|  |  |
| --- | --- |
| Title | Overview of lexical resources |
| Version | 1.1 |
| Author(s) | Darja Fišer, Jakob Lenardič |
| Date | 23-07-2019 |
| Status | For distribution |
| Distribution | BoD, NCF, UI |
| ID | CE-2019-1464 |

**Table of contents**

[1. Introduction 2](#_Toc14188521)

[2. Lexical resources in the CLARIN infrastructure 2](#_Toc14188522)

[2.1. Lexica 3](#_Toc14188523)

[2.1.1. VLO 3](#_Toc14188524)

[2.1.2. Availability 3](#_Toc14188525)

[2.1.3. Language 5](#_Toc14188526)

[2.1.4. Size 5](#_Toc14188527)

[2.1.5. Linguistic information 6](#_Toc14188528)

[2.1.6. Licence 6](#_Toc14188529)

[2.1.7. Overview of the lexica in the CLARIN infrastructure 7](#_Toc14188530)

[2.2. Dictionaries 16](#_Toc14188531)

[2.2.1. VLO 16](#_Toc14188532)

[2.2.2. Availability 16](#_Toc14188533)

[2.2.3. Language 18](#_Toc14188534)

[2.2.4. Size 19](#_Toc14188535)

[2.2.5. Linguistic information 19](#_Toc14188536)

[2.2.6. Licence 20](#_Toc14188537)

[2.2.7. Overview of the dictionaries in the CLARIN infrastructure 20](#_Toc14188538)

[2.3. Conceptual resources 30](#_Toc14188539)

[2.3.1. VLO 30](#_Toc14188540)

[2.3.2. Availability 30](#_Toc14188541)

[2.3.3. Language 30](#_Toc14188542)

[2.3.4. Size 31](#_Toc14188543)

[2.3.5. Linguistic information 31](#_Toc14188544)

[2.3.6. Licence 31](#_Toc14188545)

[2.3.7. Overview of the conceptual resources in the CLARIN infrastructure 32](#_Toc14188546)

[2.4. Glossaries 36](#_Toc14188547)

[2.4.1. VLO 36](#_Toc14188548)

[2.4.2. Availability 36](#_Toc14188549)

[2.4.3. Language 37](#_Toc14188550)

[2.4.4. Size 37](#_Toc14188551)

[2.4.5. Linguistic information 37](#_Toc14188552)

[2.4.6. Licence 38](#_Toc14188553)

[2.4.7. Overview of the glossaries in the CLARIN infrastructure 38](#_Toc14188554)

[2.5. Wordlists 42](#_Toc14188555)

[2.5.1. VLO 42](#_Toc14188556)

[2.5.2. Availability 42](#_Toc14188557)

[2.5.3. Language 43](#_Toc14188558)

[2.5.4. Size 44](#_Toc14188559)

[2.5.5. Linguistic information 44](#_Toc14188560)

[2.5.6. Licence 44](#_Toc14188561)

[2.5.7. Overview of the wordlists in the CLARIN infrastructure 45](#_Toc14188562)

[3. Conclusions 52](#_Toc14188563)

# Introduction

In the following report, we present an overview of lexical resources that are made available through the CLARIN infrastructure.

The report was conducted in two steps:

1. by manually searching the national repositories with the following keywords: “lexicon”, “dictionary”, “glossary”, “wordlist”
2. with input provided by CLARIN National Coordinators and User Involvement Representatives

In contrast with the previous overviews, we did not conduct the survey by searching the VLO, because it returns – unlike the national repositories – a disproportionately large number of resources harvested from external infrastructures (e.g., the LDC Catalogue or ELRA) that are not directly available (either for download or online browsing) to the end user.

Full results are available in a [Google Docs Spreadsheet](https://docs.google.com/spreadsheets/d/1uewa8CHsFd6vvPYMx42ZQMhWE08GYVA38h55yYv3aQE/edit#gid=0).[[1]](#footnote-1) A total of 246 resources were originally identified, with an additional 59 resources added on the basis of input by User Involvement representation and National Coordinators, whom we would like to thank for their invaluable input. In this report, we include almost all the listed resources (290 out of the 306 listed in the Spreadsheet). The reason for the small discrepancy is that for some resources included in the spreadsheet subsets are listed separately, whereas this report only includes the umbrella entries. The monolingual resources cover 26 languages (Arabic, Croatian, Czech, Dagaare, Danish, Dutch, English, Estonian, Finnish, French, Frisian, German, Greek, Icelandic, Italian, Latvian, Maltese, Ngbugu, Polish, Portuguese, Serbian, Slovenian, Swedish), whereas the multilingual corpora account for more than 50 language combinations. In this survey we did not investigate the encoding of the lexical resources, which would be an interesting future exercise.

# Lexical resources in the CLARIN infrastructure

Given the size of this resource family, we organized the resources into to the following 5 categories for a better overview:

1. [Lexica](#_Lexica) (76 resources)
2. [Dictionaries](#_Dictionaries) (94 resources)
3. [Conceptual resources](#_Conceptual_resources) (29 resources)
4. [Glossaries](#_Glossaries) (33 resources)
5. [Wordlists](#_Wordlists) (56 resources)

The classification of the resources was pragmatic and while some resources could fit in more than one category we only list it in the primary one. For the classification we applied a combination of content-, structure- and usage-related criteria:

* We classify a resource as a lexicon if its primary use is in NLP applications. Such a resource typically contains an extensive lexical inventory with specific linguistic information (e.g., morphosyntax, sentiment).
* We classify a resource as a dictionary if it was originally created for human use. Such a resource is typically organized around words and contains information on their senses, definitions, pronunciation, etc.
* We combine concept-based resources, such as wordnets, framenets, thesauri and ontologies into conceptual resources. Such resources are typically semantics-focused and interlinked with semantic relations (e.g. hypernymy, hyponymy).
* We classify a resource as a glossary if its structure is simpler than those of dictionaries and contains domain-specific terminology and/or expressions.
* We classify resources which only provide alphabetical or frequency-based lexical inventories as wordlists.

## Lexica

There are 76 lexica available in the CLARIN infrastructure. They are presented in detail in Table [1](#Lexica).

### VLO

All the lexica in in Table [1](#Lexica) can be found in the VLO expect for the following 4:

1. [A machine-readable dictionary of Egyptian Arabic](http://hdl.handle.net/11022/0000-0007-C266-B)
2. [A machine-readable Persian-English dictionary](http://hdl.handle.net/11022/0000-0000-0029-2)
3. [A machine-readable glossary of Egyptian Arabic](http://hdl.handle.net/11022/0000-0000-001D-0)
4. [A machine-readable Persian-English glossary of verbs](http://hdl.handle.net/11022/0000-0000-0027-4)

### Availability

The following 18 lexica are available for download and online browsing:

1. [Beseda Corpus Lemmatisation Lexicon](http://hdl.handle.net/11356/1141)
2. [CzEngClass 0.2](http://hdl.handle.net/11234/1-2824)
3. [CzEngVallex](http://hdl.handle.net/11234/1-1512)
4. [Database of Modern Icelandic Inflections](http://clarin.is/en/resources/dmii/)
5. [DeriNet 1.6](http://hdl.handle.net/11234/1-2873)
6. [Emoji Sentiment Ranking 1.0](http://hdl.handle.net/11356/1048)
7. [EngVallex](http://hdl.handle.net/11858/00-097C-0000-0023-4337-2)
8. [Lexicon of historical Slovene imp25k 1.1](http://hdl.handle.net/11356/1032)
9. [Morphological lexicon Sloleks 2.0](http://hdl.handle.net/11356/1230)
10. [Old Swedish morphology (2017-10-16)](http://hdl.handle.net/10794/25)
11. [Old Swedish morphology (2017-10-16)](http://hdl.handle.net/10794/25)
12. [Parole+ (2017-10-16)](http://hdl.handle.net/10794/33)
13. [PDT-Vallex](http://hdl.handle.net/11858/00-097C-0000-0023-4338-F)
14. [SALDO's morphology (2017-10-16](http://hdl.handle.net/10794/36))
15. [Simple lexicon](http://hdl.handle.net/10794/38)
16. [The Database of Estonian Multi-Word Expressions](https://metashare.ut.ee/repository/browse/the-database-of-estonian-multi-word-expressions/4d8252e8463411e2a6e4005056b400243ed5ec91ec5044bbb0e85b2ce16f472b/)
17. [VALLEX 3.0](http://hdl.handle.net/11234/1-2307)
18. [Emotional Annotations Dictionary](https://clarin-pl.eu/dspace/handle/11321/553)

The following 50 lexica are available for download only:

1. [A machine-readable dictionary of Egyptian Arabic](http://hdl.handle.net/11022/0000-0007-C266-B)
2. [A machine-readable glossary of Egyptian Arabic](http://hdl.handle.net/11022/0000-0000-001D-0)
3. [A machine-readable Persian-English dictionary](http://hdl.handle.net/11022/0000-0000-0029-2)
4. [A machine-readable Persian-English glossary of verbs](http://hdl.handle.net/11022/0000-0000-0027-4)
5. [Automatically constructed multiword lexicon hrMWELex v0.5](http://hdl.handle.net/11356/1177)

1. [Automatically constructed multiword lexicon slMWELex v0.5](http://hdl.handle.net/11356/1179" \t "_parent)
2. [Automatically constructed multiword lexicon srMWELex v0.5](http://hdl.handle.net/11356/1178" \t "_parent)
3. [Automatically stress labelled morphological lexicon Sloleks 1.2, version 1.1](http://hdl.handle.net/11356/1186)
4. [Basilex Lexicon](https://ivdnt.org/downloads/taalmaterialen/tstc-basilex-lexicon)
5. [Basiscript Lexicon](https://ivdnt.org/downloads/taalmaterialen/tstc-basiscript-lexicon)
6. [BioLexicon](https://portulanclarin.net/repository/browse/biolexicon/389ebe92d63711e1a404080027e73ea2a4b6c6e4ac464aa8976d233a6f26790b/)
7. [CombiLex](https://ivdnt.org/downloads/taalmaterialen/tstc-combilex)
8. [Concreteness and imageability lexicon MEGA.HR-Crossling](http://hdl.handle.net/11356/1187)
9. [Démonette](https://hdl.handle.net/11403/demonette/v2)
10. [Dicovalence](https://hdl.handle.net/11403/dicovalence/v1)
11. [Dutch Electronic Lexicon of Multiword Expressions](http://duelme.clarin.inl.nl/)
12. [Extended dictionary of named entities NELexicon connected with Linked Open Data](http://hdl.handle.net/11321/529)
13. [ILSP PsychoLinguistic Resource](http://hdl.grnet.gr/11500/ATHENA-0000-0000-23EC-F)
14. [Italian Content Words v3](http://hdl.handle.net/11372/LRT-2894)
15. [Italian Function Words v3](http://hdl.handle.net/11372/LRT-2893)
16. [LEX-MWE-PT: Word Combination in Portuguese Language](http://metashare.elda.org/repository/browse/lex-mwe-pt-word-combination-in-portuguese-language/8c13600ccd0711e1a404080027e73ea2f9cfd28f51d5437b8f5827c516c348fe/)
17. [LX-Rare Word Similarity Dataset](http://metashare.elda.org/repository/browse/lx-rare-word-similarity-dataset/f8dd0332e6d911e6a2aa782bcb074135a226cf379cf746a8976dd3420f5a2813/)
18. [LX-SimLex-999](http://metashare.elda.org/repository/browse/lx-simlex-999/4ab1ea58e6d311e6a2aa782bcb0741351e920e18429e4d3e9d229a58030812fe/)
19. [LX-WordSim-353](http://metashare.elda.org/repository/browse/lx-wordsim-353/c4e08b72e6dd11e6a2aa782bcb074135a5ac38ba70a14fb3adbd5782b21dacb0/)
20. [MarsaLex](https://hdl.handle.net/11403/sldr000850/v1)
21. [MorfFlex CZ](http://hdl.handle.net/11234/1-1673)
22. [Morphalou](https://hdl.handle.net/11403/morphalou/v3.1)
23. [MULTEXT-East free lexicons 4.0](http://hdl.handle.net/11356/1041)
24. [MULTEXT-East non-commercial lexicons 4.0](http://hdl.handle.net/11356/1042)
25. [Multifunctional Computational Lexicon of Contemporary Portuguese](http://metashare.elda.org/repository/browse/multifunctional-computational-lexicon-of-contemporary-portuguese/23452ca8a35a11e1a404080027e73ea2d0884e62ca8f466fa0f6be6e1e397cd4/)
26. [MWELexicon 1.1](https://clarin-pl.eu/dspace/handle/11321/508)
27. [OMBI Arabic-Dutch](https://ivdnt.org/downloads/taalmaterialen/tstc-ombi-arab-ned)
28. [OMBI Dutch-Arabic](https://ivdnt.org/downloads/taalmaterialen/tstc-ombi-ned-arab)
29. [OMBI Dutch-Danish](https://ivdnt.org/downloads/taalmaterialen/tstc-ombi-ned-deens)
30. [OMBI Dutch-Indonesian](https://ivdnt.org/downloads/taalmaterialen/tstc-ombi-ned-ind)
31. [OpeNER Sentiment Lexicon Italian - LMF](https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/ILC-73)
32. [ParaDi 2.0](http://hdl.handle.net/11234/1-2605)
33. [PAROLE Lexicon](https://ivdnt.org/downloads/taalmaterialen/tstc-parole-lexicon)
34. [QTLeap specialized lexicons](https://portulanclarin.net/repository/browse/qtleap-specialized-lexicons/9a410010830e11e5a2aa782bcb0741357950092357fd4b7c97f39c12ea6a81cc/)
35. [Reference Lexicon for Belgian-Dutch (RBBN)](https://ivdnt.org/downloads/taalmaterialen/tstc-referentiebestand-belgisch-nederlands)
36. [Reference Lexicon for Dutch](https://ivdnt.org/downloads/taalmaterialen/tstc-referentiebestand-nederlands)
37. [Slovene sentiment lexicon JOB 1.0](http://hdl.handle.net/11356/1112)
38. [Slovene sentiment lexicon KSS 1.1](http://hdl.handle.net/11356/1097)
39. [STO morphology (v2) - LMF format](http://hdl.handle.net/20.500.12115/22)
40. [STO syntax (v2) - LMF format](http://hdl.handle.net/20.500.12115/23)
41. [VfrLPL](https://hdl.handle.net/11403/sldr000533/v2)
42. [Walenty (2018-06-29)](https://clarin-pl.eu/dspace/handle/11321/586)
43. [Word embeddings CLARIN.SI-embed.hr 1.0](http://hdl.handle.net/11356/1205)
44. [Word embeddings CLARIN.SI-embed.sl 1.0](http://hdl.handle.net/11356/1204)
45. [Word embeddings CLARIN.SI-embed.sr 1.0](http://hdl.handle.net/11356/1206)

