

10 gadi ar CLARIN - sapņi un īstenība

Inguna Skadiņa

CLARIN-LV nacionālā koordinatore

05.03.2026.



Finansē
Eiropas Savienība
NextGenerationEU



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**



- Sagatavošanās (2008 – 2010)

CLARIN organizācijas izveide,
plānošana, prototipa izveide



- Izveide (2011 – 2015)

CLARIN tehniskās infrastruktūras izveide
un papildīšana ar valodas resursiem un rīkiem



- Izmantošana (2016 –)

CLARIN ir izmantojams kā pakalpojums





- **Sagatavošanās (2008 – 2010)**

CLARIN organizācijas izveide,
plānošana, prototipa izveide

- **Izveide (2011 – 2015)**

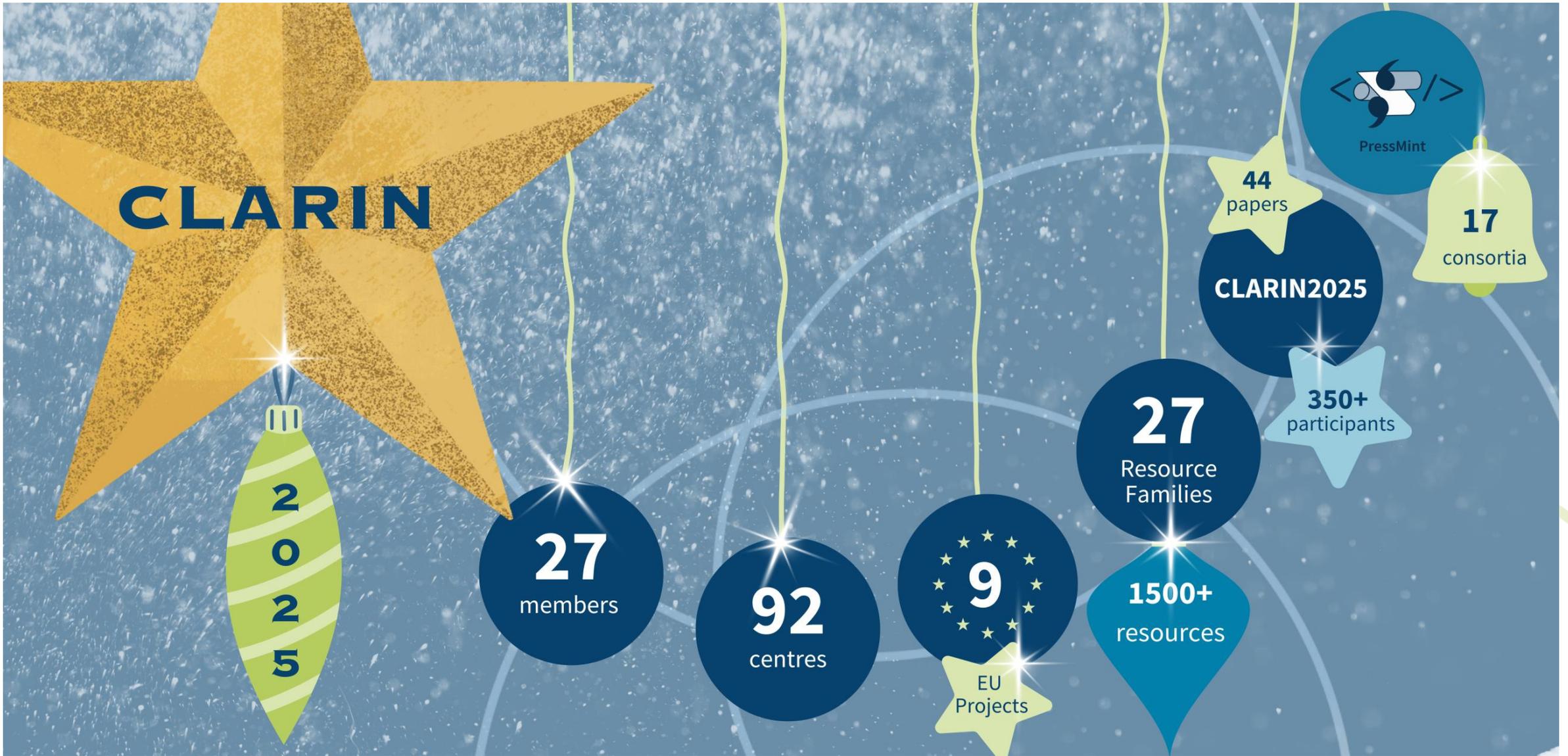
CLARIN tehniskās infrastruktūras izveide
un papildīšana ar valodas resursiem un rīkiem

