

Vienota valodas resursu un tehnoloģiju infrastruktūra CLARIN

Inguna Skadiņa

CLARIN-LV nacionālā koordinatore

26.10.2021.



NATIONAL
DEVELOPMENT
PLAN 2020



EUROPEAN UNION
European Regional
Development Fund

INVESTING IN YOUR FUTURE

Vienota valodas resursu un tehnoloģiju infrastruktūra

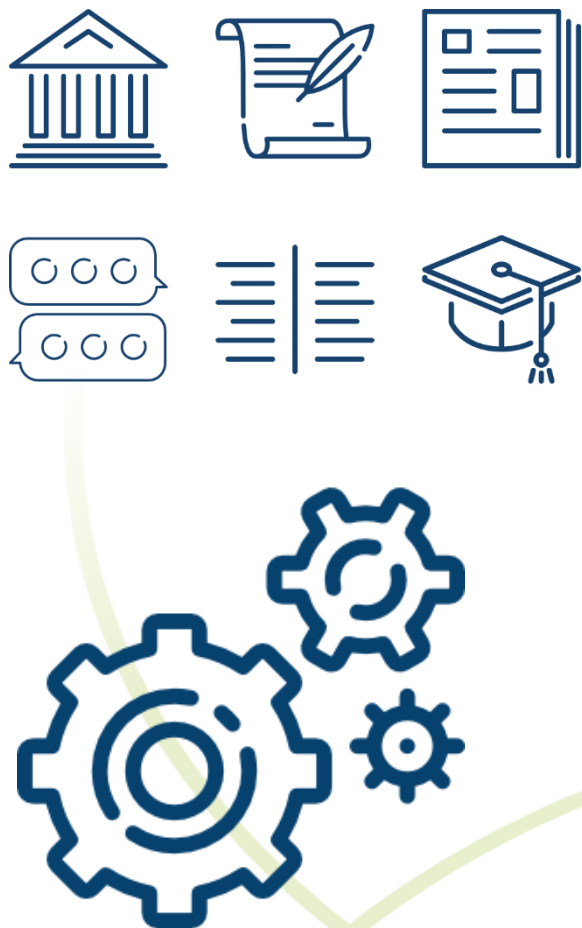
CLARIN (Common Language Resources and Technology Infrastructure) ir digitāla **Eiropas pētniecības infrastruktūra** (ERIC – European Research Infrastructure Consortium)

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: Vienota valodas resursu un tehnoloģiju infrastruktūra