The following 7 lexica are unavailable for download or online browsing:

1. [LX-Abbreviations](http://metashare.elda.org/repository/browse/lx-abbreviations/27b9084aa35a11e1a404080027e73ea2cbbed158889c419d80bae3f160cba3e5/" \t "_parent)
2. [LX-DSemVectors](https://portulanclarin.net/repository/browse/lx-dsemvectors/b325bb8495e611e6a2aa782bcb0741356b5403f31a544dfa8a7a114182a14964/)
3. [LX-StopWords](http://metashare.elda.org/repository/browse/lx-stopwords/29892e16a35a11e1a404080027e73ea22e53349e39f348a7944b0b5bef6e9c41/)
4. [Maltese Speech Engine Lexicon](https://portulanclarin.net/repository/browse/maltese-speech-engine-lexicon/03dac3a2615d11e2a2aa782bcb074135e14195826bfd40c6a8d67b1448380342/)
5. [PAROLE Portuguese Lexicon](https://portulanclarin.net/repository/browse/parole-portuguese-lexicon/f9be5b86cf3011e1a404080027e73ea246e34a4274c54c068f14a2d622f4e4d4/)
6. [Porlex](https://portulanclarin.net/repository/browse/porlex/a7ea7a8e2cb811e2a2aa782bcb074135ffcac606a29341d581999e6ff55ed5b8/)
7. [Simple Portuguese Lexicon](http://metashare.elda.org/repository/browse/simple-portuguese-lexicon/fafce15ccf3011e1a404080027e73ea2bbc3f718cfbf452989472b8f2bd5f392/)

### Language

64 lexica are monolingual and cover the following 17 languages:

1. Arabic (2 resources)
2. Croatian (2 resource)
3. Czech (5 resources)
4. Danish (2 resources)
5. Dutch (7 resources)
6. English (3 resources)
7. Estonian (1 resource)
8. French (5 resources)
9. Icelandic (1 resource)
10. Italian (4 resources)
11. Greek (1 resource)
12. Maltese (1 resource)
13. Polish (4 resources)
14. Portuguese (11 resources)
15. Serbian (2 resources)
16. Slovenian (9 resources)
17. Swedish (4 resources)

12 lexica are multilingual and cover the following 9 language combinations:

1. Afrikaans, Albanian, Arabic, Armenian, Azerbaijani, Basque, Belarusian, Bengali, Bosnian, Bulgarian, Burmese, Catalan, Cebuano, Central Khmer, Chinese, Croatian, Czech, Danish, Dutch, English, Esperanto, Estonian, Finnish, French, Galician, Georgian, German, Gujarati, Hebrew, Hindi, Hungarian, Icelandic, Indonesian, Italian, Japanese, Kannada, Kazakh, Kirghiz, Korean, Latin, Latvian, Lithuanian, Luxembourgish, Macedonian, Malagasy, Malay (macrolanguage), Malayalam, Marathi, Modern Greek (1453-), Mongolian, Nepali, Norwegian, Panjabi, Persian, Polish, Portuguese, Romanian, Russian, Serbian, Serbo-Croatian, Sinhala, Slovak, Slovenian, Spanish, Swedish, Tagalog, Tajik, Tamil, Telugu, Thai, Turkish, Ukrainian, Urdu, Uzbek, Vietnamese, Welsh, Western Frisian (1 resource)
2. Bulgarian, Czech, English, Estonian, French, Hungarian, Romanian, Slovak, Slovenian, Ukrainian (1 resource)
3. Czech, English (2 resources)
4. Dutch, Arabic (2 resources)
5. Dutch, Danish (1 resource)
6. Dutch, Indonesian (1 resource)
7. English, Spanish, Castilian, Bulgarian, Basque, Dutch, Flemish, Czech, Portuguese (1 resource)
8. Macedonian, Persian, Polish, Russian, Serbian (1 resource)
9. Persian-English (2 resources)

### Size

Size of the lexica varies considerably; the largest lexicon in terms of the number of entries is the collection of [MULTEXT-East free lexicons 4.0](http://hdl.handle.net/11356/1041) (3.6 million entries), while the smallest is [Emoji Sentiment Ranking 1.0](http://hdl.handle.net/11356/1048) (751 entries).

The following 2 lexica lack information on size:

1. [Basilex Lexicon](https://ivdnt.org/downloads/taalmaterialen/tstc-basilex-lexicon)

1. [Basiscript Lexicon](https://ivdnt.org/downloads/taalmaterialen/tstc-basiscript-lexicon" \t "_parent)

### Linguistic information

The most types of linguistic information included in the lexica are the following:

1. Morphosyntactic information (27 lexica)
2. Lemmas (18 lexica)
3. Sentiment (3 lexica)

The following 29 resources do not provide explicit information on the type of included linguistic information in the resource description field, though for some this can be inferred from the name of the resource:

1. [Basilex Lexicon](https://ivdnt.org/downloads/taalmaterialen/tstc-basilex-lexicon)
2. [Basiscript Lexicon](https://ivdnt.org/downloads/taalmaterialen/tstc-basiscript-lexicon)
3. [BioLexicon](https://portulanclarin.net/repository/browse/biolexicon/389ebe92d63711e1a404080027e73ea2a4b6c6e4ac464aa8976d233a6f26790b/)
4. [CombiLex](https://ivdnt.org/downloads/taalmaterialen/tstc-combilex)
5. [DeriNet 1.6](http://hdl.handle.net/11234/1-2873)
6. [Dutch Electronic Lexicon of Multiword Expressions](http://duelme.clarin.inl.nl/)

1. [Emotional Annotations Dictionary](https://clarin-pl.eu/dspace/handle/11321/553)

1. [Extended dictionary of named entities NELexicon connected with Linked Open Data](http://hdl.handle.net/11321/529)
2. [Italian Content Words v3](http://hdl.handle.net/11372/LRT-2894" \t "_parent)
3. [Italian Function Words v3](http://hdl.handle.net/11372/LRT-2893)
4. [MarsaLex](https://hdl.handle.net/11403/sldr000850/v1)
5. [Name Component Lexicon](http://urn.fi/urn:nbn:fi:lb-20140730188)
6. [Old Swedish morphology (2017-10-16)](http://hdl.handle.net/10794/25)
7. [Old Swedish morphology (2017-10-16)](http://hdl.handle.net/10794/25)
8. [OMBI Arabic-Dutch](https://ivdnt.org/downloads/taalmaterialen/tstc-ombi-arab-ned)
9. [OMBI Dutch-Arabic](https://ivdnt.org/downloads/taalmaterialen/tstc-ombi-ned-arab)
10. [OMBI Dutch-Danish](https://ivdnt.org/downloads/taalmaterialen/tstc-ombi-ned-deens)
11. [OMBI Dutch-Indonesian](https://ivdnt.org/downloads/taalmaterialen/tstc-ombi-ned-ind)
12. [Parole+ (2017-10-16)](http://hdl.handle.net/10794/33)
13. [PAROLE-SIMPLE-CLIPS](https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/ILC-88)
14. [QTLeap specialized lexicons](https://portulanclarin.net/repository/browse/qtleap-specialized-lexicons/9a410010830e11e5a2aa782bcb0741357950092357fd4b7c97f39c12ea6a81cc/)
15. [Reference Lexicon for Belgian-Dutch (RBBN)](https://ivdnt.org/downloads/taalmaterialen/tstc-referentiebestand-belgisch-nederlands)
16. [SALDO's morphology (2017-10-16](http://hdl.handle.net/10794/36))
17. [Simple lexicon](http://hdl.handle.net/10794/38)
18. [Simple Portuguese Lexicon](http://metashare.elda.org/repository/browse/simple-portuguese-lexicon/fafce15ccf3011e1a404080027e73ea2bbc3f718cfbf452989472b8f2bd5f392/)
19. STO morphology (v2) - LMF format
20. [STO syntax (v2) - LMF format](http://hdl.handle.net/20.500.12115/23)
21. [The Database of Estonian Multi-Word Expressions](https://metashare.ut.ee/repository/browse/the-database-of-estonian-multi-word-expressions/4d8252e8463411e2a6e4005056b400243ed5ec91ec5044bbb0e85b2ce16f472b/)
22. [Walenty (2018-06-29)](https://clarin-pl.eu/dspace/handle/11321/586)

### Licence

Information on licence is available for all lexica except the following:

1. [VfrLPL](https://hdl.handle.net/11403/sldr000533/v2)

The most common licence types are as follows:

1. CC-BY (46 lexica)
2. MS-NC (8 lexica)

### Overview of the lexica in the CLARIN infrastructure

Table 1: CLARIN Lexica and their characteristics, sorted alphabetically by language