- **Izmantošana (2016 –)**

CLARIN ir izmantojams kā pakalpojums

- Latvija CLARIN pievienojas 2016.gadā
- Kopš 2018.gada CLARIN-LV saņem nelielu finansējumu (ar pārtraukumiem) atsevišķām CLARIN-LV aktivitātēm



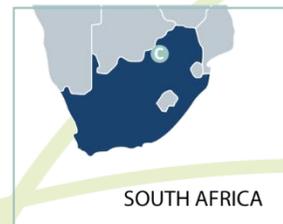


<https://www.clarin.eu/news/2025-review>

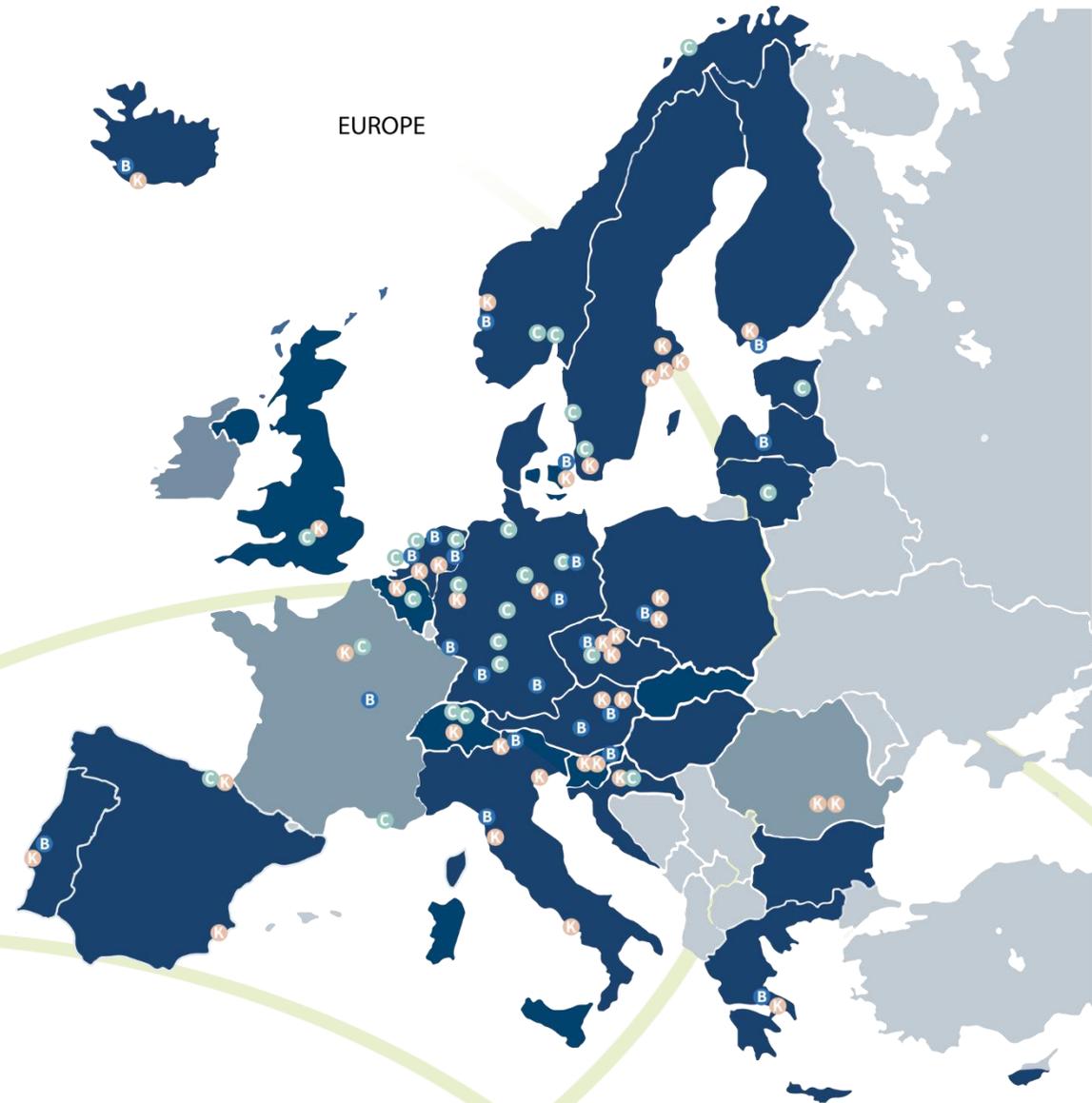
CLARIN 2026



- ERIC members
- Observers
- Countries with participating centres
- B Centre Providing Data
- C Centre Providing Metadata
- K Knowledge Centre