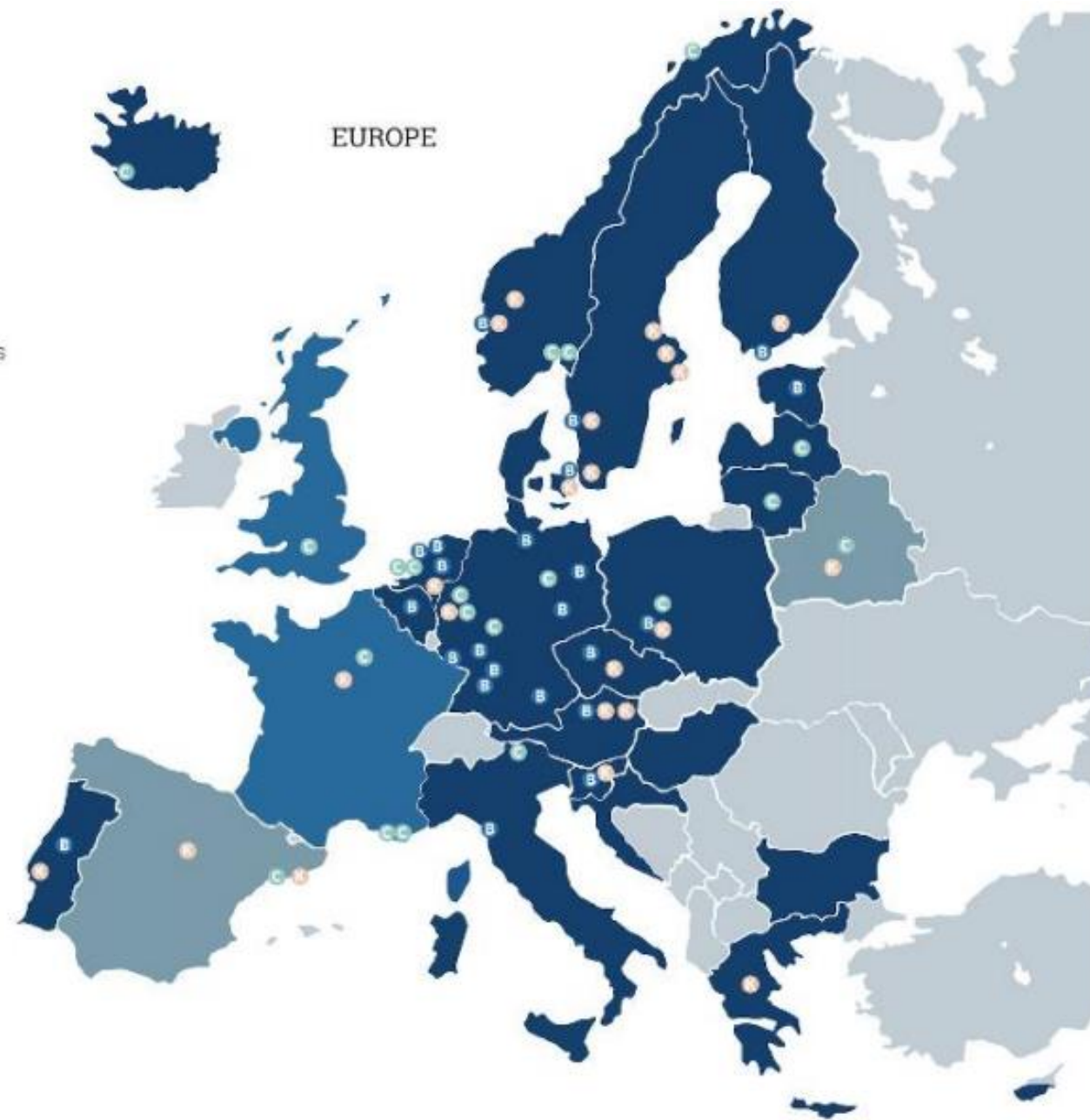


- Vienkārša un ilgtspējīga piekļuve **digitālajiem valodas datiem** (rakstiskā, runātā vai multimodālā formā)
- Mūsdienīgi **valodas apstrādes rīki**, valodas datu izpētei un analīzei neatkarīgi no to atrašanās vietas
- **Zināšanu infrastruktūras** mērķis ir nodrošināt, ka infrastruktūras mēroga zināšanas un ekspertīze būtu pieejamas