|  |  |  |
| --- | --- | --- |
| **Resource** | **Language** | **Description** |
| [Concreteness and imageability lexicon MEGA.HR-Crossling](http://hdl.handle.net/11356/1187" \t "_parent)  **Size:** 7,237,589 entries  **Linguistic information:** concreteness prediction, imageability prediction  **Licence:** CC-BY-SA 4.0 | 77 languages | These lexica contain concreteness and imageability predictions for 77 languages. They are available for download from CLARIN.SI. |
| [Emoji Sentiment Ranking 1.0](http://hdl.handle.net/11356/1048)  **Size:** 751 entries (emojis)  **Linguistic information:** sentiment labels  **Licence:** CC-BY-SA 4.0 | Albanian, Bulgarian, English, German, Hungarian, Polish, Portuguese, Russian, Serbo-Croatian, Slovak, Slovenian, Spanish, Swedish | This is a lexicon of emojis available for download from CLARIN.SI and for online browsing through a dedicated environment. |
| [A machine-readable dictionary of Egyptian Arabic](http://hdl.handle.net/11022/0000-0007-C266-B)  **Size:** 2,418 entries  **Licence:** CC-BY-NC-SA 3.0  **Linguistic information:** basic morphological information, usage examples | Arabic (Egyptian) | This lexicon presents a more comprehensive version of [A machine-readable glossary of Egyptian Arabic](http://hdl.handle.net/11022/0000-0000-001D-0). The resource is available for download from ARCHE. |
| [A machine-readable glossary of Egyptian Arabic](http://hdl.handle.net/11022/0000-0000-001D-0)  **Size:** 2,204 entries  **Linguistic information:** basic morphological information, usage examples  **Licence:** CC-BY-NC-SA 3.0 | Arabic (Egyptian) | This lexicon has been compiled for comparative as well as didactic purposes in the on-going [VICAV](https://vicav.acdh.oeaw.ac.at/) project. The resource is available for download from ARCHE. |
| [OMBI Dutch-Arabic](https://ivdnt.org/downloads/taalmaterialen/tstc-ombi-ned-arab)  **Size:** 37,000 entries  **Licence:** other | Arabic, Dutch | This is a bilingual lexicon that is suitable for language technology applications such as automatic translation, e-learning, multilingual information retrieval, etc. The resource is available for download from the Dutch Language Institute (INT). |
| [MULTEXT-East free lexicons 4.0](http://hdl.handle.net/11356/1041)  **Size:** 3,665,864 entries **Linguistic information:** MSD-tags, lemmas **Licence:** CC-BY-SA 4.0 | Bulgarian, Czech, English, Estonian, French, Hungarian, Romanian, Slovak, Slovenian, Ukrainian | These are morphological lexica available for download from the CLARIN.SI repository. |
| [Automatically constructed multiword lexicon hrMWELex v0.5](http://hdl.handle.net/11356/1177)  **Size:** 43,730 entries  **Linguistic information:** MWEs  **Licence:** CC-BY 4.0 | Croatian | This is a lexicon of multiword expressions available for download from CLARIN.SI. |
| [Word embeddings CLARIN.SI-embed.hr 1.0](http://hdl.handle.net/11356/1205)  **Size:** 3,147,352 entries  **Linguistic information:** PoS-tags, lemmas  **Licence:** CC-BY 4.0 | Croatian | This lexicon contains word embeddings extracted from the Croatian web corpus [hrWaC](http://hdl.handle.net/11356/1064) and a 400-million-token-heavy collection of newspaper texts. The resource is available for download from CLARIN.SI. |
| [DeriNet 1.6](http://hdl.handle.net/11234/1-2873)  **Size:** 1,027,832 entries  **Licence:** CC-BY-NC-SA 3.0 | Czech | This is a lexicon of derivational relations (both compounding and inflections). The resource is available for download and online browsing through LINDAT. |
| [MorfFlex CZ](http://hdl.handle.net/11234/1-1673)  **Size:** 124,259,099 lexical types  **Linguistic information:** MSD-tags, derivational, semantic, NER information  **Licence:** CC-BY-NC-SA 3.0 | Czech | This is a morphological lexicon available for download from LINDAT. |
| [ParaDi 2.0](http://hdl.handle.net/11234/1-2605)  **Size:** 1,621 entries  **Linguistic information:** MSD-tags, syntactic/semantic features  **Licence:** CC-BY 4.0 | Czech | This is a lexicon of single-word paraphrases of Czech verbal multiword expressions. The resource is available for download from LINDAT. |
| [PDT-Vallex](http://hdl.handle.net/11858/00-097C-0000-0023-4338-F)  **Size:** 7,121 entries, 11,933 frames  **Linguistic information:** verb, adjective and noun valency  **Licence:** CC-BY-NC-SA 4.0 | Czech | This is a valency lexicon linked to several Czech corpora (PDT, PCEDT Cz side, PDTSC, Faust). The resource is available for download and online browsing through LINDAT. |
| [VALLEX 3.0](http://hdl.handle.net/11234/1-2307)  **Size:** 2,722 entries, 6,711 units, 6,711 frames, 4,586 words  **Linguistic information:** verb senses (characterized by glosses and examples)  **Licence:** CC-BY-NC-SA 4.0 | Czech | This is a valency lexicon available for download and online browsing through LINDAT. |
| [CzEngClass 0.2](http://hdl.handle.net/11234/1-2824" \t "_parent)  **Size:** 200 classes, 3,525 entries  **Linguistic information:** valency and synonymy  **Licence:** CC-BY-NC-SA 4.0 | Czech, English | This is a valency lexicon linked to  PDT-Vallex, EngVallex and external resources, such as FrameNet, VerbNet, WordNet, etc. The resource is available for download and online browsing through LINDAT. |
| [CzEngVallex](http://hdl.handle.net/11234/1-1512)  **Size:** 20,835 pairs (verb senses)  **Linguistic information:** verb valency  **Licence:** CC-BY-NC-SA 4.0 | Czech, English | This is a valency lexicon linked to the parallel PCEDT corpus. The resource is available for download and online browsing through LINDAT. |
| [STO morphology (v2) - LMF format](http://hdl.handle.net/20.500.12115/22" \t "_parent)  **Size:** 87,209 entries  **Licence:** CC BY-SA 4.0 | Danish | This morphological lexicon is available for download from the CLARIN-DK repository. It is also available in the .[csv](http://hdl.handle.net/20.500.12115/21) format. |
| [STO syntax (v2) - LMF format](http://hdl.handle.net/20.500.12115/23)  **Size:** 84,159 entries  **Licence:** CC BY-SA 4.0 | Danish | This syntactic lexicon is available for download from the CLARIN-DK repository. |
| [Basilex Lexicon](https://ivdnt.org/downloads/taalmaterialen/tstc-basilex-lexicon) | Dutch | This is a lexicon that comprises all lemmas from the [Basilex Corpus](https://ivdnt.org/downloads/tstc-basilex-corpus). The Basilex Corpus is an annotated collection of texts written for children in elementary school. The resource is available for download from the Dutch Language Institute (INT). |
| [Basiscript Lexicon](https://ivdnt.org/downloads/taalmaterialen/tstc-basiscript-lexicon)  **Licence:** other | Dutch | This is a lexicon that comprises all lemmas from the [Basiscript Corpus](https://ivdnt.org/downloads/tstc-basiscript-corpus). The Basiscript Corpus is an annotated collection of texts written by children in elementary school. The resource is available for download from the Dutch Language Institute (INT). |
| [CombiLex](https://ivdnt.org/downloads/taalmaterialen/tstc-combilex)  **Size:** 213,000 lemmas  **Linguistic information:** lemmas and word forms  **Licence:** other | Dutch | This is a lexicon of words and word forms available for download from the Dutch Language Institute (INT). |
| [Dutch Electronic Lexicon of Multiword Expressions](http://duelme.clarin.inl.nl/)  **Size:** 5,000 expressions  **Licence:** other | Dutch | This is a lexicon of multiword expressions available for download from the Dutch Language Institute (INT). |
| [PAROLE Lexicon](https://ivdnt.org/downloads/taalmaterialen/tstc-parole-lexicon)  **Size:** 20,000 entries  **Linguistic information:** MSD-tags and syntactic complementation patterns  **Licence:** other | Dutch | This morphosyntactic lexicon is available for download from the Dutch Language Institute (INT). |
| [Reference Lexicon for Belgian-Dutch (RBBN)](https://ivdnt.org/downloads/taalmaterialen/tstc-referentiebestand-belgisch-nederlands)  **Size:** 4,000 words and expressions  **Licence:** other | Dutch | This lexicon, which contains words and expressions typically of Dutch spoken in Belgium, is available for download from the Dutch Language Institute (INT). |
| [Reference Lexicon for Dutch](https://ivdnt.org/downloads/taalmaterialen/tstc-referentiebestand-nederlands)  **Size:** 50,000 lemmas  **Linguistic information:** dialectical information  **Licence:** other | Dutch | This is a corpus-based monolingual lexicon available for download the Dutch Language Institute (INT). |
| [OMBI Arabic-Dutch](https://ivdnt.org/downloads/taalmaterialen/tstc-ombi-arab-ned)  **Size:** 37,000 entries  **Licence:** other | Dutch, Arabic | This is a bilingual lexicon for language technology applications such as automatic translation, e-learning, multilingual information retrieval, etc. The resource is available for download from the Dutch Language Institute (INT). |
| [OMBI Dutch-Danish](https://ivdnt.org/downloads/taalmaterialen/tstc-ombi-ned-deens)  **Size:** 46,000 entries  **Licence:** other | Dutch, Danish | This is a bilingual lexicon for language technology applications such as automatic translation, e-learning, multilingual information retrieval, etc. The resource is available for download from the Dutch Language Institute (INT). |
| [OMBI Dutch-Indonesian](https://ivdnt.org/downloads/taalmaterialen/tstc-ombi-ned-ind)  **Size:** 50,000 entries  **Licence:** other | Dutch, Indonesian | This is a bilingual lexicon for language technology applications such as automatic translation, e-learning, multilingual information retrieval, etc. The resource is available for download from the Dutch Language Institute (INT). |
| [BioLexicon](https://portulanclarin.net/repository/browse/biolexicon/389ebe92d63711e1a404080027e73ea2a4b6c6e4ac464aa8976d233a6f26790b/)  **Size:** over 2.2 million entries (over 3.3 million semantic relations)  **Licence:** ELRA END USER | English | This is a large-scale, wide-coverage computational lexicon covering the biomedical domain. The resource is unavailable for download or online browsing, but can be accessed by contacting the resource manager. |
| [EngVallex](http://hdl.handle.net/11858/00-097C-0000-0023-4337-2)  **Size:** 4,337 entries, 7,148 frames  **Linguistic information:** verb valency  **Licence:** CC-BY-NC-SA 4.0 | English | This is a valency lexicon linked to the English side of the PCEDT corpus (WSJ corpus). The resource is available for download from LINDAT and for online browsing. |
| [QTLeap specialized lexicons](https://portulanclarin.net/repository/browse/qtleap-specialized-lexicons/9a410010830e11e5a2aa782bcb0741357950092357fd4b7c97f39c12ea6a81cc/)  **Size:** 231,516 entries  **Licence:** CC-BY | English, Spanish, Castilian, Bulgarian, Basque, Dutch, Flemish, Czech, Portuguese | This lexicon is used for the automatic translation of specific IT domain expressions and is available for download from CLARIN PORTULAN. |
| [The Database of Estonian Multi-Word Expressions](https://metashare.ut.ee/repository/browse/the-database-of-estonian-multi-word-expressions/4d8252e8463411e2a6e4005056b400243ed5ec91ec5044bbb0e85b2ce16f472b/)  **Size:** 12,500 words  **Licence:** proprietary | Estonian | This is a collection of lexica that contain multi-word expressions consisting of a verb and a particle or a verb and its complements. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [Démonette](https://hdl.handle.net/11403/demonette/v2)  **Size:** 96,027 entries  **Linguistic information:** MSD-tags (grace format), semantic types  **Licence:** CC-BY 4.0 | French | This is a morphological lexicon available for download from ORTOLANG. |
| [Dicovalence](https://hdl.handle.net/11403/dicovalence/v1)  **Size:** 8,000 entries  **Linguistic information:** c- and s-selectional restrictions,  **Licence:** Licence Publique Générale Amoindrie GNU | French | This is a verb-valency lexicon.  The lexicon specifies certain selectional restrictions, possible term manifestations (pronominal, phrasal), and whether the valency frames can be used in various passive constructions, as well as references to other valency frames for the same infinitive. The resource is available for download from ORTOLANG. |
| [MarsaLex](https://hdl.handle.net/11403/sldr000850/v1)  **Size:** 595,000,000 inflected forms  **Licence:** CC-BY 4.0 | French | This is a morphological lexicon available for download from ORTOLANG. |
| [Morphalou](https://hdl.handle.net/11403/morphalou/v3.1)  **Size:** 159,261 entries  **Linguistic information:** spelling, phonetics, mood, tense, MSD-tags, spelling variant, feminine variation, pronominal  **Licence:** Publique Générale Amoindrie GNU | French | This is a morphological lexicon available for download from ORTOLANG. |
| [VfrLPL](https://hdl.handle.net/11403/sldr000533/v2)  **Size:** 8,800 entries  **Linguistic information:** conjugation forms, phonetic forms, use frequencies | French | This is a morphosyntactic lexicon available for download from ORTOLANG. |
| [ILSP PsychoLinguistic Resource](http://hdl.grnet.gr/11500/ATHENA-0000-0000-23EC-F)  **Size:** 217,664 entries  **Linguistic information:** phonetic transcription, frequency of usage  **Licence**: CC-BY-NC-SA | Greek | This is a lexicon for psycholinguistic research. The resource is available for download from clarin:el. |
| [Database of Modern Icelandic Inflections](http://clarin.is/en/resources/dmii/)  **Size:** 278,994 entries  **Linguistic information:** MSD-tags  **Licence:** other | Icelandic | This is a morphological lexicon available for download and online browsing through CLARIN-IS. |
| [Italian Content Words v3](http://hdl.handle.net/11372/LRT-2894)  **Size:** 2,342,120 items  **Licence:** CC-BY-NC-SA 4.0 | Italian | This is a morphological lexicon. The resource is available for download from LINDAT. |
| [Italian Function Words v3](http://hdl.handle.net/11372/LRT-2893)  **Size:** 3,510 entries  **Licence:** CC-BY-NC-SA 4.0 | Italian | This is a morphological lexicon. The resource is available for download from LINDAT. |
| [OpeNER Sentiment Lexicon Italian - LMF](https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/ILC-73)  **Size:** 24,293 entries  **Linguistic information:** positive/negative/neutral polarity.  **Licence:** CC-BY 4.0 | Italian | This is a sentiment lexicon available for download from ILC4CLARIN. |
| [PAROLE-SIMPLE-CLIPS](https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/ILC-88)  **Size:** 37,406 syntactic units  **Licence:** CC-BY-SA 4.0 | Italian | This is a morphological lexicon available for download from LC4CLARIN. |
| [MULTEXT-East non-commercial lexicons 4.0](http://hdl.handle.net/11356/1042)  **Size:** 2,288,228 entries **Linguistic information:** MSD-tags, lemmas **Licence:** CC-BY-NC 4.0 | Macedonian, Persian, Polish, Russian, Serbian | These are morphological lexica available for download from the CLARIN.SI repository. |
| [Maltese Speech Engine Lexicon](https://portulanclarin.net/repository/browse/maltese-speech-engine-lexicon/03dac3a2615d11e2a2aa782bcb074135e14195826bfd40c6a8d67b1448380342/)  **Size:** 39,242 entries  **Linguistic information:** PoS-tags, orthographic transcription, phonetic forms, syllables, stress position  **Licence:** MS-BY-NC-SA | Maltese | This is a speech lexicon that is useful for building speech-to-text systems. It is available for download from CLARIN PORTULAN. |
| [A machine-readable Persian-English dictionary](http://hdl.handle.net/11022/0000-0000-0029-2)  **Size:** 1,892 entries  **Licence:** CC-BY-NC-SA 3.0  **Linguistic information:** morphological information, usage examples | Persian-English | This bilingual lexicon has been compiled for comparative as well as didactic purposes in the on-going [VICAV](https://vicav.acdh.oeaw.ac.at/) project. The resource is available for download from ARCHE. |
| [A machine-readable Persian-English glossary of verbs](http://hdl.handle.net/11022/0000-0000-0027-4)  **Size:** 429 entries  **Linguistic information:** basic morphological information  **Licence:** CC-BY-NC-SA 3.0 | Persian-English | This lexicon of single-word verbs in Modern Persian is available for download from ARCHE. |
| [Emotional Annotations Dictionary](https://clarin-pl.eu/dspace/handle/11321/553)  **Size:** 178,514 elements  **Licence:** CC-BY 4.0 | Polish | This is a lexicon with emotional annotation extracted from Polish Wordnet. The resource is available for download from the CLARIN-PL repository and online browsing. |
| [Extended dictionary of named entities NELexicon connected with Linked Open Data](http://hdl.handle.net/11321/529)  **Size:** 103,585 entries  **Licence:** GNU LGPL 3.0 | Polish | This lexicon contains Polish named entities connected with terminology from available resources within Linked Open Data (e.g. WordNet, DBPedia, Wikipedia, etc.). The resource is available for download from the CLARIN-PL repository. |
| [MWELexicon 1.1](http://hdl.handle.net/11321/508)  **Size:** 56,500 lexical units  **Linguistic information:** “syntactic behaviour”  **Licence:** plWordNet | Polish | This is a lexicon of multiword expressions available for download from CLARIN.PL. |
| [Walenty (2018-06-29)](https://clarin-pl.eu/dspace/handle/11321/586)  **Size:** 18,236 entries  **Licence:** CC BY SA 4.0 | Polish | This is a lexicon of verb valency that is available for download from the CLARIN-PL repository. |
| [LEX-MWE-PT: Word Combination in Portuguese Language](http://metashare.elda.org/repository/browse/lex-mwe-pt-word-combination-in-portuguese-language/8c13600ccd0711e1a404080027e73ea2f9cfd28f51d5437b8f5827c516c348fe/)  **Size:** 1,198 entries 12,753 multi word unit  **Linguistic information:** lemmas  **Licence:** MS NC-NoReD-ND | Portuguese | This is a lexicon of multiword expressions available for download from CLARIN PORTULAN. |
| [LX-Abbreviations](http://metashare.elda.org/repository/browse/lx-abbreviations/27b9084aa35a11e1a404080027e73ea2cbbed158889c419d80bae3f160cba3e5/)  **Size:** 208 words  **Linguistic information:** MSD-tags  **Licence:** MS NC-NoReD-ND | Portuguese | This is a lexicon of abbreviations. It is unavailable for download or online browsing, but can be accessed by contacting the resource manager. |
| [LX-DSemVectors](https://portulanclarin.net/repository/browse/lx-dsemvectors/b325bb8495e611e6a2aa782bcb0741356b5403f31a544dfa8a7a114182a14964/)  **Size:** 17,572 words  **Linguistic information:** word embeddings  **Licence:** MS NC-NoReD-ND | Portuguese | This lexicon provides distributional semantic representations of Portuguese words and is available for download from CLARIN PORTULAN. |
| [LX-Rare Word Similarity Dataset](http://metashare.elda.org/repository/browse/lx-rare-word-similarity-dataset/f8dd0332e6d911e6a2aa782bcb074135a226cf379cf746a8976dd3420f5a2813/)  **Size:** 2,034 words  **Linguistic information:** synonyms  **Licence:** MS NC-NoReD-ND | Portuguese | This is a word-similarity lexicon that is unavailable for download or online browsing, but can be accessed by contacting the resource manager. |
| [LX-SimLex-999](http://metashare.elda.org/repository/browse/lx-simlex-999/4ab1ea58e6d311e6a2aa782bcb0741351e920e18429e4d3e9d229a58030812fe/)  **Size:** 1,998 words  **Linguistic information:** MSD-tags, linguistic standardness  **Licence:** MS NC-NoReD-ND | Portuguese | This is a word-similarity lexicon that is unavailable for download or online browsing, but can be accessed by contacting the resource manager. |
| [LX-StopWords](http://metashare.elda.org/repository/browse/lx-stopwords/29892e16a35a11e1a404080027e73ea22e53349e39f348a7944b0b5bef6e9c41/)  **Size:** 2,631 words  **Linguistic information:** MSD-tags, MWEs  **Licence:** MS NC-NoReD-ND | Portuguese | This is a manually compiled exhaustive list of closed-class words in Portuguese. The resource is unavailable for download or online browsing, but can be accessed by contacting the resource manager. |
| [LX-WordSim-353](http://metashare.elda.org/repository/browse/lx-wordsim-353/c4e08b72e6dd11e6a2aa782bcb074135a5ac38ba70a14fb3adbd5782b21dacb0/)  **Size:** 706 words  **Linguistic information:** synonyms, antonyms, identical, hypernym-hyponym, sibling terms, meronym-holonym  **Licence:** MS NC-NoReD-ND | Portuguese | This is a word-similarity lexicon that is unavailable for download or online browsing, but can be accessed by contacting the resource manager. |
| [Multifunctional Computational Lexicon of Contemporary Portuguese](http://metashare.elda.org/repository/browse/multifunctional-computational-lexicon-of-contemporary-portuguese/23452ca8a35a11e1a404080027e73ea2d0884e62ca8f466fa0f6be6e1e397cd4/)  **Size:** 26,443 entries  **Linguistic information:** lemmas, MWEs, PoS-tags  **Licence:** CC-BY - SA | Portuguese | This is a frequency lexicon suitable for NLP specific purposes (information extraction, lemmatization, PoS tagging). The resource is available for download from META-SHARE (CLARIN PORTULAN distribution). |
| [PAROLE Portuguese Lexicon](https://portulanclarin.net/repository/browse/parole-portuguese-lexicon/f9be5b86cf3011e1a404080027e73ea246e34a4274c54c068f14a2d622f4e4d4/)  **Size:** 20,000 entries  **Linguistic information:** MSD tags, lemma  **Licence:** ELRA EVALUATION | Portuguese | This is a morphosyntactic lexicon available for download from CLARIN PORTULAN |
| [Porlex](https://portulanclarin.net/repository/browse/porlex/a7ea7a8e2cb811e2a2aa782bcb074135ffcac606a29341d581999e6ff55ed5b8/)  **Size:** 27,374 words  **Linguistic information:** orthographic and phonological/phonetic transcriptions, phonetic, MSD-tags, and frequency information  **Licence:** MS NC-NoReD-ND | Portuguese | This is a lexicon that provides psycholinguistic and cognitive information that is useful to select stimulus materials for experiments and/or training vocabularies. The resource is available for download from CLARIN PORTULAN. |
| [Simple Portuguese Lexicon](http://metashare.elda.org/repository/browse/simple-portuguese-lexicon/fafce15ccf3011e1a404080027e73ea2bbc3f718cfbf452989472b8f2bd5f392/)  **Size:** 10,438 entries  **Linguistic information:** qualia structure, semantic relations (hyponymy, synonymy, etc.) **Licence:** MS-BY-NC-SA | Portuguese | This semantic lexicon is available for download from CLARIN PORTULAN. |
| [Automatically constructed multiword lexicon srMWELex v0.5](http://hdl.handle.net/11356/1178)  **Size:** 22,290 entries  **Linguistic information:** MWEs  **Licence:** CC-BY 4.0 | Serbian | This is a lexicon of multiword expressions available for download from CLARIN.SI. |
| [Word embeddings CLARIN.SI-embed.sr 1.0](http://hdl.handle.net/11356/1206)  **Size:** 1,480,566 entries  **Linguistic information:** PoS-tags, lemmas  **Licence:** CC-BY 4.0 | Serbian | The lexicon contains word embeddings from the [srWaC web corpus](http://hdl.handle.net/11356/1063). The resource is available for download from CLARIN.SI. |
| [Automatically constructed multiword lexicon slMWELex v0.5](http://hdl.handle.net/11356/1179)  **Size:** 47,579 entries  **Linguistic information:** MWEs  **Licence:** CC-BY 4.0 | Slovenian | This is a lexicon of multiword expressions available for download from CLARIN.SI. |
| [Automatically stress labelled morphological lexicon Sloleks 1.2, version 1.1](http://hdl.handle.net/11356/1186)  **Size:** 100,805 entries; 2,774,745 words  **Linguistic information:** wordforms, PoS-tags, lemmas, frequency, prosody  **Licence:** CC-BY-NC-SA 4.0 | Slovenian | This is an extended version of the morphological lexicon [Sloleks 1.2](http://hdl.handle.net/11356/1039) with added information about the stress of each word form. The resource is available for download from CLARIN.SI. |
| [Beseda Corpus Lemmatisation Lexicon](http://hdl.handle.net/11356/1141)  **Size:** 3,228,127 entries  **Linguistic information:** wordforms**,** PoS-tags, lemmas, frequency  **Licence:** CC-BY 4.0 | Slovenian | This lexicon contains inflected open class words from the [Dictionary of Standard Slovenian](http://bos.zrc-sazu.si/sskj_en.html) that are augmented by wordforms, their part of speech tags and their lemmas used during the PoS tagging and lemmatization of the [Beseda corpus](http://bos.zrc-sazu.si/beseda.html). The resource is available for download from CLARIN.SI and for online browsing. |
| [Lexicon of historical Slovene imp25k 1.1](http://hdl.handle.net/11356/1032)  **Size:** 28,034 entries **Linguistic information:** MSD-tags, lemmas, etymological glosses  **Licence:** CC-BY 4.0 | Slovenian | This is a morphological lexicon available for download from CLARIN.SI and for online browsing through a dedicated environment. |
| [Morphological lexicon Sloleks 2.0](http://hdl.handle.net/11356/1230)  **Size:** 100,805 entries  **Linguistic information:** wordforms, PoS-tags, lemmas, frequency, phonology  **Licence:** CC-BY-NC-SA 4.0 | Slovenian | This is a reference morphological lexicon of the Slovenian language developed to be used in NLP applications and language manuals. The resource is available for download from CLARIN.SI and for online browsing. |
| [Slovene sentiment lexicon JOB 1.0](http://hdl.handle.net/11356/1112)  **Size:** 25,524 entries  **Linguistic information:** sentiment tags  **Licence:** CC-BY-S15A 4.0 | Slovenian | This is a lexicon of sentiment labels available for download from the CLARIN.SI repository. |
| [Slovene sentiment lexicon KSS 1.1](http://hdl.handle.net/11356/1097)  **Size**: 90,620 lexica  **Linguistic information:** lemmas, sentiment tags  **Licence:** CC-BY 4.0 | Slovenian | This is a lexicon of sentiment labels available for download from the CLARIN.SI repository. |
| [Word embeddings CLARIN.SI-embed.sl 1.0](http://hdl.handle.net/11356/1204)  **Size:** 4,560,444entries  **Linguistic information:** PoS-tags, lemmas  **Licence:** CC-BY 4.0 | Slovenian | This is a lexicon of word embeddings that is available for download from CLARIN.SI. |
| [Old Swedish morphology (2017-10-16)](http://hdl.handle.net/10794/25)  **Size:** 41,958 entries  **Licence:** CC-BY 4.0 | Swedish | This is a glossary of Old Swedish that is available for download from the SWE-CLARIN repository and can be queried online through KARP. |
| [Parole+ (2017-10-16)](http://hdl.handle.net/10794/33)  **Size:** 24,523 entries  **Licence:** CC-BY 4.0 | Swedish | This is a lexicon for language technologies which offers access to syntactic information and is connected to SALDO senses. The resource can be download from the SWE-CLARIN repository and can be queried online through KARP. |
| [SALDO's morphology (2017-10-16](http://hdl.handle.net/10794/36))  **Size:** 128,036 entries  **Licence:** CC-BY 4.0 | Swedish | This is a semantic and morphological lexicon for language technologies. The resource can be download from the SWE-CLARIN repository and can be queried online through KARP. |
| [Simple lexicon](http://hdl.handle.net/10794/38)  **Size:** 11,624 entries  **Licence:** CC-BY 4.0 | Swedish | This is a semantic lexicon that is available for download from the SWE-CLARIN repository and can be queried online through KARP. |