27 dalībvalstis

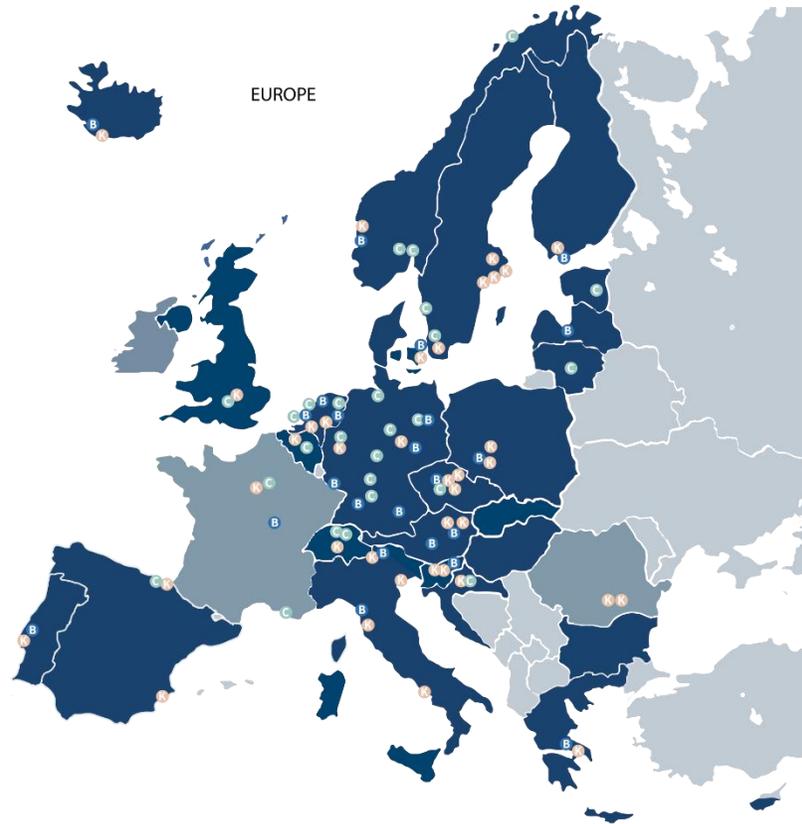


<https://www.clarin.eu/content/national-consortia>

CLARIN 2026



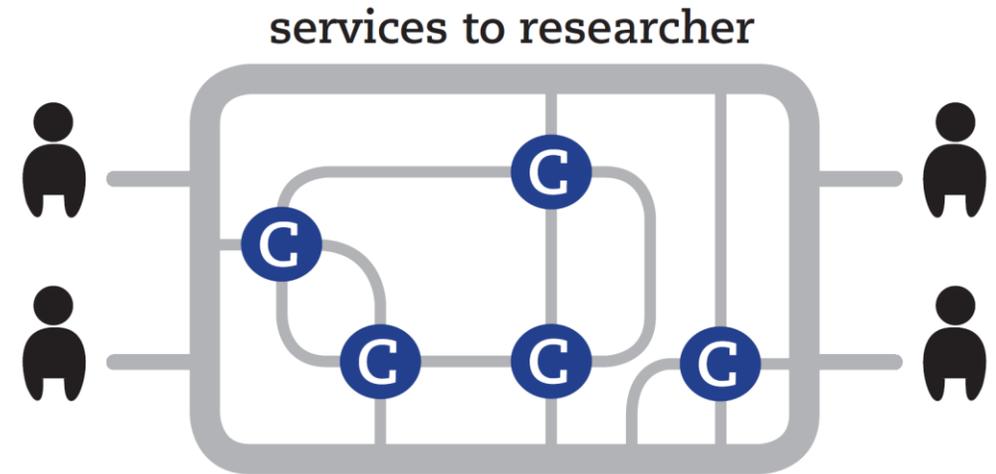
- ERIC members
- Observers
- Countries with participating centres
- ⓑ Centre Providing Data
- Ⓒ Centre Providing Metadata
- Ⓚ Knowledge Centre



- Vairāk nekā **60 tehnisko centru**, kas piedāvā vienkāršu un ilgtspējīgu piekļuvi digitālajiem valodas datiem un valodas apstrādes rīkiem, neatkarīgi no to atrašanās vietas
- Zināšanu centru tīkls (K-centri)
- Cieša sasaiste ar **Atvērtās zinātnes** programmu, fokuss uz sadarbību un **FAIR** principu ievērošanu

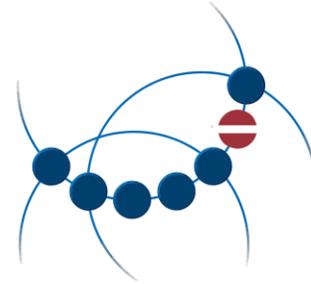
FAIR principles for open language resources

- **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

CLARIN-LV



LAMI

LU LITERĀRAS,
FOLKLORAS
UN MĀKSLAS
INSTITŪTS



Latvijas Universitātes
Matemātikas un informātikas institūts



**LATVIJAS
UNIVERSITĀTE**



L N B

LATVIJAS
NACIONĀLĀ
BIBLIOTĒKA



RTU LIEPĀJA



RTU RĒZEKNE



**RĪGAS STRADIŅA
UNIVERSITĀTE**

Par CLARIN

CLARIN (Common Language Resources and Technology Infrastructure –Vienota valodas resursu un tehnoloģiju infrastruktūra) ir daļa 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, savienošanai 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 piedalās trīs zināšanu centros (K-centros): SAFMORIL (zināšanu centrs morfoloģiski bagātām valodām), ELEXIS (leksikogrāfijas zināšanu centrs) un LLMs4SSH (zināšanu centrs par lielajiem valodas modeļiem humanitārajām un sociālajām zinātnēm). 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.

