

CLARIN: Vienota valodas resursu un tehnoloģiju infrastruktūra

Inguna Skadiņa

CLARIN Latvija nacionālā koordinatore

Common Language Resources and Technology
Infrastructure



Kas ir CLARIN?



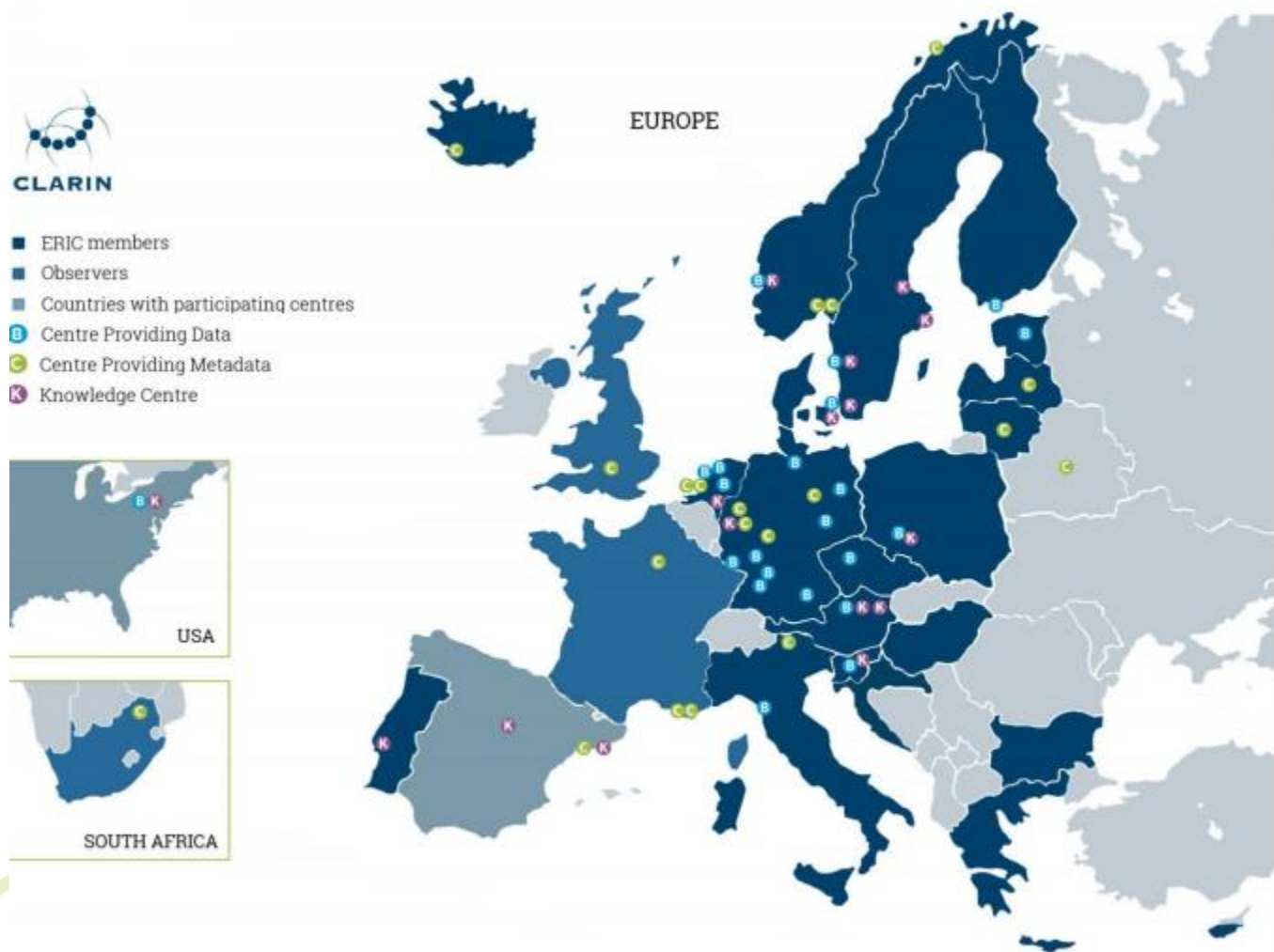
CLARIN ir **vienota** valodas resursu un tehnoloģiju Eiropas pētniecības infrastruktūra (ERIC),

- kas nodrošina visu jomu, īpaši **humanitāro un sociālo zinātņu**, pētniekiem vienkāršu un pastāvīgu piekļuvi
- **digitāliem valodas datiem** (rakstītā, runātā, video vai multimodālā formā)
- un **atbilstošiem rīkiem**, lai tos pētītu, izmantotu, analizētu vai kombinētu, neatkarīgi no tā, kur tie atrodas.