CLARIN ERIC - dalīta Eiropas pētniecības infrastruktūra

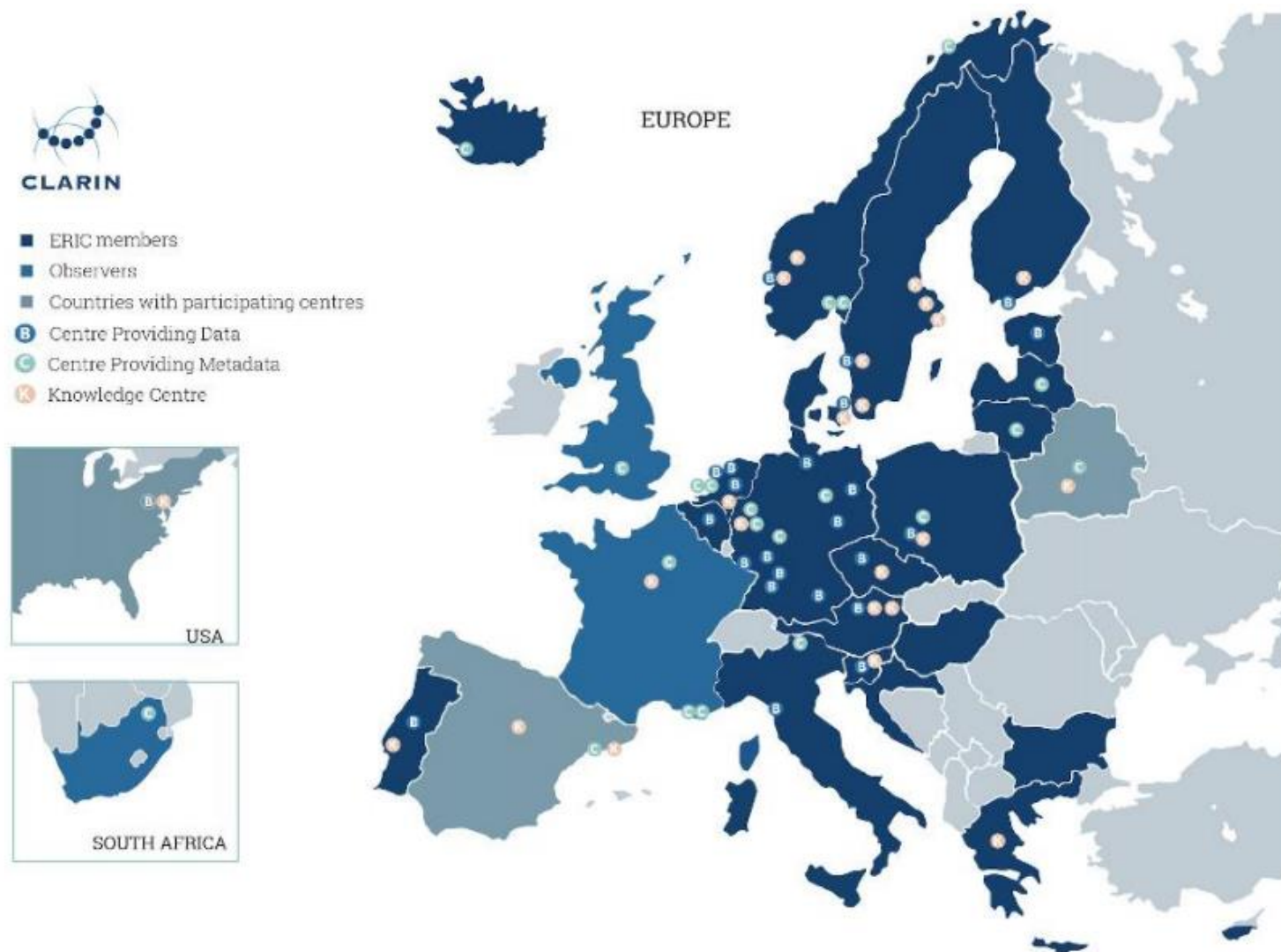


- ERIC members
- Observers
- Countries with participating centres
- B Centre Providing Data
- C Centre Providing Metadata
- K Knowledge Centre



CLARIN ERIC - dalīta Eiropas pētniecības infrastruktūra

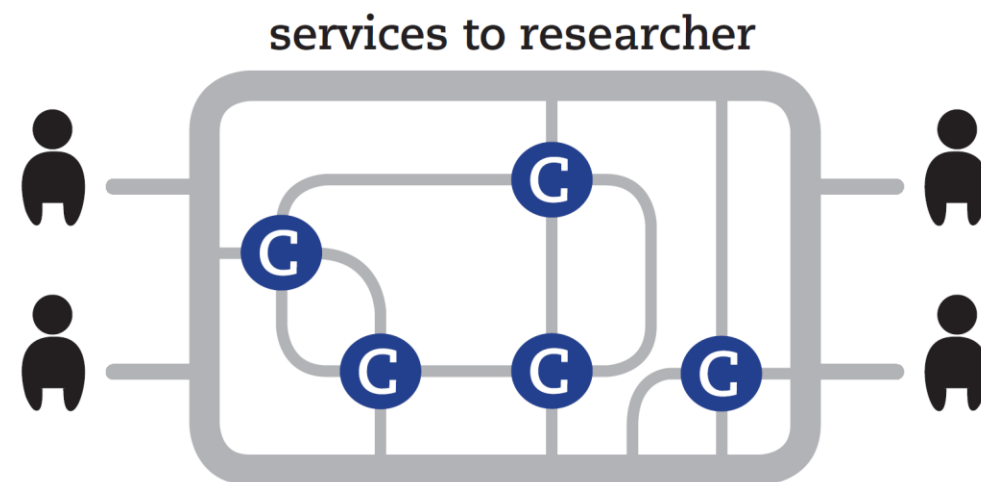
- 22 + 3 valstis
- Federētu servisu piedāvājums, piekļuve
 - FAIR datiem
 - Rīki darbplūsmu veidošanai
- > 60 reģistrētu centru (t.sk., 25 sertificēti B-centri)
- Zināšanu centru tīkls (K-centri)
- Cieša sasaiste ar Atvērtās zinātnes programmu, fokuss uz **sadarbspēju**



https://www.clarin.eu/sites/default/files/CLARIN2021_StateoftheInfrastructure_DeJong.pdf

FAIR principles for open language resources

- **F** – *findability*:
existence at a data **provider**, discovery, **persistence**, long-term **preservation** e.g. via CLARIN, ELG, ELE, LingHub, etc.
- **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), 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 pierakstīšanos. CLARIN piedāvā ilgtermiņa risinājumus un tehnoloģiju servissus 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 servisi tiek nepārtraukti atjaunināti un pilnveidoti. Dažās valstīs, kas CLARIN pievienojās nesen, CLARIN infrastruktūra tiek veidota. Latvija pievienojās CLARIN ERIC 2016. gada 1. jūnijā.



Atvērtās piekļuves nedēļā iepazīstinām ar CLARIN pētniecības infrastruktūru

Atvērtās piekļuves nedēļas ietvaros
26. oktobrī 10.00 – 10.30 Zoom
platformā interesenti tiks
iepazīstināti ar CLARIN pētniecības
infrastruktūru un CLARIN-LV
repozitoriju. Reģistrēties pasākumam
var šeit.