CLARIN-LV darbība tiek finansēta no bāzes finansējuma un ERAF projekta "Latvijas Universitāte un institūti Eiropas pētniecības telpā-izcilība pētniecībā un sadarbībā" (kopš 2025. gada septembra), kā arī no Valsts pētījumu programmas "Digitālās humanitārās zinātnes" projektā "Atvērtas un FAIR principiem atbilstīgas digitālo humanitāro zinātņu ekosistēmas attīstība Latvijā" (kopš 2022. gada decembra) un Eiropas Savienības Atvērto zināšanu un noturības mehānisma investīciju projekts "Valodu tehnoloģiju iniciatīva" (kopš 2023. gada). CLARIN Latvijā darbību atbalstījis ERAF projekts "Latvijas Universitāte un institūti Eiropas pētniecības telpā - ekselence, aktivitāte, mobilitāte, kapacitāte" (2018.-2022), Valsts pētījumu programmas "Humanitāro zinātņu digitālie resursi" projekts "Humanitāro zinātņu digitālie resursi: integrācija un attīstība" (5.10.2020-5.10.2022) un Valsts pētījumu programmas "Letonika latviskas un eiropeiskas sabiedrības attīstībai" projekts "Mūsdienu latviešu valodas izpēte un valodas tehnoloģiju attīstība" (20.12.2022-19.12.2024).



Reģistrācija CLARIN & DARIAH pavasara konferencei

Aicinām piedalīties CLARIN & DARIAH pavasara konferencei, kas notiks **2026. gadā 5. martā LU Sapere Aude zālē, Kalpaka bulvārī 4**. Reģistrācija dalībai konferencei: <https://ej.uz/registracija-clarin-dariah-konferencei-2026>. Vairāk informācijas digitalhumanities.lv un [CLARIN-LV lapās](http://CLARIN-LV.lapās).



www.clarin.lv
info@clarin.lv



CLARIN-LV darbību atbalsta ERAF projekts "Latvijas Universitāte un institūti Eiropas pētniecības telpā-izcilība

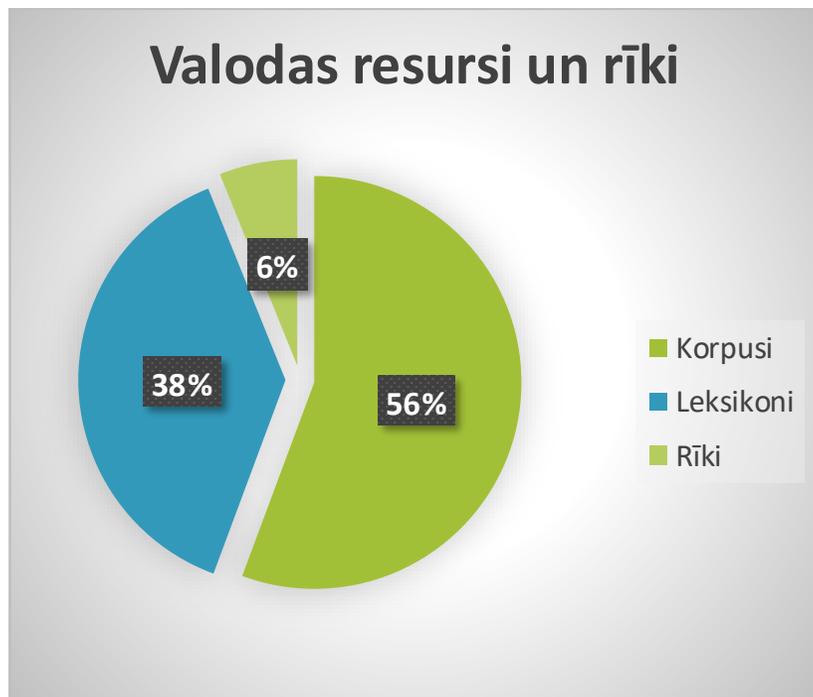
pētniecībā un sadarbībā"

CLARIN-LV valodas resursu repozitorijs



- 2020. gada 5. martā pirmo reizi tiek iepazīstināta DHZ kopiena
 - 6 korpusiem
 - 1 rīku
 - 1 leksikais resurss
- 2026. gada 5. martā:
 - 64 korpusi
 - 44 leksikoni (to versijas)
 - 7 rīki
- 2023. gada 16. janvārī iegūts B-centra statuss

CLARIN-LV valodas resursu repozitorijs



repository.clarin.lv

Kas jauns

Corpus

CLARIN Centre Of Latvian Language Resources And Tools

Corpus of Contemporary Latgalian Speech (MuLaR) (2026-03-02)

Autors (-i):

Martena, Sanita ; et al.