## Dictionaries

There are 94 dictionaries available in the CLARIN infrastructure. They are presented in detail in Table [2](#Dictionaries).

### VLO

Of the 94 identified dictionaries, the following 14 cannot be found in the VLO:

1. [A digital dictionary of Dagaare](http://hdl.handle.net/11022/0000-0007-C267-A)
2. [A digital dictionary of Damascus Arabic](http://hdl.handle.net/11022/0000-0007-C093-9)
3. [A digital dictionary of Modern Standard Arabic](http://hdl.handle.net/11022/0000-0007-C265-C)
4. [A Digital Dictionary of Tunis Arabic](https://id.acdh.oeaw.ac.at/uuid/175b8cdf-5d04-f4d3-a778-67910aa8fd37)
5. [Dictionary of Modern Latvian (MLVV)](http://tezaurs.lv/mlvv/)
6. [English-Latvian Dictionary](https://www.letonika.lv/groups/default.aspx?g=2&r=10331062&f=1)
7. [Historical Dictionary of Latvian (16th-17th Centuries) (LVVV)](http://tezaurs.lv/lvvv/)
8. [Latvian-Latgalian Dictionary](http://vuordineica.lv/)
9. [Latvian-Russian Dictionary](https://www.letonika.lv/groups/default.aspx?g=2&r=10621049)
10. [Lithuanian-Latvian Dictionary](https://www.letonika.lv/ltlv/)
11. [Lithuanian-Latvian-Latgalian Dictionary (LLL)](http://hipilatlit.ru.lv/dictionary/eng/dictionary.html)
12. [Mühlenbach­ Endzelin Latvian Dictionary (MEV)](http://tezaurs.lv/mev/)
13. [Rendering Dictionary of Personal Names](https://www.letonika.lv/groups/default.aspx?g=2&r=1109)
14. [Russian-Latvian Dictionary](https://www.letonika.lv/groups/default.aspx?g=2&r=10491062&f=1)

### Availability

The following 14 dictionaries are available for download and online browsing:

1. [Comprehensive Estonian-French Dictionary](https://metashare.ut.ee/repository/browse/comprehensive-estonian-french-dictionary/6479410ef4e311e4a6e4005056b4002443080069cb424332b1b68168450efa3f/" \t "_parent)
2. [Dictionary of New Slovenian Words](http://hdl.handle.net/11356/1091)
3. [Dictionary of Slovenian Particles](http://hdl.handle.net/11356/1128)
4. [Dictionary of Slovenian Phrasemes](http://hdl.handle.net/11356/1129)
5. [Dictionary of the Slovenian Language in the Works of Janez Svetokriški](http://hdl.handle.net/11356/1092)
6. [First Estonian Slang Dictionary](https://metashare.ut.ee/repository/browse/first-estonian-slang-dictionary/609a141695f811e6a6e4005056b400244db51ea9f9504fbdb4568bbc615e82cc/)
7. [ISLEX](http://clarin.is/en/resources/islex/)
8. [Slovenian-German Dictionary of Maks Pleteršnik (1894-1895)](https://www.clarin.si/repository/xmlui/handle/11356/1114)
9. [A Digital Dictionary of Tunis Arabic](https://id.acdh.oeaw.ac.at/uuid/175b8cdf-5d04-f4d3-a778-67910aa8fd37)
10. [A digital dictionary of Dagaare](http://hdl.handle.net/11022/0000-0007-C267-A)
11. [Dalin's Dictionary (2017-10-16)](http://hdl.handle.net/10794/16)
12. [Hellquist's Swedish etymology (2017-10-16)](http://hdl.handle.net/10794/26)
13. [Dictionary of Twitterese Janes-Dict 1.0](http://hdl.handle.net/11356/1172)
14. [Japanese-Slovene learner's dictionary jaSlo 3.1](http://hdl.handle.net/11356/1050)

The following 17 dictionaries are available for download only:

1. [A digital dictionary of Damascus Arabic](http://hdl.handle.net/11022/0000-0007-C093-9)
2. [A digital dictionary of Modern Standard Arabic](http://hdl.handle.net/11022/0000-0007-C265-C)
3. [Dictionary of Contemporary Dutch (ANW)](http://anw.inl.nl)
4. [Dictionary of Lesser Used Slovenian Words](http://hdl.handle.net/11356/1130)
5. [English-Greek Dictionary of False Friends](http://hdl.grnet.gr/11500/UOA-0000-0000-25EC-D)
6. [French-Greek lexicon of idioms](http://hdl.grnet.gr/11500/UOA-0000-0000-2574-4)
7. [Jysk Ordbog (a-h), 17. April, 2011, xml-edition](http://hdl.handle.net/20.500.12115/27)
8. [Kven-Norwegian Bokmål dictionary](http://hdl.handle.net/11509/89)
9. [Multilingual dictionaries](http://hdl.grnet.gr/11500/ATHENA-0000-0000-23C7-8)
10. [North Saami-Finnish dictionary](http://hdl.handle.net/11509/97)
11. [Norwegian Bokmål-North Saami dictionary](http://hdl.handle.net/11509/94)
12. [Norwegian Bokmål-South Saami dictionary](http://hdl.handle.net/11509/93)
13. [Norwegian-Vietnamese Dictionary](http://hdl.handle.net/11509/22)
14. [Pronunciation Dictionary](http://clarin.is/en/resources/prondict/)
15. [Reverse dictionary of Slovenian language](http://hdl.handle.net/11356/1094)
16. [Swedish-South Saami dictionary](http://hdl.handle.net/11509/99)
17. [Words of the 16th-Century Slovenian Literary Language](http://hdl.handle.net/11356/1127)

The following 61 dictionaries are available only for online browsing:

1. [Dictionaries of the CLARIN-PL Lexical Platform](http://lexp.clarin-pl.eu/)
2. [Dictionary of Contemporary Finnish](http://urn.fi/urn:nbn:fi:lb-201411261)
3. [Dictionary of Finnish Dialects](http://urn.fi/urn:nbn:fi:lb-201411263)
4. [Dictionary of Middle Dutch (MNW)](http://gtb.ivdnt.org/search/?owner=mnw)
5. [Dictionary of Modern Latvian (MLVV)](http://tezaurs.lv/mlvv/)
6. [Dictionary of Old Dutch (ONW)](http://gtb.ivdnt.org/search/?owner=onw)
7. [Dictionary of Old Literary Finnish](http://urn.fi/urn:nbn:fi:lb-201411262)
8. [Dictionary of Standard Latvian Language (LLVV)](http://tezaurs.lv/llvv/)
9. [Digital Dictionary of the German Language (DWDS)](https://www.dwds.de/)
10. [Early Middle Dutch Dictionary (VMNW)](http://gtb.ivdnt.org/search/?owner=vmnw)
11. [English Hask Collocation Database](http://pelcra.pl/hask_en/)
12. [English-Estonian Machine Translation Dictionary](https://metashare.ut.ee/repository/browse/english-estonian-machine-translation-dictionary/442c5658463411e2a6e4005056b40024b1c86daf75e34c59aa8940bdc7754046/)
13. [English-Latvian Dictionary](https://www.letonika.lv/groups/default.aspx?g=2&r=10331062&f=1)
14. [Estonian-Latvian Dictionary](https://www.letonika.lv/eelv/)
15. [Estonian-Russian Learner's Orthographic Dictionary](https://metashare.ut.ee/repository/browse/estonian-russian-learners-orthographic-dictionary/404962cc9f4c11e6a6e4005056b40024640fb7652c0a4bce9667a4c46e1021ba/)
16. [Etymological Database of the Saami Languages](http://urn.fi/urn:nbn:fi:lb-20140730111)
17. [Finnish-Estonian Dictionary](https://metashare.ut.ee/repository/browse/finnish-estonian-dictionary/8899502cd5c011e7a6e4005056b40024719f97f36908495cbad4fd0795c63005/)
18. [Historical Dictionary of Latvian (16th-17th Centuries) (LVVV)](http://tezaurs.lv/lvvv/)
19. [iWNT](http://gtb.ivdnt.org/search/?owner=wnt)
20. [Karelian Dictionary](http://urn.fi/urn:nbn:fi:lb-20140730157)
21. [Latvian-English Dictionary](https://www.letonika.lv/groups/default.aspx?g=2&r=10621033&f=1)
22. [Latvian-Latgalian Dictionary](http://vuordineica.lv/)
23. [Latvian-Lithuanian Dictionary](https://www.letonika.lv/lvlt/)
24. [Latvian-Russian Dictionary](https://www.letonika.lv/groups/default.aspx?g=2&r=10621049)
25. [Lithuanian-Latvian Dictionary](https://www.letonika.lv/ltlv/)
26. [Lithuanian-Latvian-Latgalian Dictionary (LLL)](http://hipilatlit.ru.lv/dictionary/eng/dictionary.html)
27. [Livonian-Estonian-Latvian Dictionary](http://www.livones.net/lingua/en)
28. [Low-German Loanwords in the Estonian Language](https://metashare.ut.ee/repository/browse/low-german-loanwords-in-the-estonian-language/6648d7aa95f611e6a6e4005056b400247e0a300f7081444296b3c05daf92eed5/)
29. [Mühlenbach­ Endzelin Latvian Dictionary (MEV)](http://tezaurs.lv/mev/)
30. [Online Dictionary of Standard Estonian ÕS 2018](https://metashare.ut.ee/repository/browse/online-dictionary-of-standard-estonian-os-2018/36168830ef1711e8a6e4005056b40024fe3bf119ca574414b7e54174523cb430/)
31. [Online Lexicon of Veps Language](http://urn.fi/urn:nbn:fi:lb-20140730178)
32. [Polish Hask Collocation Database](http://pelcra.pl/hask_pl/)
33. [Rendering Dictionary of Personal Names](https://www.letonika.lv/groups/default.aspx?g=2&r=1109)
34. [Russian-Latvian Dictionary](https://www.letonika.lv/groups/default.aspx?g=2&r=10491062&f=1)
35. [Tezaurs.lv](http://tezaurs.lv/)
36. [The Advisory Dictionary of Document Language](https://metashare.ut.ee/repository/browse/the-advisory-dictionary-of-document-language/0fcc2bf495f411e6a6e4005056b400248b8847a2889f48bf9ce9cc429def135e/)
37. [The Basic Estonian Dictionary](https://metashare.ut.ee/repository/browse/the-basic-estonian-dictionary/d312a56c692a11e6a6e4005056b400242d3e87d443564a978a3646abb4adc977/)
38. [The Dialectological Dictionary of Estonian](https://metashare.ut.ee/repository/browse/the-dialectological-dictionary-of-estonian/d89fb86469eb11e6a6e4005056b40024263595b1e3af473a9c18d246e662575f/)
39. [The Dictionary of Estonian Antonyms](https://metashare.ut.ee/repository/browse/the-dictionary-of-estonian-antonyms/12402e08f57311e4a6e4005056b40024506393c8e07846718038461607252195/)
40. [The Dictionary of Estonian Dialects (2013-...)](https://metashare.ut.ee/repository/browse/the-dictionary-of-estonian-dialects-2013-lopetisnisu/1c5982c4e7b911e6a6e4005056b400240e2540ce1f5941f3b69193da88ffaad4/)
41. [The Dictionary of Estonian Phraseology](https://metashare.ut.ee/repository/browse/the-dictionary-of-estonian-phraseology/69ac9020f3c811e4a6e4005056b4002428255d31ead24b85abdd8c1ee0ef0c10/)
42. [The Dictionary of Estonian Synonyms](https://metashare.ut.ee/repository/browse/the-dictionary-of-estonian-synonyms/27c1646c95f511e6a6e4005056b40024799b94d50fde4807b94aeeb6abb186b3/)
43. [The Dictionary of Estonian Word Families - Index](https://metashare.ut.ee/repository/browse/the-dictionary-of-estonian-word-families-index/04bc1190ed1d11e6a6e4005056b400242473071cf99b4485bda8d88185a908c4/)
44. [The Dictionary of Foreign Words](https://metashare.ut.ee/repository/browse/the-dictionary-of-foreign-words/f8ef8e5495f211e6a6e4005056b40024838d5cfd92004b75ba384f57098aa12b/)
45. [The Dictionary of Old Võru Dialect](https://metashare.ut.ee/repository/browse/the-dictionary-of-old-voru-dialect/dd4f944c355b11e7a6e4005056b400240c86a3dcb1694dcfbd0e7c07e6c9b3b5/)
46. [The Dictionary of the Eastern Dialect](https://metashare.ut.ee/repository/browse/the-dictionary-of-the-eastern-dialect/3294fb089f4911e6a6e4005056b40024d2f5e7ce781a4950b39f5845a3072336/)
47. [The Dictionary of the Hiiu Dialect](https://metashare.ut.ee/repository/browse/the-dictionary-of-the-hiiu-dialect/032fafec9f4a11e6a6e4005056b40024d8e5fa2111d14c6ab7821ba8fd0a83e9/)
48. [The Dictionary of the Kihnu Dialect](https://metashare.ut.ee/repository/browse/the-dictionary-of-the-kihnu-dialect/b759d4c2e39a11e6a6e4005056b400245d8225446e764abcb01b7c09632dcc6b/)
49. [The Dictionary of the Mulgi Dialect](https://metashare.ut.ee/repository/browse/the-dictionary-of-the-mulgi-dialect/1fcab366e7b811e6a6e4005056b40024ce57d9d042734b9c8a670f76479f7f98/)
50. [The Dictionary of the Seto Dialect](https://metashare.ut.ee/repository/browse/the-dictionary-of-the-seto-dialect/9a1640d89f4a11e6a6e4005056b40024edd62c01a31d4f33b0ac5c498baa4f7d/)
51. [The Estonian Etymological Dictionary](https://metashare.ut.ee/repository/browse/the-estonian-etymological-dictionary/0116f124ee6c11e4a6e4005056b4002469250a0f627744cbac4ff5ac88f7547a/)
52. [The Estonian-Latvian-Estonian Dictionary](https://metashare.ut.ee/repository/browse/the-estonian-latvian-dictionary/0413a04e9f4f11e6a6e4005056b400242d5f1aa4a7064ab4b7ad6f7a5f0822a5/)
53. [The Estonian-Mari Dictionary](https://metashare.ut.ee/repository/browse/estonian-mari-dictionary/22c9344e186d11e7a6e4005056b40024ce5bb509f3cb40598e1ab9c3fb5ce4b4/)
54. [The Estonian-Norwegian-Estonian Dictionary](https://metashare.ut.ee/repository/browse/the-estonian-norwegian-dictionary/32815b2e9b6c11e6a6e4005056b40024efdedde1208545beaf897766577bf1d4/)
55. [The Estonian-Russian Dictionary](https://metashare.ut.ee/repository/browse/the-estonian-russian-dictionary/4ba9654c463411e2a6e4005056b40024e786814b938a4085b8b9969dd2ba5e78/)
56. [The Estonian-Udmurt Dictionary](https://metashare.ut.ee/repository/browse/the-estonian-udmurt-dictionary/2588d5329f6d11e6a6e4005056b400249915c93c7ac54369a9e72cc06b65efb0/)
57. [The Explanatory Dictionary of Estonian Language](https://metashare.ut.ee/repository/browse/the-explanatory-dictionary-of-the-estonian-language/8df79018f3cf11e4a6e4005056b400245fba483853d142e0adafb7a8044aecd5/)
58. [The Livonian-German Dictionary](https://metashare.ut.ee/repository/browse/the-livonian-german-dictionary/683f3e0ce7bb11e6a6e4005056b40024dcccb23809bf48b893737339a8263717/)
59. [The Online Dictionary of Estonian Sign Language - Estonian](https://metashare.ut.ee/repository/browse/the-online-dictionary-of-the-estonian-sign-language-estonian/428cf3c6e7ba11e6a6e4005056b400249ab92441a63244e3a3c4799d69b542ee/)
60. [The Russian-Estonian Dictionary](https://metashare.ut.ee/repository/browse/the-russian-estonian-dictionary/9229be7895e911e6a6e4005056b4002455bf19cefdee47c6b9658cdfa54933cc/)
61. [Wurdboek fan de Frsyke taal](http://gtb.ivdnt.org/search/?owner=wft)

The following dictionary is unavailable for download or online browsing, but can seemingly be accessed by contacting the resource manager. Its inaccessibility may be due to outdated metadata (i.e., the link to the landing page of [Basic English-Maltese Dictionary](https://portulanclarin.net/repository/browse/basic-english-maltese-dictionary/13fc5802abc511e1a404080027e73ea2a210be7dd5c44a3b9dd47afb4b2a34ef/) is broken).

1. [Basic English-Maltese Dictionary](https://portulanclarin.net/repository/browse/basic-english-maltese-dictionary/13fc5802abc511e1a404080027e73ea2a210be7dd5c44a3b9dd47afb4b2a34ef/)

### Language

56 dictionaries are monolingual and cover the following 14 languages:

1. Arabic (3 dictionaries)
2. Dagaare (1 dictionary)
3. Danish (1 dictionary)
4. Dutch (5 dictionaries)
5. Estonian (20 dictionaries)
6. English (1 dictionary)
7. Finnish (3 dictionaries)
8. Frisian (1 dictionary)
9. German (1 dictionary)
10. Icelandic (1 dictionary)
11. Latvian (5 dictionaries)
12. Slovenian (9 dictionaries)
13. Swedish (2 dictionaries)
14. Veps (1 dictionary)

38 dictionaries are multilingual and cover the following 31 language combinations:

1. Bulgarian, English, German, Romanian, Greek, and Polish
2. English-Greek
3. Estonian, Estonian Sign Language
4. Estonian, English
5. Estonian, Finnish
6. Estonian, Latvian (2 dictionaries)
7. Estonian, Mari
8. Estonian, Norwegian
9. Estonian, Russian (3 dictionaries)
10. Estonian, Udmurt
11. Greek, English, Bulgarian, Romanian, German, Polish
12. Icelandic, Danish, Norwegian (Bokmal), Norwegian (Nynorsk), Swedish, Faroese
13. Karelian, Finnish
14. Kven Finnish - Norwegian Bokmål
15. Latvian (German)
16. Latvian, English (2 dictionaries)
17. Latvian, Latgalian
18. Latvian, Lithuanian (2 dictionaries)
19. Latvian, Russian (2 dictionaries)
20. Lithuanian, Latvian, Latgalian
21. Livonian, Estonian, Latvian
22. Livonian, German
23. Maltese, English
24. Northern Sami, Finnish
25. Norwegian Bokmal, Northern Sami
26. Norwegian Bokmal, Southern Sami
27. Norwegian, Vietnamese
28. Saami languages, Finnish
29. Slovenian, German
30. Slovenian, Japanese
31. Swedish, Southern Sami

### Size

The following 3 dictionaries lack information on size:

1. [Multilingual dictionaries](http://hdl.grnet.gr/11500/ATHENA-0000-0000-23C7-8)

1. [Online Lexicon of Veps Language](http://urn.fi/urn:nbn:fi:lb-20140730178" \t "_parent)

1. [The Dictionary of Estonian Dialects (2013-...)](https://metashare.ut.ee/repository/browse/the-dictionary-of-estonian-dialects-2013-lopetisnisu/1c5982c4e7b911e6a6e4005056b400240e2540ce1f5941f3b69193da88ffaad4/" \t "_parent)

### Linguistic information

Most (80 out of 94) dictionaries lack explicit information about their linguistic information in the description field. Only the following 14 dictionaries provide this:

1. [A digital dictionary of Dagaare](http://hdl.handle.net/11022/0000-0007-C267-A) (PoS-tags, nominal and verbal valency)
2. [Digital Dictionary of the German Language (DWDS)](https://www.dwds.de/) (morphosyntactic information, pronunciation, etymology, lexical relations, examples of use, typical word connections)
3. [ISLEX](http://clarin.is/en/resources/islex/) (phonology and MSD-tags)

1. [Japanese-Slovene learner's dictionary jaSlo 3.1](http://hdl.handle.net/11356/1050" \t "_parent) (headwords in all Japanese scripts, PoS-tags, difficulty level, Slovene glosses, Slovenian translation equivalents, examples of use)
2. [Multilingual dictionaries](http://hdl.grnet.gr/11500/ATHENA-0000-0000-23C7-8) (MSD-tags, semantic information, multi-word units)