CLARIN-LV valodas resursu repozitorijs



Repozitorijs

Meklēšana korpusā

Par vietni



 Pieteikšanās



Atrast

Lingvistiskie dati un valodas apstrādes rīki

Citēšanas atbalsts (ar pastāvīgajiem ID)



Meklēt

[Detalizētā meklēšana](#)

Autors	Temats	Valoda (ISO)
Darģis, Roberts (10)	text (7)	Latvian (28)
Grūzītis, Normunds (10)	morphology (6)	English (2)
Levāne-Petrova, Kristīne (10)	dictionary (3)	German (2)
Auziņa, Ilze (7)	learner (3)	Latgalian (2)
Saulīte, Baiba (7)	treebank (3)	Lithuanian (2)

repository.clarin.lv

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

Corpus

CLARIN Centre Of Latvian Language Resources And Tools

Latvian Treebank v2.6

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 (16.44 MB).



Publicly Available

pository.clarin.lv/repository/xmlui/?locale-attribute=lv#

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

Meklēšana repozitorijā



korpus

[Detalizētā meklēšana](#)

Rāda 1 līdz 10 no 15 rezultātiem

1

2

>



Corpus

[CLARIN Centre Of Latvian Language Resources And Tools](#)

Latvian Treebank v2.8

(AiLab IMCS UL / 2021)

Autors (-i):

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

Šajā vienumā ir 1 fails (19.28 MB).

Publicly Available



Corpus

[CLARIN Centre Of Latvian Language Resources And Tools](#)

Rainis

(AiLab IMCS UL / 2018)

Autors (-i):

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

Šajā vienumā ir 2 faili (56.83 MB).

Publicly Available



Detalizētā meklēšana

Meklēšanas ierobežošana

Autors

Temats

Valoda (ISO)

Latvian (28)

English (2)

German (2)

Latgalian (2)

Lithuanian (2)

Tips

corpus (21)

lexicalConceptualResource (6)

toolService (2)

letver failus

Rāda 1 līdz 10 no 29 rezultātiem

1

2

3

>

Corpus

CLARIN Centre Of Latvian Language Resources

Latvian Treebank v2.8

(AiLab IMCS UL / 2021)

Autors (-i):
Rituma, Laura ; Pretkalniņa, Lauma ; Saulīte, Baiba ; Nešpore-Bērzk
Gunta ; Grūzītis, Normunds

Šajā vienumā ir 1 fails (19.28 MB).

Publicly

Corpus

CLARIN Centre Of Latvian Language Resources

Lithuanian-Latvian-Lithuanian Parallel Corpus (LILA)

(Ailab IMCS UL; Vytautas Magnus University / 2013)

Autors (-i):
Utka, Andrius ; Levāne-Petrova, Kristīne ; Vēvere, Daira ; Rābante-Buša,
Guna ; Kovalevskaitē, Jolanta ; Rimkutē, Erika ; Levāne-Petrova, Kristīne ;
Levāne-Petrova, Kristīne

Šajā vienumā nav failu.

Deponēt

Citēt

Iesniegš

Kārtot vienumus pēc

✓ Atbilstība

Nosaukums (augošā secībā)

Nosaukums (dilstošā secībā)

Izdošanas datums (augoši)

Izdošanas datums (dilstoši)

Rezultāti/lapa

5

✓ 10

20

40

60

80

100

Detalizētā meklēšana

Meklēšanas ierobežošana

Autors

Darģis, Roberts (10)
Grūzītis, Normunds (10)
Levāne-Petrova, Kristīne (10)
Auziņa, Ilze (7)
Saulīte, Baiba (7)
Pokratniece, Kristīne (6)
Znotiņš, Artūrs (6)
Pretkalniņa, Lauma (5)
Rituma, Laura (5)
Spektors, Andrejs (5)
Andronova, Everita (4)
Nešpore-Bērzkalne, Gunta (4)
Paikens, Pēteris (4)
Vēvere, Daira (4)
Rābante-Buša, Guna (3)
Frīdenberga, Anna (2)
Siliņa-Piņķe, Renāte (2)
Saulīte, Baiba (2)

Rāda 1 līdz 10 no 29 rezultātiem

1 2 3 >



Corpus

CLARIN Centre Of Latvian Language Resources And Tools

Latvian Treebank v2.8

(AiLab IMCS UL / 2021)



Autors (-i):

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

Šajā vienumā ir 1 fails (19.28 MB).

Publicly Available

Corpus

CLARIN Centre Of Latvian Language Resources And Tools

Lithuanian-Latvian-Lithuanian Parallel Corpus (LILA)

(Ailab IMCS UL; Vytautas Magnus University / 2013)



Autors (-i):

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

Šajā vienumā nav failu.

