latvian (Andronova, Everita et al) (20.500.12574/3)	1828 [+bots: 1906]
Latvian Language of Science (Dubova, Agnese et al) (20.500.12574/30)	1788 [+bots: 1842]

Populārākie valodas resursi pēdējā nedēļā

Visvairāk skatītie vienumi

Populārākie pēdējā nedēļā

	Skatījumi
Tēzaurs.lv 2022	60
Balanced Corpus of Modern Latvian (LVK2022)	21
Tēzaurs.lv 2023 (Summer Edition)	20
Tēzaurs.lv 2020	13
Ilvars - Latvian Male VITS Text-to-Speech Model (vers. 2022)	13
LVTB - Latvian Treebank v2.12 (2023-06-05)	10
LVBERT - Latvian BERT	10
Balanced Corpus of Modern Latvian (LVK2013)	10
Dictionary of Contemporary Latvian Language (MLVV)	10
LVTB - Latvian Treebank v2.11	9

CLARIN-LV valodas resursi VLO

 Virtual Language Observatory

Search

Contributors

Help

CLARIN



VLO / [Faceted search](#) / [Search results](#)

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

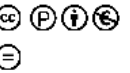
Full Stack of Latvian Language Resources for NLU

(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](#)





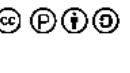
Tēzaurus.lv (Winter 2020)

(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](#)



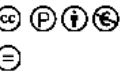


The Balanced Corpus of Modern Latvian

(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:





<https://vlo.clarin.eu>

17

CLARIN vienotā pierakstīšanās

CLARIN ERIC pamatprincips ir **vienota pierakstīšanās**, t.i., lai piekļūtu tiem CLARIN ERIC infrastruktūrā iekļautajiem valodas resursiem, kas nav brīvpieejas, lietotājam nepieciešams lietotāja konts, kas atbilst CLARIN ERIC izvirzītajām lietotāju identifikācijas un autorizācijas sistēmas prasībām un ir iekļauts CLARIN ERIC identitāšu federācijā (Latvijā tas ir LAIFE)



[eng](#) | [nob](#) | Sign in: [CLA](#)

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 in collaboration with colleagues at the University of Bergen. The work has been supported by [gr 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)

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.

Corpuscle :: ICAME – BROWN Family :: Metadata

Brown family (extended)

Full metadata record:
[hdl:11495/D92A-6FA8-C53E-1](https://hdl.handle.net/11495/D92A-6FA8-C53E-1)

Persistent identifier for the resource:
[hdl:11495/DF2E-8AB4-BEF5-1](https://hdl.handle.net/11495/DF2E-8AB4-BEF5-1)

Links:
<http://clarino.uib.no/korpuskel/landing-page?identifier=brown&view=short> (landing page @ Corpuscle)

Contact Person: [Hofland, Knut](#)

This resource is licensed under the following terms:

[CLARIN ACA](#) 

Please click on the link to read the license terms.

By accepting the terms of the license you will be granted access to the resource.

Vienotā pierakstīšanās ar licenci aizsargātiem valodas resursiem



eng | nob | [Sign out \(Inguna Skadina\)](#)

Corpuscle :: ICAME – BROWN Family :: Concordance

Corpuscle Home
Getting started
Documentation
FAQ
Publications
Links

Corpus list

Query

[Concordance](#)

Collocations

Distribution

Word List

Text

Metadata

Overview

Variables

Corpus doc.

Localization

Basic search | [switch to Advanced search](#)

cat

[Run Query](#) | [Refine](#) window: [sentence](#) | [Stop](#) | [Saved queries ...](#)

Done. Running time: 0.02 sec. (0.01 CPU sec.)