1. [Norwegian-Vietnamese Dictionary](http://hdl.handle.net/11509/22" \t "_parent) (phonetic transcription, PoS-tags, translation and example sentences)
2. [Online Dictionary of Standard Estonian ÕS 2018](https://metashare.ut.ee/repository/browse/online-dictionary-of-standard-estonian-os-2018/36168830ef1711e8a6e4005056b40024fe3bf119ca574414b7e54174523cb430/) (standardised orthography, phonology, MSD-tags)
3. [Pronunciation Dictionary](http://clarin.is/en/resources/prondict/) (IPA and SAMPA phonetic transcription)

1. [The Dialectological Dictionary of Estonian](https://metashare.ut.ee/repository/browse/the-dialectological-dictionary-of-estonian/d89fb86469eb11e6a6e4005056b40024263595b1e3af473a9c18d246e662575f/" \t "_parent) (geographical distribution)
2. [The Dictionary of Estonian Dialects (2013-...)](https://metashare.ut.ee/repository/browse/the-dictionary-of-estonian-dialects-2013-lopetisnisu/1c5982c4e7b911e6a6e4005056b400240e2540ce1f5941f3b69193da88ffaad4/" \t "_parent) (geographical distribution; phonetic/phonological markup)
3. [The Dictionary of Foreign Words](https://metashare.ut.ee/repository/browse/the-dictionary-of-foreign-words/f8ef8e5495f211e6a6e4005056b40024838d5cfd92004b75ba384f57098aa12b/) (orthography, MSD-tags, semantics and word etymology)

1. [The Dictionary of the Seto Dialect](https://metashare.ut.ee/repository/browse/the-dictionary-of-the-seto-dialect/9a1640d89f4a11e6a6e4005056b40024edd62c01a31d4f33b0ac5c498baa4f7d/" \t "_parent) (dialect orthography and simplified spelling)

1. [The Estonian Etymological Dictionary](https://metashare.ut.ee/repository/browse/the-estonian-etymological-dictionary/0116f124ee6c11e4a6e4005056b4002469250a0f627744cbac4ff5ac88f7547a/" \t "_parent) (word etymology)
2. [Words of the 16th-Century Slovenian Literary Language](http://hdl.handle.net/11356/1127) (PoS tags, phonological features)

### Licence

The most common licence types are as follows:

1. CC-BY (40 dictionaries)
2. CLARIN ACA (22 dictionaries)

Information on licence is missing for the following 19 dictionaries:

1. [Dictionary of Contemporary Dutch (ANW)](http://anw.inl.nl" \t "_parent)

1. [Dictionary of Middle Dutch (MNW)](http://gtb.ivdnt.org/search/?owner=mnw" \t "_parent)
2. [Dictionary of Modern Latvian (MLVV)](http://tezaurs.lv/mlvv/" \t "_parent)
3. [Dictionary of Standard Latvian Language (LLVV)](http://tezaurs.lv/llvv/)

1. [Digital Dictionary of the German Language (DWDS)](https://www.dwds.de/)
2. [English-Latvian Dictionary](https://www.letonika.lv/groups/default.aspx?g=2&r=10331062&f=1)
3. [Estonian-Latvian Dictionary](https://www.letonika.lv/eelv/)
4. [Historical Dictionary of Latvian (16th-17th Centuries) (LVVV)](http://tezaurs.lv/lvvv/)
5. [Latvian-English Dictionary](https://www.letonika.lv/groups/default.aspx?g=2&r=10621033&f=1)
6. [Latvian-Latgalian Dictionary](http://vuordineica.lv/)
7. [Latvian-Lithuanian Dictionary](https://www.letonika.lv/lvlt/)
8. [Latvian-Russian Dictionary](https://www.letonika.lv/groups/default.aspx?g=2&r=10621049)
9. [Lithuanian-Latvian Dictionary](https://www.letonika.lv/ltlv/)
10. [Lithuanian-Latvian-Latgalian Dictionary (LLL)](http://hipilatlit.ru.lv/dictionary/eng/dictionary.html)
11. [Livonian-Estonian-Latvian Dictionary](http://www.livones.net/lingua/en)
12. [Mühlenbach­ Endzelin Latvian Dictionary (MEV)](http://tezaurs.lv/mev/)
13. [Rendering Dictionary of Personal Names](https://www.letonika.lv/groups/default.aspx?g=2&r=1109)
14. [Russian-Latvian Dictionary](https://www.letonika.lv/groups/default.aspx?g=2&r=10491062&f=1)
15. [Tezaurs.lv](http://tezaurs.lv/)

### Overview of the dictionaries in the CLARIN infrastructure

Table 2: CLARIN dictionaries and their characteristics, sorted alphabetically by language