Citēšana

CLARIN-LV Repository Home / View Item

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:



CLARIN Centre Of Latvian Language Resources And Tools



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>.



bibtex

```
@misc{20.500.12574/41,  
  title = {Rainis},  
  author = {Spektors,  
    Andrejs and Grūzītis,  
    Normunds and Dargis, Roberts and Auziņš,  
    Ilze and Saulīte,  
    Baiba and Levāne-Petrova,  
    Kristīne},  
  url = {http://hdl.handle.net/20.500.12574/41},  
  note = {{CLARIN}-LV digital library at IMCS, University of Latvia},  
  copyright = {Creative Commons - Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)},  
  year = {2018} }
```



Share:



Citēšana

CLARIN-LV Repository Home / View Item

Rainis



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

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>.

BIBTEX

CMDI



Share:

Google

Rainis clarin



CLARIN Cent

Viss

Attēli

Ziņas

Maps

Video

Vēl

Rīki

Aptuveni 22 000 rezultāti (0,42 sekundes)

<https://repository.clarin.lv> > repository > xmlui > handle ▾

Rainis - Repository

autors A Spektors · 2018 — Spektors, Andrejs; Grūzītis, Normunds; Dargis, Roberts; et al., 2018, **Rainis**, CLARIN-LV digital library at IMCS, University of Latvia, ...

<https://vlo.clarin.eu> > record > europeana_58_item_47_... ▾

Raiņa Vasaras princīši un princītes : dzejoļi | CLARIN VLO

Raiņa Vasaras princīši un princītes : dzejoļi. Creator. **Rainis**, Jānis, 1865-1929. Description. Pirmizdevums. Collection. Europeana - National Library of ...

Lietojumi: meklēšana korpusā

Repozitorijs

Meklēšana korpusā

Par vietni



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

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

1 / 1

Emuāri	8M		Details
Latviešu valodas tīmekļa korpus	123M		Details
LVK2018 (beta)	12M		Details
morphocorpus	220k		Details
MuLa	1M		Details
Rainis	2M		Details
Saeima	22M		Details
Vidusskolēnu pārspridumi	226k		Details