Type: [kwic](#) | ☐ Show line filter | [Attributes ...](#) | [Structures ...](#) (☐ show in match) | Page size: | Context size: [500px](#)

Hit 1 – 30 of 148 | [Previous](#) | [Next](#) | Go to: | [Download](#) (☐ Excel mode) | [Copy query PID](#)

match

sister. Yet over the coming, tearstained months, she adopts her mother's idea of whether your idea of perfection is a blonde or brunette, male or female, and how long legs, it is impossible for a human to keep up with, say, a three-legged dog. They went out walking. George was sorry to hear that the pony is ill and the cat. The remaining 42 per cent were ambidextrous. Many owners think their cats when they were briefly joined in turn by their owner, an unfamiliar human, are an efficient layer of insulation. In hot weather, licking compensates for the cat's lack of sweat glands and helps to cool down the fur. Historically, cats have purr may be a self-healing mechanism. Cats purr at between 25 and 50 Hz. Some cats have between one and eight kittens per litter, though a litter may contain as many as 13 kittens. In one case, a cat was discovered to be carrying 18 fetuses. One cat was recorded as having 420 kittens over a period of 17 years. The oldest cat is 30 - almost four times as many as cats, which have approximately 470. A

cat , takes to wearing her hand-knitted jumpers and learns to bake. At the same time, a dog, by understanding how to clone you could be one step ahead in the game. Compare this with the risible image of an ill-fated ancestor attempting to be 'dirty'; Robert was anxious that his mother should learn baby to say 'George' senses their arrival home in the car: The truth is more likely to be that its owner 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 Hz. Some cats have between one and eight kittens per litter, though a litter may contain as many as 13 kittens. In one case, a cat was discovered to be carrying 18 fetuses. One cat was recorded as having 420 kittens over a period of 17 years. The oldest cat is 30 - almost four times as many as cats, which have approximately 470. A

CLARIN zināšanu centrs morfoloģiski bagātām valodām

SAFMORIL

- SAFMORIL ir CLARIN zināšanu centrs (knowledge centre) morfoloģiski bagātām valodām
- SAFMORIL konsultē digitālo humanitāro zinātņu pētniekus un valodu tehnoloģiju izstrādātāju par valodas resursiem un rīkiem morfoloģiski bagātām valodām
- SAFMORIL apvienojušās Helsinku Universitāte, Trumses Universitāte, Latvijas Universitātes Matemātikas un Informātikas institūts un Vītauta Dižā Universitāte

CLARIN
K CENTRE



Seminārs "Runas korpusu izveide un to izmantojums"

2023. gada 18. maijā notika *digitalhumanities.lv* seminārs sadarbībā ar Clarin-LV un DARIAH-EU "Runas korpusu izveide un to izmantojums". Semināra materiāli:

- Ilze Auziņa. Runas korpusu nozīme valodu tehnoloģiju izstrādē.
- Antra Kļavinska. Mūsdienu latgaliešu runas korpusa izveide mazāk lietoto valodu dokumentēšanas kontekstā.
- Ilze Auziņa un Guna Rābante-Buša. Runas datu apstrāde programmā ELAN.



Praktiskais seminārs par jaunākajiem Mākslīgā intelekta laboratorijā radītajiem valodas resursiem

2022. gada 8. decembrī Latvijas Universitātes Matemātikas un informātikas institūts rīkoja CLARIN-LV praktisko semināru par jaunākajiem Mākslīgā intelekta laboratorijā radītajiem valodas resursiem. Semināra materiāli:

- Baiba Saulīte, Ilze Auziņa un Roberts Dargis. Nacionālā korpusu kolekcija.
- Pēteris Paikens un Laura Rituma. Latviešu valodas leksiskais tīkls (Latvian Wordnet)
- Kristīne Levāne-Petrova. Līdzsvarotais mūsdienu latviešu valodas tekstu korpus 2022

- 4. konference
- 6 semināri

Konsorcijs



Latvijas Universitātes
Matemātikas un informātikas institūts

LMI

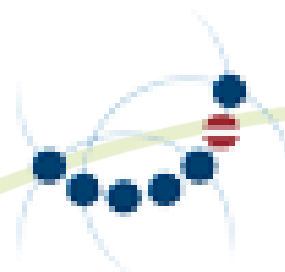
LU LITERATŪRAS,
FOLKLORAS
UN MĀKSLAS
INSTITŪTS



L N B

L A T V I J A S
N A C I O N Ā L Ā
B I B L I O T Ē K A

CLARIN-LV



**LIEPĀJAS
UNIVERSITĀTE**



**LATVIJAS
UNIVERSITĀTE**



**RĪGAS STRADIŅA
UNIVERSITĀTE**

Paldies par uzmanību!



CLARIN-LV