|  |  |  |
| --- | --- | --- |
| **Resource** | **Language** | **Description** |
| [A digital dictionary of Damascus Arabic](http://hdl.handle.net/11022/0000-0007-C093-9)  **Size:** 4,218 entries  **Linguistic information:** basic morphological information, usage examples  **Licence:** CC-BY-NC-SA 3.0 | Arabic (Damascus) | This small-sized dictionary of Spoken Damascus Arabic has mainly been compiled for comparative and didactic purposes, and comprises all words in Prochazka's 2-volume textbook *Lehrbuch des Syrisch-Arabischen*. The resource is available for download from ARCHE. |
| [A digital dictionary of Modern Standard Arabic](http://hdl.handle.net/11022/0000-0007-C265-C)  **Size:** 2,418 entries  **Linguistic information:** basic morphological information, a few usage examples  **Licence:** CC-BY-NC-SA 3.0 | Arabic (Modern Standard) | This small-sized dictionary of Modern Standard Arabic has mainly been compiled for comparative and didactic purposes. The resource is available for download from ARCHE. |
| [A Digital Dictionary of Tunis Arabic](https://id.acdh.oeaw.ac.at/uuid/175b8cdf-5d04-f4d3-a778-67910aa8fd37)  **Size:** 7,543 entries  **Licence:** CC-BY-NC-SA 3.0 | Arabic (Tunis) | This dictionary includes entries from the TuniCo corpus and is additionally based on data elicited from complementary interviews with young Tunisians and lexical material taken from published historical sources dating back to the middle of the 20th century. The resource is available for download from ARCHE. |
| [A digital dictionary of Dagaare](http://hdl.handle.net/11022/0000-0007-C267-A)  **Size:** 1252 entries  **Linguistic information:** PoS-tags, nominal and verbal valency  **Licence:** CC-BY-NC-SA 3.0 | Dagaare | This dictionary, which illustrates salient aspects of the Dagaare lexico-grammatical system, is available for download from ARCHE. |
| [Jysk Ordbog (a-h), 17. April, 2011, xml-edition](http://hdl.handle.net/20.500.12115/27)  **Size:** 16,000 entries  **Licence:** CC BY-SA 4.0 | Danish | This is a dictionary of Danish dialects spoken in Jutland between 1700 and 1920. The resource is available for download from CLARIN-DK. |
| [Dictionary of Contemporary Dutch (ANW)](http://anw.inl.nl)  **Size:** 255,239 words | Dutch | This dictionary describes Dutch vocabulary from 1970 onwards. The resource is available for download from the Dutch Language Institute (INT). |
| [Dictionary of Middle Dutch (MNW)](http://gtb.ivdnt.org/search/?owner=mnw)  **Size:** 60,000 lemmas | Dutch | This is a historical dictionary of Medieval Dutch vocabulary from the period between 1250 and 1550. The resource is available for online browsing through the Dutch Language Institute (INT). |
| [Dictionary of Old Dutch (ONW)](http://gtb.ivdnt.org/search/?owner=onw)  **Licence:** other | Dutch | This is a historical dictionary of Old Dutch vocabulary from the period from the period between 500 and 1200. The resource is available for online browsing through the Dutch Language Institute (INT). |
| [Early Middle Dutch Dictionary (VMNW)](http://gtb.ivdnt.org/search/?owner=vmnw)  **Size:** 25,000 lemmas  **Licence:** other | Dutch | This is a historical dictionary of Early Medieval Dutch vocabulary from the period between 1200 and 1300. The resource is available for online browsing through the Dutch Language Institute (INT). |
| [iWNT](http://gtb.ivdnt.org/search/?owner=wnt)  **Size:** 95,000 lemmas  **Licence:** other | Dutch | The dictionary describes the meaning and history of hundred thousands of words in written Dutch, from the period 1500-1976. The resource is available for online browsing through the Dutch Language Institute (INT). |
| [English Hask Collocation Database](http://pelcra.pl/hask_en/)  **Size:** 150,045 entries containing 2,875,407 word combinations  **Licence:** CC-BY-NC | English | HASK collocation dictionaries are phraseological databases used by linguists, language teachers, lexicographers, language materials developers, translators and other language professionals and casual dictionary users. The resource is available for online browsing through the PELCRA environment. |
| [English-Greek Dictionary of False Friends](http://hdl.grnet.gr/11500/UOA-0000-0000-25EC-D)  **Size:** 1,000 entries  **Licence:** CC-BY-NC | English-Greek | This is a dictionary of false friends available for download from clarin:el. |
| [First Estonian Slang Dictionary](https://metashare.ut.ee/repository/browse/first-estonian-slang-dictionary/609a141695f811e6a6e4005056b400244db51ea9f9504fbdb4568bbc615e82cc/)  **Size:** 7,500 keywords  **Licence:** proprietary | Estonian | This is a dictionary of Estonian slang that  is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [Low-German Loanwords in the Estonian Language](https://metashare.ut.ee/repository/browse/low-german-loanwords-in-the-estonian-language/6648d7aa95f611e6a6e4005056b400247e0a300f7081444296b3c05daf92eed5/)  **Size:** 1,150 keywords  **Licence:** CLARIN ACA | Estonian | This is a dictionary of Low-German loanwords. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [Online Dictionary of Standard Estonian ÕS 2018](https://metashare.ut.ee/repository/browse/online-dictionary-of-standard-estonian-os-2018/36168830ef1711e8a6e4005056b40024fe3bf119ca574414b7e54174523cb430/" \t "_parent)  **Size:** 59,275 keywords  **Linguistic information:** standardised orthography, phonology, MSD-tags  **Licence:** CC-BY | Estonian | This is a normative dictionary of modern standard Estonian. The resource is available for download from META-SHARE (CELR distribution). |
| [The Advisory Dictionary of Document Language](https://metashare.ut.ee/repository/browse/the-advisory-dictionary-of-document-language/0fcc2bf495f411e6a6e4005056b400248b8847a2889f48bf9ce9cc429def135e/)  **Size:** 1,200 keywords  **Licence:** CLARIN ACA | Estonian | This is a dictionary for simple, plain and precise expression in official communication. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Basic Estonian Dictionary](https://metashare.ut.ee/repository/browse/the-basic-estonian-dictionary/d312a56c692a11e6a6e4005056b400242d3e87d443564a978a3646abb4adc977/)  **Size:** 50,000 keywords  **Licence:** CC-BY | Estonian | The is a monolingual dictionary that includes important Estonian words explained in simple language. The resource is available for download from META-SHARE (CELR distribution). |
| [The Dialectological Dictionary of Estonian](https://metashare.ut.ee/repository/browse/the-dialectological-dictionary-of-estonian/d89fb86469eb11e6a6e4005056b40024263595b1e3af473a9c18d246e662575f/)  **Size:** 73,397 keywords  **Linguistic information:** geographical distribution  **Licence:** proprietary | Estonian | This is a dictionary that provides a survey of Estonian dialect words and their areal distribution. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Dictionary of Estonian Antonyms](https://metashare.ut.ee/repository/browse/the-dictionary-of-estonian-antonyms/12402e08f57311e4a6e4005056b40024506393c8e07846718038461607252195/)  **Size:** 2,000 antonym pairs  **Licence:** CLARIN ACA | Estonian | The dictionary contains antonym pairs, most of which belong to general language. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Dictionary of Estonian Dialects (2013-...)](https://metashare.ut.ee/repository/browse/the-dictionary-of-estonian-dialects-2013-lopetisnisu/1c5982c4e7b911e6a6e4005056b400240e2540ce1f5941f3b69193da88ffaad4/)  **Linguistic information:** geographical distribution, phonetic/phonological information  **Licence:** CLARIN ACA | Estonian | The is a dialectal dictionary with phonetic and grammatical information, information on the area of distribution, and meaning of dialectal words. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Dictionary of Estonian Phraseology](https://metashare.ut.ee/repository/browse/the-dictionary-of-estonian-phraseology/69ac9020f3c811e4a6e4005056b4002428255d31ead24b85abdd8c1ee0ef0c10/)  **Size:** 7,500 phrasemes  **Licence:** CLARIN ACA | Estonian | This is a dictionary that contains phrasemes from various stylistic registers. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Dictionary of Estonian Synonyms](https://metashare.ut.ee/repository/browse/the-dictionary-of-estonian-synonyms/27c1646c95f511e6a6e4005056b40024799b94d50fde4807b94aeeb6abb186b3/)  **Size:** 10,000 synsets  **Licence:** CLARIN ACA | Estonian | This is a comprehensive dictionary of Estonian synonyms. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Dictionary of Estonian Word Families - Index](https://metashare.ut.ee/repository/browse/the-dictionary-of-estonian-word-families-index/04bc1190ed1d11e6a6e4005056b400242473071cf99b4485bda8d88185a908c4/)  **Size:** 120,000 words  **Licence:** CLARIN ACA | Estonian | A word family consists of all the words that are linked by formative processes and share one and the same word stem. The resource is available for online browsing. |
| [The Dictionary of Foreign Words](https://metashare.ut.ee/repository/browse/the-dictionary-of-foreign-words/f8ef8e5495f211e6a6e4005056b40024838d5cfd92004b75ba384f57098aa12b/)  **Size:** 33,000 keywords  **Linguistic information:** orthography, inflection, semantics and etymology of the words; for many unadapted foreign words the pronunciation is given, both audio and written; the understanding and use of the words is facilitated by labels and cross-references.  **Licence:** CC-BY | Estonian | This is a dictionary that contains widespread Estonian words of foreign origin. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Dictionary of Old Võru Dialect](https://metashare.ut.ee/repository/browse/the-dictionary-of-old-voru-dialect/dd4f944c355b11e7a6e4005056b400240c86a3dcb1694dcfbd0e7c07e6c9b3b5/)  **Size:** 10,000 keywords  **Licence:** CC-BY | Estonian | The dictionary is on the archaic Võru dialect spoken in the second half of the 19th and in the first half of the 20th century. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Dictionary of the Eastern Dialect](https://metashare.ut.ee/repository/browse/the-dictionary-of-the-eastern-dialect/3294fb089f4911e6a6e4005056b40024d2f5e7ce781a4950b39f5845a3072336/)  **Size:** 7,000 keywords  **Licence:** CLARIN ACA | Estonian | This is a dialectal dictionary. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Dictionary of the Hiiu Dialect](https://metashare.ut.ee/repository/browse/the-dictionary-of-the-hiiu-dialect/032fafec9f4a11e6a6e4005056b40024d8e5fa2111d14c6ab7821ba8fd0a83e9/)  **Size:** 6,300 entries  **Licence:** CLARIN ACA | Estonian | This dialectal dictionary is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Dictionary of the Kihnu Dialect](https://metashare.ut.ee/repository/browse/the-dictionary-of-the-kihnu-dialect/b759d4c2e39a11e6a6e4005056b400245d8225446e764abcb01b7c09632dcc6b/)  **Size:** 6,000 keywords  **Licence:** CLARIN ACA | Estonian | This is the first dictionary of the Hiiu vernacular. The entries also give sentence examples with reference to their source parish. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Dictionary of the Mulgi Dialect](https://metashare.ut.ee/repository/browse/the-dictionary-of-the-mulgi-dialect/1fcab366e7b811e6a6e4005056b40024ce57d9d042734b9c8a670f76479f7f98/)  **Size:** 4,000 keywords  **Linguistic information:** sentence examples, synonymy  **Licence:** CLARIN ACA | Estonian | This is a dialectal dictionary searchable by keyword (in dialect spelling), standard Estonian equivalent (meaning) or style label. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Dictionary of the Seto Dialect](https://metashare.ut.ee/repository/browse/the-dictionary-of-the-seto-dialect/9a1640d89f4a11e6a6e4005056b40024edd62c01a31d4f33b0ac5c498baa4f7d/)  **Size:** 6,000 keywords  **Linguistic information:** dialect orthography and simplified spelling without diacritical marks  **Licence:** CLARIN ACA | Estonian | This is the first online Seto dictionary accessible to search both in dialect orthography and in simplified spelling (without diacritical marks). The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Estonian Etymological Dictionary](https://metashare.ut.ee/repository/browse/the-estonian-etymological-dictionary/0116f124ee6c11e4a6e4005056b4002469250a0f627744cbac4ff5ac88f7547a/)  **Size:** 6,600 keywords  **Linguistic information:** origin of Estonian word stems and the genetic links between words (loanwords, genuine stems, artificially coined stems etc)  **Licence:** CLARIN ACA | Estonian | This etymological dictionary is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Explanatory Dictionary of Estonian Language](https://metashare.ut.ee/repository/browse/the-explanatory-dictionary-of-the-estonian-language/8df79018f3cf11e4a6e4005056b400245fba483853d142e0adafb7a8044aecd5/)  **Size:** 154,000 keywords  **Licence:** CLARIN ACA | Estonian | This is the biggest dictionary of the Estonian language, which covers Estonian standard language as a whole, including spoken language and special languages. The resource is available for download from META-SHARE (CELR distribution). |
| [English-Estonian Machine Translation Dictionary](https://metashare.ut.ee/repository/browse/english-estonian-machine-translation-dictionary/442c5658463411e2a6e4005056b40024b1c86daf75e34c59aa8940bdc7754046/)  **Size:** 90,000 entries  **Licence:** BSD - Style | Estonian, English | This is a constantly incremented English-Estonian online dictionary, which is helpful in both human and machine translation. The resource is available for online browsing. |
| [The Online Dictionary of Estonian Sign Language - Estonian](https://metashare.ut.ee/repository/browse/the-online-dictionary-of-the-estonian-sign-language-estonian/428cf3c6e7ba11e6a6e4005056b400249ab92441a63244e3a3c4799d69b542ee/)  **Size:** 4,000 keywords  **Licence:** CLARIN ACA | Estonian, Estonian Sign Language | The dictionary contains over 4,000 entry words, enabling search either by an Estonian word or a by a sign of Estonian sign language.  The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [Finnish-Estonian Dictionary](https://metashare.ut.ee/repository/browse/finnish-estonian-dictionary/8899502cd5c011e7a6e4005056b40024719f97f36908495cbad4fd0795c63005/)  **Size:** 90,000 keywords  **Licence:** CC-BY | Estonian, Finnish | This is the web-version of the two-volume Finnish-Estonian Dictionary (2003). The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [Comprehensive Estonian-French Dictionary](https://metashare.ut.ee/repository/browse/comprehensive-estonian-french-dictionary/6479410ef4e311e4a6e4005056b4002443080069cb424332b1b68168450efa3f/)  **Size:** 15,300 keywords  **Licence:** undefined | Estonian, French | This is a bilingual dictionary that is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [Estonian-Latvian Dictionary](https://www.letonika.lv/eelv/)  **Size:** 25,000 entries | Estonian, Latvian | This is a bilingual dictionary that is available for online browsing. |
| [The Estonian-Latvian-Estonian Dictionary](https://metashare.ut.ee/repository/browse/the-estonian-latvian-dictionary/0413a04e9f4f11e6a6e4005056b400242d5f1aa4a7064ab4b7ad6f7a5f0822a5/)  **Size:** 47,000 keywords  **Licence:** CLARIN ACA | Estonian, Latvian | This bilingual dictionary is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Estonian-Mari Dictionary](https://metashare.ut.ee/repository/browse/estonian-mari-dictionary/22c9344e186d11e7a6e4005056b40024ce5bb509f3cb40598e1ab9c3fb5ce4b4/)  **Size:** 9,800 keywords  **Licence:** CLARIN ACA | Estonian, Mari | This bilingual dictionary is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Estonian-Norwegian-Estonian Dictionary](https://metashare.ut.ee/repository/browse/the-estonian-norwegian-dictionary/32815b2e9b6c11e6a6e4005056b40024efdedde1208545beaf897766577bf1d4/)  **Size:** 19,000 Estonian keywords, 21,000 Norwegian keywords  **Licence:** CLARIN ACA | Estonian, Norwegian | This bilingual dictionary is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [Estonian-Russian Learner's Orthographic Dictionary](https://metashare.ut.ee/repository/browse/estonian-russian-learners-orthographic-dictionary/404962cc9f4c11e6a6e4005056b40024640fb7652c0a4bce9667a4c46e1021ba/)  **Size:** 23,000 keywords  **Licence:** CLARIN ACA | Estonian, Russian | The Estonian-Russian Learner's dictionary is mainly meant for those learners of Estonian whose mother tongue is Russian or who learn it through Russian. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Estonian-Russian Dictionary](https://metashare.ut.ee/repository/browse/the-estonian-russian-dictionary/4ba9654c463411e2a6e4005056b40024e786814b938a4085b8b9969dd2ba5e78/)  **Size:** 66,423 keywords  **Licence:** CC-BY | Estonian, Russian | This is the most comprehensive Estonian-Russian dictionary. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Russian-Estonian Dictionary](https://metashare.ut.ee/repository/browse/the-russian-estonian-dictionary/9229be7895e911e6a6e4005056b4002455bf19cefdee47c6b9658cdfa54933cc/)  **Size:** 74,000 keywords  **Licence:** CC-BY | Estonian, Russian | This bilingual dictionary is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Estonian-Udmurt Dictionary](https://metashare.ut.ee/repository/browse/the-estonian-udmurt-dictionary/2588d5329f6d11e6a6e4005056b400249915c93c7ac54369a9e72cc06b65efb0/)  **Size:** 10,000 keywords  **Licence:** CLARIN ACA | Estonian, Udmurt | This is the first-ever bilingual dictionary of Estonian and Udmurt that also accounts for dialectal variation. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [Dictionary of Contemporary Finnish](http://urn.fi/urn:nbn:fi:lb-201411261)  **Size:** 100,000 entries  **Licence:** CLARIN RES | Finnish | This is a dictionary of modern Finnish that is available for online browsing through the Language Bank of Finland. |
| [Dictionary of Finnish Dialects](http://urn.fi/urn:nbn:fi:lb-201411263)  **Size:** accruing  **Licence:** CLARIN RES | Finnish | This is a dialect dictionary that is available for online browsing through the Language Bank of Finland. |
| [Dictionary of Old Literary Finnish](http://urn.fi/urn:nbn:fi:lb-201411262)  **Size:** accruing  **Licence:** CC-BY 4.0 | Finnish | This is a dictionary of Old Literary Finish. The resource is available for online browsing through the Language Bank of Finland. |
| [Wurdboek fan de Frsyke taal](http://gtb.ivdnt.org/search/?owner=wft)  **Size:** 120,000 lemmas  **Licence:** other | Frisian | This dictionary describes the meaning and history of hundred thousands of words in written Frisian, from the period 1800-1976. The resource is available for online browsing through the Dutch Language Institute (INT). |
| [Digital Dictionary of the German Language (DWDS)](https://www.dwds.de/)  **Size:** approx. 500,000 entries  **Linguistic information:** morphosyntactic information, pronunciation, etymology, lexical relations, examples of use, typical word connections | German | This collection consists of the following digitized dictionaries:  [Dictionary of contemporary German](https://www.dwds.de/d/wb-wdg)  [DWDS-Dictionary](https://www.dwds.de/d/wb-dwdswb)  [Etymological dictionary of German](https://www.dwds.de/d/wb-etymwb)  [The German Dictionary by Jacob Grimm and Wilhelm Grimm](https://www.dwds.de/d/wb-1dwb)  [The Great Dictionary of the German Language in 10 Volumes](https://www.dwds.de/d/wb-duden)   1. [OpenThesaurus](https://www.dwds.de/d/wb-openthesaurus)   This collection is available for online browsing from a dedicated environment associated with CLARIN-D. |
| [Multilingual dictionaries](http://hdl.grnet.gr/11500/ATHENA-0000-0000-23C7-8)  **Linguistic information:** MSD-tags, semantic information, multi-word units  **Licence:** CC-BY-NC | Greek, English, Bulgarian, Romanian, German, Polish | This is a collection of dictionaries created on the basis of Wikipedia articles. The collection is available for download from clarin:el. |
| [French-Greek lexicon of idioms](http://hdl.grnet.gr/11500/UOA-0000-0000-2574-4)  **Size:** 293 entries  **Licence:** CC-BY-ND | Greek, French | This is a glossary of idioms that supplies their wider contextual information (the context, the level of language, the connotation, the feeling that it transfers). The resource is available for download from clarin:el. |
| [Pronunciation Dictionary](http://clarin.is/en/resources/prondict/)  **Size:** 65,000 words  **Linguistic information:** IPA and SAMPA phonetic transcription  **Licence:** CC-BY 3.0 | Icelandic | This is a dictionary with pronunciation information for common Icelandic words. The resource is available for download from CLARIN-IS. |
| [ISLEX](http://clarin.is/en/resources/islex/)  **Size:** 48,700 entries  **Linguistic information:** Icelandic headwords with pronunciation (sound files) and inflection; translations into Danish, Norwegian (Bokmål and Nynorsk), Swedish, Faroese  **Licence:** CC-BY-NC-ND | Icelandic, Danish, Norwegian (Bokmål), Norwegian (Nynorsk), Swedish, Faroese | This multilingual dictionary is available for download from CLARIN-IS as well as for online browsing. |
| [Karelian Dictionary](http://urn.fi/urn:nbn:fi:lb-20140730157)  **Size:** 83,000 entries  **Licence:** CC-BY 4.0 | Karelian | This is a monolingual dictionary that is available for download and for online browsing through the Language Bank of Finland. |
| [Kven-Norwegian Bokmål dictionary](http://hdl.handle.net/11509/89)  **Size:** 8,468 entries  **Licence:** CC-BY 3.0 | Kven Finnish - Norwegian Bokmål | This is a bilingual dictionary that is available for download from CLARINO. |
| [Dictionary of Modern Latvian (MLVV)](http://tezaurs.lv/mlvv/)  **Size:** 45,877 entries | Latvian | The resource is available for online browsing through the tezaurus.lv repository. |
| [Dictionary of Standard Latvian Language (LLVV)](http://tezaurs.lv/llvv/)  **Size:** 64,200 entries | Latvian | This is a dictionary that contains examples of the standard Latvian language from the 1870s to the 1990s. The resource is available for online browsing. |
| [Historical Dictionary of Latvian (16th-17th Centuries) (LVVV)](http://tezaurs.lv/lvvv/)  **Size:** 1,800 entries | Latvian | This is a dictionary of Latvian from the period between the 16th and 17th centuries with a reference to the “Historical Corpus of the Latvian Language”. The resource is available for online browsing. |
| [Rendering Dictionary of Personal Names](https://www.letonika.lv/groups/default.aspx?g=2&r=1109)  **Size:** 4,000 entries | Latvian | This dictionary contains person names in four languages (Czech, French, Italian and German) and their renderings in Latvian in accordance with the rendering rules of proper nouns. The resource is available for online browsing. |
| [Tezaurs.lv](http://tezaurs.lv/)  **Size:** 309,016 entries | Latvian | This is a comprehensive corpus-based dictionary of the Latvian language. The resourceis available for online browsing. |
| [Mühlenbach­ Endzelin Latvian Dictionary (MEV)](http://tezaurs.lv/mev/)  **Size:** 120,000 entries | Latvian (German) | This is a bilingual dictionary that is available for online browsing through the tezaurus.lv repository. |
| [English-Latvian Dictionary](https://www.letonika.lv/groups/default.aspx?g=2&r=10331062&f=1)  **Size:** 70,000 entries | Latvian, English | This is a bilingual dictionary that is available for online browsing through the tezaurus.lv repository. |
| [Latvian-English Dictionary](https://www.letonika.lv/groups/default.aspx?g=2&r=10621033&f=1)  **Size:** 80,000 entries | Latvian, English | This is a bilingual dictionary that is available for online browsing through the tezaurus.lv repository. |
| [Latvian-Latgalian Dictionary](http://vuordineica.lv/)  **Size:** 25,700 entries | Latvian, Latgalian | This is a bilingual dictionary that is available for online browsing through the tezaurus.lv repository. |
| [Latvian-Lithuanian Dictionary](https://www.letonika.lv/lvlt/)  **Size:** 43,000 entries | Latvian, Lithuanian | This is a bilingual dictionary that is available for online browsing through the tezaurus.lv repository. |
| [Latvian-Russian Dictionary](https://www.letonika.lv/groups/default.aspx?g=2&r=10621049)  **Size:** 44,000 entries | Latvian, Russian | This is a bilingual dictionary that is available for online browsing through the tezaurus.lv repository. |
| [Lithuanian-Latvian Dictionary](https://www.letonika.lv/ltlv/)  **Size:** 60,000 entries | Lithuanian, Latvian | This is a bilingual dictionary that is available for online browsing through the tezaurus.lv repository. |
| [Lithuanian-Latvian-Latgalian Dictionary (LLL)](http://hipilatlit.ru.lv/dictionary/eng/dictionary.html)  **Size:** 42,061 headwords | Lithuanian, Latvian, Latgalian | This is a trilingual dictionary that is available for online browsing through a dedicated environment. |
| [Livonian-Estonian-Latvian Dictionary](http://www.livones.net/lingua/en) | Livonian, Estonian, Latvian | This is a trilingual dictionary that is available for online browsing through the tezaurus.lv repository. |
| [The Livonian-German Dictionary](https://metashare.ut.ee/repository/browse/the-livonian-german-dictionary/683f3e0ce7bb11e6a6e4005056b40024dcccb23809bf48b893737339a8263717/)  **Licence:** CLARIN ACA | Livonian, German | This bilingual dictionary is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [Basic English-Maltese Dictionary](https://portulanclarin.net/repository/browse/basic-english-maltese-dictionary/13fc5802abc511e1a404080027e73ea2a210be7dd5c44a3b9dd47afb4b2a34ef/)  **Size:** 5,458 entries  **Licence:** MS-BY-NC-SA | Maltese, English | This is a bilingual dictionary of alphabetically ordered English lemmas with their Maltese translation and Maltese pronunciation. The resource is unavailable for download or online browsing, but can be accessed by contacting the resource manager. |
| [North Saami-Finnish dictionary](http://hdl.handle.net/11509/97)  **Size:** 11,297 entries  **Licence:** CC-BY 3.0 | Northern Sami, Finnish | This is a bilingual dictionary that is available for download from CLARINO. |
| [Norwegian Bokmål-North Saami dictionary](http://hdl.handle.net/11509/94)  **Size:** 25,902 entries  **Licence:** CC-BY 3.0 | Norwegian Bokmål, Northern Sami | This is a bilingual dictionary that is available for download from CLARINO. |
| [Norwegian Bokmål-South Saami dictionary](http://hdl.handle.net/11509/93)  **Size:** 12,951 entries  **Licence:** CC-BY 3.0 | Norwegian Bokmål, Southern Sami | This is a bilingual dictionary that is available for download from CLARINO. |
| [Norwegian-Vietnamese Dictionary](http://hdl.handle.net/11509/22)  **Size:** 11,635 entries  **Linguistic information:** phonetic transcription, PoS-tags, translation and example sentences  **Licence:** CC-BY-NC-ND 3.0 | Norwegian, Vietnamese | This is a bilingual dictionary that is available for download from CLARINO. |
| [Dictionaries of the CLARIN-PL Lexical Platform](http://lexp.clarin-pl.eu/)  **Size:** 512,973 entries  **Licence:** Other - Institute of Computer Science PAS | Polish | This is a collection of the following 4 dictionaries, which are available through the same online environment:   1. [Grammatical Dictionary of Polish](http://lexp.clarin-pl.eu/) (334,845 entries) 2. [Vilnius Dictionary](https://eswil.ijp.pan.pl/index.php) (108,513 entries) 3. [The Polish Electronic Dictionary of the 16th c](http://spxvi.edu.pl/). (30,700 entries) 4. [The Polish Electronic Dictionary of the 17th and 18th c.](http://spxvi.edu.pl/) (38,915 entries)   All the 4 dictionaries are available through the same online lexical platform developed by CLARIN-PL. |
| [Polish Hask Collocation Database](http://pelcra.pl/hask_pl/)  **Size:** 106,723 entries containing 5,636,857 combinations  **Licence:** CC-BY-NC | Polish | HASK collocation dictionaries are phraseological databases used by linguists, language teachers, lexicographers, language materials developers, translators and other language professionals and casual dictionary users. The resource is available for online browsing through the PELCRA environment. |
| [Russian-Latvian Dictionary](https://www.letonika.lv/groups/default.aspx?g=2&r=10491062&f=1) | Russian, Latvian | This is a bilingual dictionary that is available for online browsing through the tezaurus.lv repository. |
| [Etymological Database of the Saami Languages](http://urn.fi/urn:nbn:fi:lb-20140730111)  **Size:** accruing  **Licence:** EUPL | Saami languages, Finnish | This is an etymological knowledge base available for online browsing through FIN-CLARIN. |
| [Dictionary of Lesser Used Slovenian Words](http://hdl.handle.net/11356/1130)  **Size:** 178,457 entries  **Linguistic information:** MSD-tags  **Licence:** CC-BY 4.0 | Slovenian | This is a dictionary that contains entries that were not included in the Dictionary of the Slovenian Standard Language. The resource is available for download from CLARIN.SI and for online browsing. |
| [Dictionary of Lesser Used Slovenian Words](http://hdl.handle.net/11356/1130)  **Size:** 178,457 entries  **Licence:** CC-BY 4.0 | Slovenian | This is a dictionary of infrequently used lexemes. The resource is available for download from CLARIN.SI. |
| [Dictionary of New Slovenian Words](http://hdl.handle.net/11356/1091)  **Size:** 6,399 entries  **Licence:** CC-BY 4.0 | Slovenian | This is a dictionary of new words and neologisms. The resource is available for download from CLARIN.SI and for online browsing. |
| [Dictionary of Slovenian Particles](http://hdl.handle.net/11356/1128)  **Size:** 429 entries  **Licence:** CC-BY 4.0 | Slovenian | This is a dictionary of connectives, disjuncts, emphasisers; not to be confused with English adverbial particles despite its name. The resource is available for download from CLARIN.SI and for online browsing. |
| [Dictionary of Slovenian Phrasemes](http://hdl.handle.net/11356/1129)  **Size:** 3,002 entries  **Linguistic information:** descriptions and explanations of Slovenian phrasemes  **Licence:** CC-BY-NC 4.0 | Slovenian | This is a multiword lexicon available for download from CLARIN.SI and for online browsing through a dedicated environment. |
| [Dictionary of the Slovenian Language in the Works of Janez Svetokriški](http://hdl.handle.net/11356/1092)  **Size:** 8,540 entries  **Licence:** CC-BY 4.0 | Slovenian | This is a historical dictionary compiling the works of Janez Svetokriški. The resource is available for download from CLARIN.SI and for online browsing. |
| [Dictionary of Twitterese Janes-Dict 1.0](http://hdl.handle.net/11356/1172)  **Size:** 1,002 entries  **Licence:** CC-BY-SA 4.0 | Slovenian | This is a dictionary of computer-mediated communication appearing on Twitter. The resource is available for download from CLARIN.SI and for online browsing. |
| [Reverse dictionary of Slovenian language](http://hdl.handle.net/11356/1094)  **Size:** 115,355 entries  **Licence:** CC-BY 4.0 | Slovenian | This is a dictionary where headwords are sorted by last-to-first letter. The resource is available for download from CLARIN.SI. |
| [Words of the 16th-Century Slovenian Literary Language](http://hdl.handle.net/11356/1127)  **Size:** 27,759 entries  **Linguistic information:** PoS-tags, phonological features  **Licence:** CC-BY 4.0 | Slovenian | This is a historical dictionary that provides comprehensive information on the vocabulary used in the Slovenian literary language during the period of the Reformation. The resource is available for download from CLARIN.SI and for online browsing. |
| [Slovenian-German Dictionary of Maks Pleteršnik (1894-1895)](https://www.clarin.si/repository/xmlui/handle/11356/1114)  **Size:** 103,185 entries  **Licence:** CC-BY 4.0 | Slovenian, German | This is a historical dictionary compiling the works of Maks Pleteršnik. The resource is available for download from CLARIN.SI and for online browsing. |
| [Japanese-Slovene learner's dictionary jaSlo 3.1](http://hdl.handle.net/11356/1050)  **Size:** 9,891 entries  **Linguistic information:** headwords in all Japanese scripts, PoS-tags, difficulty level, Slovene glosses, Slovenian translation equivalents, examples of use  **Licence:** CC-BY-SA 4.0 | Slovenian, Japanese | This is a learner’s dictionary for Slovenian students of Japanese. The resource is available for download from CLARIN.SI and for online browsing. |
| [Dalin's Dictionary (2017-10-16)](http://hdl.handle.net/10794/16)  **Size:** 62,975 entries  **Licence:** CC-BY 4.0 | Swedish | Dalin's Dictionary of 19th century Swedish. The resource can be download from the SWE-CLARIN repository and can be queried online through KARP. |
| [Hellquist's Swedish etymology (2017-10-16)](http://hdl.handle.net/10794/26)  **Size:** 12,376 entries  **Licence:** CC-BY 4.0 | Swedish | This is an etymological dictionary that is available for download from the SWE-CLARIN repository and can be queried online through KARP. |
| [Swedish-South Saami dictionary](http://hdl.handle.net/11509/99)  **Size:** 3,351 entries  **Licence:** CC-BY 3.0 | Swedish, Southern Sami | This is a bilingual dictionary that is available for download from CLARINO. |
| [Online Lexicon of Veps Language](http://urn.fi/urn:nbn:fi:lb-20140730178)  **Licence:** EUPL | Veps | This is a dictionary of the Veps Language available for online browsing through FIN-CLARIN. |