Latviešu valodas resursi citu valstu CLARIN repozitorijos

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

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"] --> adv1["adv  
brīdī  
ncmsl2  
[brīdis]"]
    sent --> pred["pred  
crdParts"]
    sent --> punct["punct  
.  
zs  
[.]"]
    adv1 --> attr["attr  
Kādā  
pq0msln  
[kāds]"]
    adv1 --> attrC1["attrC1  
subrC1"]
    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  
vmypdmsnasnnp  
[sacelties]"]
    crdPart1 --> auxVerb1["auxVerb  
reduction:  
vcnrpii00an(esot)"]
    crdPart2 --> basElem2["basElem  
iepūtis  
vmnpdmsnasnnp  
[iepūst]"]
    crdPart2 --> auxVerb2["auxVerb  
reduction:  
vcnrpii00an(esot)"]
    obj["obj  
lapas  
ncfpa4  
[lapa]"] --> adv2["adv  
krūzē  
ncfsl5  
[krūze]"]
    obj --> basElem3["basElem  
iepūtis  
vmnpdmsnasnnp  
[iepūst]"]
```

Latviešu valodas
resursi citu valstu
CLARIN
repozitorijos

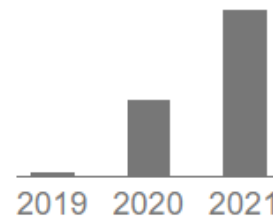
Universal Dependencies 2.5

Autori Daniel Zeman, et al.

Publicēšanas
datums 2020/11

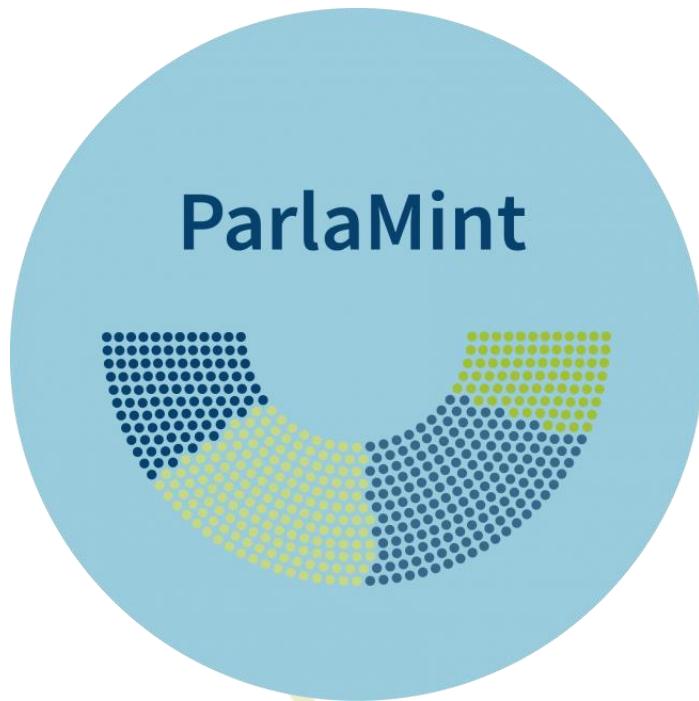
Avots LINDAT/CLARIAH-CZ, <http://hdl.handle.net/11234/1-3105>

Atsauču
kopskaits **Minēts 61**



Raksti pakalpojumā Scholar **Universal dependencies 2.5**
D Zeman, J Nivre, M Abrams, N Aeppli, Ž Agić... - LINDAT/CLARIAHCZ digital library at the Institute of ..., 2020
Minēts 61 [Saistītie raksti](#)

Latviešu valodas resursi citu valstu CLARIN repozitorijos



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 ; Darģis, 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. ⁰: 545.74 (related to the whole "parlamint20_lv") | ARF ⁰: 739.42 | Result is sorted

1 / 114 ▶▶▶

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

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

Latviešu valodas tekstu analīzes rīkkopa 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=Nom Ger	5	nsubj
2	Panks	Panks	NOUN	ncmsn1	Case=Nom Ger	1	flat:name
3	vēl	vēl	ADV	r0_	_	5	advmod
4	aizvien	aizvien	ADV	r0_	_	5	advmod
5	mitinās	mitināties	VERB	vmyip_330an	Evident=Fh Mo	0	root
6	koka	koks	NOUN	ncmsg1	Case=Gen Ger	7	nmod
7	dobumā	dobums	NOUN	ncmsl1	Case=Loc Ger	5	obl
8	Vērmānes	Vērmāne	NOUN	npfsg5	Case=Gen Ger	9	nmod
9	dārzā	dārzs	NOUN	ncmsl1	Case=Loc Ger	7	nmod
10	.	.	PUNCT	zs	_	5	punct

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

<http://nlp.ailab.lv/>

Populārākie valodas resursi (jūlijs,2020-marts, 2021)

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

Populārākie valodas resursi pēdējā nedēļā (15.oktobris-23.oktobris)

Item/Handle	Number of views
Latvian Treebank v2.8 (Rituma, Laura et al) (20.500.12574/55)	41 [+bots: 41]
Latvian Treebank v2.6 (Rituma, Laura et al) (20.500.12574/54)	28 [+bots: 28]
Corpus of the Tests of the State Language Proficiency Testing (Auziņa, Ilze et al) (20.500.12574/49)	26 [+bots: 26]
Rainis (Spektors, Andrejs et al) (20.500.12574/41)	24 [+bots: 24]
Corpus of the Saeima (the Parliament of Latvia) (Auziņa, Ilze et al) (20.500.12574/50)	21 [+bots: 21]
Tēzaurus.lv 2020 (Spektors, Andrejs et al) (20.500.12574/9)	21 [+bots: 21]
Corpus of Students' Essays (Levāne-Petrova, Kristīne et al) (20.500.12574/51)	20 [+bots: 20]

CLARIN virtuālā valodas resursu krātuve (VLO)

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

[Open](#) [Download](#) [Share](#)

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

[Open](#) [Download](#) [Share](#)

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

[Open](#) [Download](#) [Share](#)

<https://vlo.clarin.eu/>

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](#)

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 



CLARIN

27

CLARIN virtuālā valodas resursu krātuve (VLO)

Collection

latv

Europeana - National Library of Latvia: Collection 748774 (1331)
Europeana - National Library of Latvia: Collection 569516 (396)