► show everyone

Apraksts:

The corpus consists of audio recordings and their transcripts. It documents natural, spontaneous speech, including field research recordings, interviews, TV and radio broadcasts.

Šajā vienumā ir 31 faili (3.92 GB).



Academic Use

LexicalConceptualResource

CLARIN Centre Of Latvian Language Resources And Tools

Historical Dictionary of Latvian Given Names

Autors (-i):

Siliņa-Piņķe, Renāte ; Rapa, Sanda ; Jansone, Ilga and Kazakevičs, Ņikita

Apraksts:

"Historical Dictionary of Latvian Given Names" (LPVV) is an online scientific dictionary that collects and describes Latvian given names documented in written sources spanning more than eight centuries. This dictionary ...

Šajā vienumā nav failu.



LexicalConceptualResource

CLARIN Centre Of Latvian Language Resources And Tools

Tēzauris.lv 2026 (Winter 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 410,000 entries based on 350 sources. The dictionary is enriched with phonetic, morphological, derivational, semantic ...

Šajā vienumā ir 5 faili (329.1 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

Populārākie resursi pēdējā nedēļā

Visvairāk skatītie vienumi

Populārākie pēdējā nedēļā

	Skatījumi
Tēzauris.lv 2026 (Winter Edition)	110
Corpus of Contemporary Latgalian Speech (MuLaR) (2026-03-02)	54
Balanced Corpus of Modern Latvian (LVK2022)	35
Corpus of the Tests of the State Language Proficiency Testing	33
Dictionary of Contemporary Latvian Language (MLVV) (2025-12-21)	24
Corpus of Legal Acts of the Republic of Latvia (Likumi)	23
Corpus of Students' Essays	21
Historical Dictionary of Latvian Given Names	20
Corpus of Contemporary Latgalian Speech (MuLaR)	19
Corpus of Latvian Early Novels (2025-03-11)	18

Datu uzglabāšanas servisi: deponēšana

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



CLARIN Centre of Latvian language resources and tools

Description:

CLARIN Centre of Latvian language resources and tools

Collections in this community:

- [Humanities and Cultural Heritage Resources of ILFA UL](#)
Humanities and Cultural Heritage Resources of ILFA UL
- [Language and Cultural Heritage Data of the UL Faculty of Humanities](#)
- [Language resources and tools of AiLab IMCS UL](#)
- [Language resources and tools of the National Library of Latvia](#)
Language resources and tools of the National Library of Latvia
- [Language resources of Latvian Language Agency](#)
Language resources of Latvian Language Agency
- [Language Resources of RTU Liepaja](#)
- [Language resources of the Faculty of Science and Technology, University of Latvia](#)
- [Language resources of Ventspils University of Applied Sciences](#)
- [Latgalian language and culture resources of RTU RA](#)
Latgalian language and culture resources of RTU RA
- [Livonian Language resources and tools by UL LI](#)
Livonian Language resources and tools by UL LI
- [Tilde Language resources and tools](#)

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

CLARIN-LV valodas resursi VLO



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



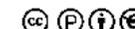
<https://vlo.clarin.eu>

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:



CLARIN-LV valodas resursi Dataverse-lv repozitorijā

The screenshot displays the Dataverse-LV repository interface. At the top, logos for the University of Latvia, Riga Stradiņš University (RSU), Riga Technical University, and CLARIN-LV are visible. The main content area shows search results for '1 to 10 of 115 Results'. On the left sidebar, there are filters for 'Dataverses (29)' and 'Datasets (328)', both highlighted with green boxes. Below these are filters for 'Files (1,174)'. The search results list includes:

- Historical Dictionary of Latvian Given Names** (Feb 18, 2026): A historical dictionary of Latvian given names documented in written sources from the 19th century.
- LATE Conversational Speech Corpus V1 (LATE-sarunas)** (Jan 23, 2026): A corpus of informal conversations, interviews, and public speeches with transcripts.
- Tēzauris.lv 2026 (Winter Edition)** (Dec 22, 2025): The largest open machine-readable dictionary for Latvian, containing over 410,000 entries.
- Tēzauris.lv 2025 (Autumn Edition)** (Dec 22, 2025): Similar to the 2026 edition, containing over 410,000 entries.

<https://dv.dataverse.lv>

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

CLARIN zināšanu centri

- CLARIN darbojas 38 zināšanu centri
- CLARIN-LV iesaistīties trīs zināšanu centros:
 - SAFMORIL - zināšanu centrs morfológiski bagātām valodām,
 - ELEXIS - leksikogrāfijas zināšanu centrs,
 - [LLMs4SSH](#) - zināšanu centrs par lielajiem valodas modeļiem humanitārajām un sociālajām zinātnēm.

CLARIN
K CENTRE



CLARIN konsorcija projekti

- Mērķis – apvienot atsevišķo valstu veidotos resursus, vienotā daudzvalodu resursā
- ParlaMint – salīdzināms 29 valstu parlamentu tekstu korpus
- Pressmint – daudzvalodu vēsturisko avīžu korpus

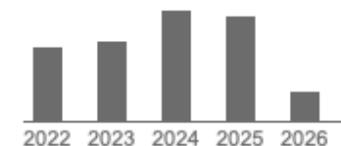


The ParlaMint corpora of parliamentary proceedings

Autori Tomaž Erjavec, Maciej Ogrodniczuk, Petya Osenova, Nikola Ljubešić, Kiril Simov, Andrej Pančur, Michał Rudolf, Matyáš Kopp, Starkaður Barkarson, Steinþór Steingrímsson, Çağrı Çöltekin, Jesse de Does, Katrien Depuydt, Tommaso Agnoloni, Giulia Venturi, María Calzada Pérez, Luciana D de Macedo, Costanza Navarretta, Giancarlo Luxardo, Matthew Coole, Paul Rayson, Vaidas Morkevičius, Tomas Krilavičius, Roberts Darģis, Orsolya Ring, Ruben van Heusden, Maarten Marx, Darja Fišer

Publicēšanas datums 2023/3
Žurnāls Language resources and evaluation

Atsauču kopskaits **Minēts 149**



CLARIN stratēģija

Our Users

- Academic User Base
- Non-Academic User Base

Our Offer

- Quality and Interoperability
- Accessibility and Usability
- Collaboration & Knowledge Exchange

Our Landscape

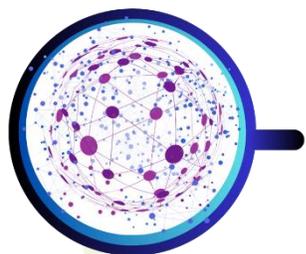
- Sustainability & Impact
- Central Hub
- National Nodes
- Collaboration Beyond CLARIN

<https://www.clarin.eu/content/clarin-strategy>



Paldies par uzmanību!

Valodu tehnoloģiju iniciatīva



DigiLATE



FLPP

FUNDAMENTĀLIE UN
LIETIŠĀJIE PĒTĪJUMU
PROJEKTI



VPP

Valsts pētījumu
programma



Finansē
Eiropas Savienība
NextGenerationEU

2027

Nacionālais
attīstības plāns

CLARIN-LV atbalsta projekts

Latvijas Universitāte un institūti Eiropas pētniecības telpā – izcilība pētniecībā un sadarbībā (Nr. 1.1.1.5/3/25/I/011)