CLARIN ERIC

Konsorciju veido

- 24+ 3 valstis
- 24 B centri



CLARIN-LV



Kontakti

CLARIN
Common Language Resources and
Technology Infrastructure

CLARIN
K CENTRE



www.clarin.lv
info@clarin.lv

CLARIN-LV valodas resursu repozitorijs



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 (4)

Levāne-Petrova, Kristīne (4)

Darģis, Roberts (3)

Pretkalniņa, Lauma (3)

Saulīte, Baiba (3)

... View More

Subject

corpus (2)

Abstract Meaning Rep ...(1)

child language (1)

collocations and idioms (1)

coreferences (1)

... View More

Language (ISO)

Latvian (9)

Latgalian (1)

Lithuanian (1)

repository.clarin.lv

CLARIN-LV valodas resursu repozitorijs

What's New

Corpus

CLARIN Centre Of Latvian Language Resources And Tools

The Corpus of Early Written Latvian



Author(s):

Andronova, Everita ; Spektors, Andrejs ; Vanags, Pēteris ; Baltiņa, Maija ; Trumpa, Anta ; Trumpa, Edmunds ; Grūzītis, Normunds ; Siliņa-Piņķe, Renāte ; Fridenberga, Anna ; Skrūzmane, Elga ; Ķauķīte, Sintija

Description:

The Corpus of early written Latvian 'SENIE'(16th-18th cc.) provides access to the texts of written Latvian of the 16th–18th century, and its aim is to facilitate studies of early Latvian in general (e.g. the lexis, morphology ...

 This item contains 1 file (31.65 KB).

Publicly Available

Corpus

CLARIN Centre Of Latvian Language Resources And Tools

The Balanced Corpus of Modern Latvian



Author(s):

Levāne-Petrova, Kristīne ; Dargis, Roberts ; Levāne-Petrova, Kristīne

Description:

LVK2018 is the 10 million general language representative corpus of contemporary Latvian. It is balanced corpus, aiming to cover the variety of existing texts in some estimated proportions. The corpus represents five ...

 This item contains no files.

Corpus

CLARIN Centre Of Latvian Language Resources And Tools

Latvian Treebank v2.5



Author(s):

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

Description:

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

Detalizēta meklēšana

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

ToolService

CLARIN LATVIA

NLP-PIPE: Latvian NLP Tool Pipeline

(Ailab IMCS UL / 2015-05-05)



Author(s):

Znotiņš, Artūrs

This item contains no files.

Search

×

Detalizēta meklēšana

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 LAT

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 LAT

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 LAT

NLP-PIPE: Latvian NLP Tool Pipeline

(Ailab IMCS UL / 2015-05-05)

Author(s):

Znotiņš, Artūrs

This item contains no files.

Detalizēta meklēšana

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 virtuālā valodas resursu krātuve (VLO)

Virtual Language Observatory

Search

Contributors

Help



VLO / Faceted search



Search through 1,198,049 records



Showing all records (781,153 results)

Results per page: 10

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

Language

Type to filter or search for more

English (146386)
Dutch (118386)
German (61065)
multiple languages (35140)
Slovenian (30203)
French (27005)
Spanish; Castilian (16613)
Bulgarian (14366)
Polish (8332)
Afrikaans (7844)
[more...](#)

Collection

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

Nganasan Spoken Language Corpus (NSLC)

(Part of Hamburger Zentrum für Sprachkorpora (HZSK))

The Nganasan Spoken Language Corpus (NSLC) has been created as part of Corpus based grammatical studies on Nganasan project (supported by the German Research Grant; WA3153/2-1). The Spoken Nganasan Corpus contains the same text samples in at least three languages: The original text in Nganasan with translations mostly ...

Nganasan Russian

Landing page for this record



EXMARaLDA Demo corpus

(Part of Hamburger Zentrum für Sprachkorpora (HZSK))

A selection of short audio and video recordings in various languages to be used for instruction or demonstration of the EXMARaLDA system.; HIAT (simplified); HIAT; free comment; suprasegmental information; accentuation/stress; English translation; Standard German translation; German translation; Englisch translation; code-switch

German English French Spanish; Cas.. Turkish ... (+8)

Landing page for this record



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

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

🔍 Ⓐ €

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 virtuālā valodas resursu krātuve (VLO)

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).

Latvian

🏠 Landing page for this record

1 📄

?

NLP-PIPE: Latvian NLP Tool Pipeline

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

⊕ NLP-PIPE is a modular toolchain that allows researchers to combine multiple natural language processing tools in a unified framework. It supports a wide range of annotation services for Latvian, including tokenization, morphological tagging, lemmatisation, universal dependency parsing, and named entity recognition. In ...

Latvian

🏠 Landing page for this record

1 📄

?

Annotated longitudinal corpus of Latvian children's language

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

⊕ The collection contains three longitudinal corpora of monolingual Latvian speaking children, and one longitudinal corpus of simultaneous Latvian-Russian bilingual child. Participants were recorded for 30 minutes each week for 16 months, resulting in 134 hours of speech. 34 hours of obtained speech samples are orthographically transcribed.

Latvian

🏠 Landing page for this record

1 📄

?