Europeana - National Library of Latvia: Collection 57760 (31)

Kristīga sātība

(Part of Europeana - National Library of Latvia: Collection 748774)

Vecā druka

Latvian

Landing page for this record at www.europeana.eu

2

© P

Mīlestības rozītes jaunībai

(Part of Europeana - National Library of Latvia: Collection 748774)

Vecā druka

Latvian

Landing page for this record at www.europeana.eu

2

© P

Jaunām meitām laipni mājieni

(Part of Europeana - National Library of Latvia: Collection 748774)

Vecā druka

Latvian

Landing page for this record at www.europeana.eu

2

© P

Zviedru spička

(Part of Europeana - National Library of Latvia: Collection 748774)

Vecā druka

Latvian

Landing page for this record at www.europeana.eu

2

© P

Vienots satura meklētājs

Text layer CQL query

Search for Text layer Contextual Query Language (CQL) and show up to hits per endpoint

▼ **est_news_2011_300K** – ASV Leipzig

Valjala põhikooli vene **keele** õpetaja ja kooli usaldusisik Ulvi Põldema on samuti optimistlikult meelestatud.

Kuid soovides pakkuda tasuta eestikeelset kõrgharidust, tuleks valitsusleppes ette näha kindlad võimalused ja vahendid eesti **keele** kui teaduskeele arenguks ja õpetamiseks iga ülikooli igal erialal.

Tallinnas on peaaegu võimatu leida eesti **keele** kõrgtaseme kursuseid - seda isegi oma raha eest, rääkimata tasuta koolitusest.

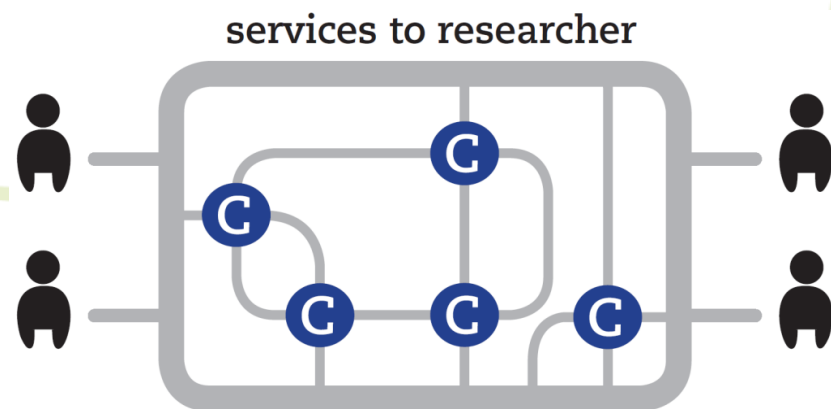
1. Linnast välja kolides osta parem kohe maastur (56) 2. Uus eesti **keele** eksam külvab koolides hirmu (33) 3. TÄISMAHUS: Lasteaiad võitlevad miljonite pärast (39) 4. Kümme või Einaste peab tooma pjedestaalikoha!

Eelmise reede palvusel Iraanis, kui imaam oli natuke aega rääkinud pärsia **keele** s, vahetas ta selle araabia keele vastu selges teadmises, et see üle kantakse, ja rääkis edasi araabia keeles, julgustades meeleavaldajaid Kairos Tahriri väljakul.

<https://contentsearch.clarin.eu/>

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)



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.

Previously chosen home organisation

University of Latvia
Latvia



Home organisation list



All countries



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



United Kingdom

Vienotā pierakstīšanās

Vienotā pierakstīšanās ar licenci aizsargātiem valodas resursiem



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.

Vienotā pierakstīšanās

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](#)  BY NC ID NORED

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](#)

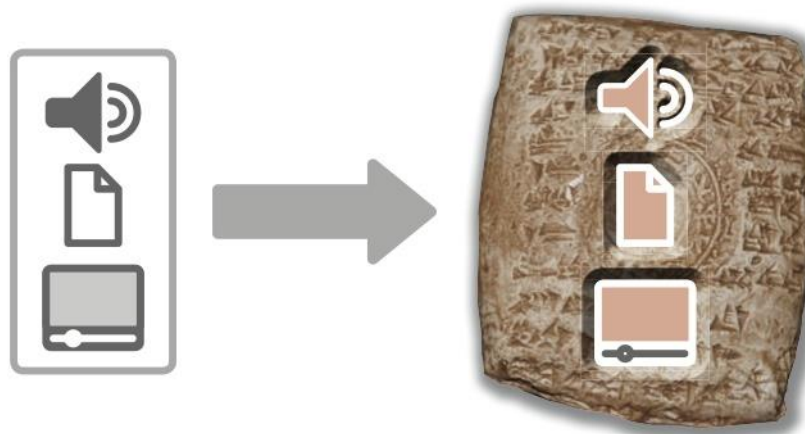
match

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 tall, 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 dog is dead. The remaining 42 per cent were ambidextrous. Many owners think their dogs are intelligent when they were briefly joined in turn by their owner, an unfamiliar human, a dog, or a cat. They are efficient layers of insulation. In hot weather, licking compensates for the lack of sweat glands and helps to cool down the fur. Historically, cats have purred or unconsciously, their internal clocks had told them to expect food. The purr is a sound that 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 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 years old - 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 cat or dog, by understanding how to clone you could be one step ahead in the game of cat or dog. Compare this with the risible image of an ill-fated ancestor attempting to clone a 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 is an unfamiliar dominant dog, a Belgian Shepherd. The dog's tail wags to compensate for its lack of sweat glands and helps to cool down the fur. Historically, cats have purred or unconsciously, their internal clocks had told them to expect food. The purr is a sound that 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 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 years old - almost four times as many as cats, which have approximately 470. A

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



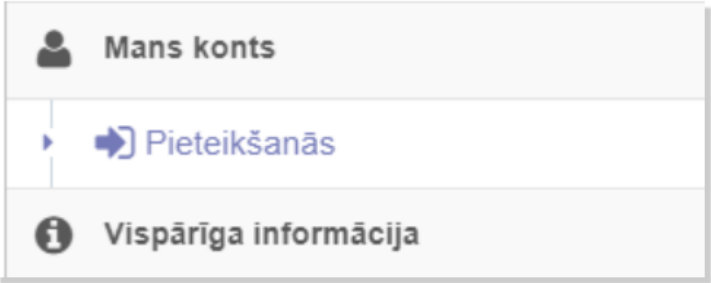
https://www.clarin.lv/attachments/CLARIN%202020_resursu_iesniegsana.pdf

Kā noguldīt

Vienumus var noguldīt tikai autentificēti lietotāji. Ja jūsu organizācija nav atrodama organizāciju sarakstā "Pieteikšanās" dialogā, reģistrējieties clarin.eu un autentificējieties, izmantojot "clarin.eu tīmekļa vietnes kontu (clarin.eu website account)". Ja nevarat izmantot nevienu iepriekšminēto autentifikācijas metodi vai ja rodas problēmas, sazinieties ar mūsu [palīdzības dienestu](#), un mēs jums varam izveidot kontu.

1. darbība: pieteikšanās

Lai sāktu jaunu iesniegumu, vispirms ir jāpiesakās. Labajā izvēlnes paneļā sadaļā "Mans kots" noklikšķiniet uz "Pieteikties".



1. attēls. Pieteikšanās izvēlne

2. darbība: jauna iesnieguma sākšana

<https://repository.clarin.lv/repository/xmlui/page/deposit?locale-attribute=lv>

?

Ko tu vari darīt?

DEPOSIT

CITE

🎯

Pārlūkot

>

Visi repozitoriji

▼

👤

Mans kots

➡

Pieteikšanās

i

Vispārīga informācija

📁

Deponēt

🗣

Citēt

🔄

Iesniegšanas dzīves cikls

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 izstrādātājus 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



CLARIN zināšanu centrs morfoloģiski bagātām valodām SAFMORIL

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

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

CLARIN valodas resursu grupas, kurās ietverta latviešu valoda

- Vēsturiskie korpusi
- Valodas apguvēju korpusi
- Parlamentāro datu korpusi
- Runātās valodas korpusi
- Leksiskie resursi

Konference "Latviešu valodas digitālie resursi un rīki izglītībai"

2021.gada 2. jūnijā CLARIN-LV un VPP "Humanitāro zinātņu digitālie resursi: integrācija un attīstība" rīkoja konferenci "Latviešu valodas digitālie resursi un rīki izglītībai".

Konferences materiāli:

- *Aleksandrs Mārtiņš Blūms* (Latvijas Republikas Izglītības un zinātnes ministrija). Atvērtā zinātne un digitālās humanitārās zinātnes izstrāde (Digital Humanities in Latvia)
- *Frančeska Frontini* (Francesca Frontini) (CLARIN lietotāju iesaistes direktore). CLARIN popularizēšana un mācīšana koronavīrusa laikā (Disseminating and Teaching CLARIN in the time of Corona. Challenges and opportunities)
- *Katrīna Marheineke* (Katrin Marheinecke) (DFKI). Ieskats Eiropas valodu tīkla (European Language Grid) un Eiropas valodu tīkla (European Language Grid and European Language Equality) projektos (European Language Grid and European Language Equality: A brief introduction)
- *Inguna Skadiņa* (LU MII). CLARIN-LV infrastruktūra pētniecībai un izglītībai
- *Sanita Reinsone* (LU LFMI). VPP „Humanitāro zinātņu digitālie resursi”: atskats uz pirmo pusgadu
- *Ilze Auziņa, Pēteris Paikens un Roberts Darģis* (LU MII). Jaunākie Mākslīgā intelekta laboratorijas valodas resursi izglītībai un pētniecībai
- *Uldis Bojārs* (LNB). Atvērtie dati un saistītie dati Latvijas Nacionālajā bibliotēkā
- *Raivis Skadiņš* (Tilde). No vārdnīcas – līdz mašintulkošanai: Tildes valodas resursi pētniecībai un izglītībai
- *Anna Briška* (RTA). Mūsdienu latgaliešu valodas korpusa „jaunā elpa”



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

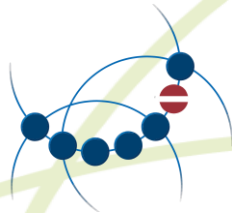
Country

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.

Paldies par uzmanību!

CLARIN-LV



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



Projekts "Humanitāro zinātņu digitālie resursi: integrācija un attīstība" (VPP-IZM-DH-2020/1-0001) tiek īstenots Valsts pētījumu programmas "Humanitāro zinātņu digitālie resursi" ietvaros.