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

CC BY

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

CC BY

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

CC BY

Zviedru spička

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

Vecā druka

Latvian

Landing page for this record at www.europeana.eu

2

CC BY

Vienots satura meklētājs

Text layer CQL query

keele



Search for

Estonian ▾

Text layer Contextual Query Language (CQL) ▾

in

3 selected collections ▾

and show up to

10

hits per endpoint

▼ **est_news_2011_300K** – ASV Leipzig

View

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.

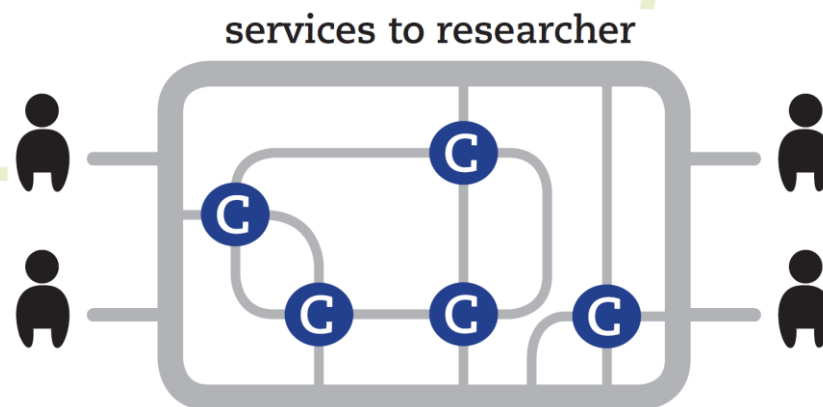
1. Linnast välja kolides osta parem kohe maastur (56) 2. Uus eesti **keele** eksam külvab koolides hirmu (33) 3. TÄISMAHUS: Lasteaiad võitlevad miljonite pärast (39) 4. Kümmel või Einaste peab tooma pjedestaalikoha!

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/>

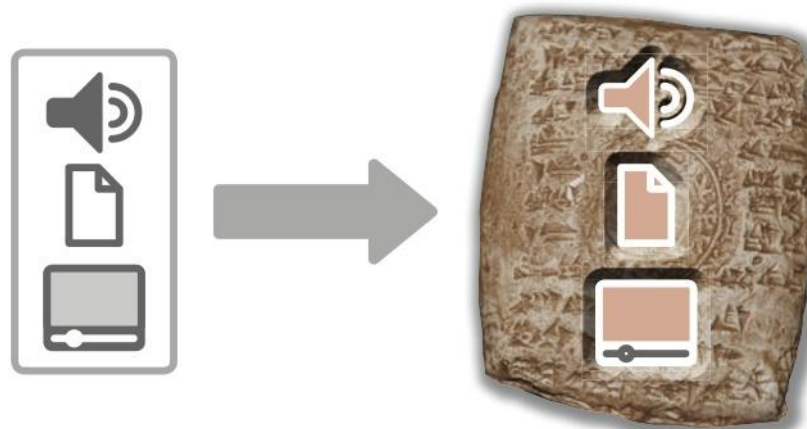
CLARIN datu un pakalpojumu centri

- CLARIN dalītās infrastruktūras pamats ir CLARIN servisu nodrošināšanas centri
- CLARIN ERIC ir noteicis stingras prasības centriem: centram jābūt **stabilam tehniskam un institucionālam pamatam**.
- Šie centri nodrošina zinātnieku kopienai **pastāvīgu** un **ilgtspējīgu** piekļuvi valodas resursiem, servisiem un zināšanām.



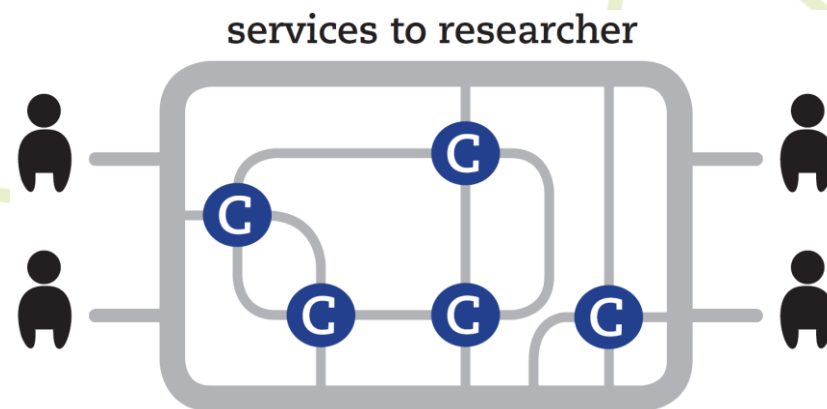
Datu uzglabāšanas servisi

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



CLARIN vienotā pierakstīšanās

CLARIN ERIC pamatprincips ir **vienota pierakstīšanās**, t.i., lai piekļūtu CLARIN ERIC infrastruktūrā iekļautajiem valodas resursiem, 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)



Vienota pierakstīšanās



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.

Accept

CLARINO

Corpuscle :: ICAME – BROWN Family :: Concordance

Localizati

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 sai 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

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



CLARIN zināšanu centrs morfolóģiski 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