## Conceptual resources

There are 29 conceptual resources – wordnets, framenets, thesauri and ontologies – available in the CLARIN infrastructure, presented in detail in Table [3](#wordnets).

### VLO

All of the conceptual resources listed in Table [3](#wordnets) can be found in the VLO.

### Availability

The following 12 resources are available for download and online browsing:

1. [Bring (2015-05-08)](http://hdl.handle.net/10794/15)
2. [Cornetto-LMF](http://cornetto.clarin.inl.nl/index.html)
3. [Finnish FrameNet](http://urn.fi/urn:nbn:fi:lb-2019040202)
4. [Finnish WordNet](http://urn.fi/urn:nbn:fi:lb-2014052714)
5. [MWNPT-International WordNet of Portuguese](https://doi.org/21.11115/0000-000B-D509-B)
6. [Saldo](http://hdl.handle.net/10794/5)
7. [Semantic lexicon of Slovene sloWNet 3.1](http://hdl.handle.net/11356/1026)
8. [Swedish FrameNet (2017-10-16)](http://hdl.handle.net/10794/41)
9. [Swesaurus (2017-10-16)](http://hdl.handle.net/10794/44)
10. [The Sanat Version of the Finnish TransFrameNet](http://urn.fi/urn:nbn:fi:lb-2019040503)
11. [WordNet-SALDO (2017-10-16)](http://hdl.handle.net/10794/47)
12. [plWordNet 4.0](http://plwordnet.pwr.wroc.pl/wordnet/)

The following 16 resources are available for download only:

1. [Estonian Wordnet 2.1](https://metashare.ut.ee/repository/browse/estonian-wordnet-21/a2d6bb20988811e7a6e4005056b40024b1261fced7734054b5e1f9650181e4cf/)
2. [Geo-Net-PT 02](https://portulanclarin.net/repository/browse/geo-net-pt-02/1d9fe90acb7811e1a404080027e73ea241d3e2b7fb3c411481d3c23b787f8072/)
3. [Hontology](https://portulanclarin.net/repository/browse/hontology/a83c9d04cb7a11e1a404080027e73ea2359e10ea62b940109aabe03684aa5ea4/)
4. [ItalWordNet Kyoto](https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/ILC-65)
5. [ItalWordNet v.2](https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/ILC-62)
6. [IWN-LOD](http://hdl.handle.net/20.500.11752/ILC-66)
7. [Maltese automatically produced distributional thesaurus](https://portulanclarin.net/repository/browse/maltese-automatically-produced-distributional-thesaurus/1390982c3aec11e2a2aa782bcb0741350903f12d7d514e119f8386683e41b6f3/)
8. [Ontology for the area of Nanoscience and Nanotechnology](https://portulanclarin.net/repository/browse/ontology-for-the-area-of-nanoscience-and-nanotechnology/34e4d1d6cf5e11e1a404080027e73ea2296e3e6e58204292a7dbc1b0aae74f85/)
9. [Open Ancient Greek WordNet 0.5](https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/ILC-56)
10. [Thesaurus of Modern Slovene 1.0](http://hdl.handle.net/11356/1166)
11. [DanNet, Danish Wordnet (v 2.2)](http://hdl.handle.net/20.500.12115/25)
12. [STO semantics (The Danish SIMPLE Lexicon) - LMF format](http://hdl.handle.net/20.500.12115/26)
13. [Prolex](https://hdl.handle.net/11403/prolex/v3.1)
14. [GermaNet](http://www.sfs.uni-tuebingen.de/GermaNet/)
15. [NE\_SUMO\_PLWN\_mapping](https://clarin-pl.eu/dspace/handle/11321/286)
16. [PLWordNet to Sumo mapping](https://clarin-pl.eu/dspace/handle/11321/529)

The following resource is unavailable for download or online browsing:

1. [Polytropon EL Conceptual Lexicon](http://hdl.grnet.gr/11500/ATHENA-0000-0000-2862-5)

### Language

21 resources are monolingual and represent the following 14 languages:

1. Ancient Greek
2. Danish (2 resources)
3. Greek
4. Brazilian Portuguese
5. Dutch
6. Estonian
7. Finnish
8. German
9. Italian (3 resources)
10. Maltese
11. Polish (2 resources)
12. Portuguese
13. Swedish (4 resources)
14. Slovenian

5 resources are multilingual and represent the following 6 language combinations:

1. Finnish, English (2 resources)
2. French, English, Polish, Serbian
3. Slovenian, English

1. [Polish,](http://plwordnet.pwr.wroc.pl/wordnet/" \t "_parent) English
2. Portuguese, English, Spanish, French
3. Swedish, English

### Size

All the resources in this category contain information on size. Among wordnets, the largest is [plWordNet 4.0](http://plwordnet.pwr.wroc.pl/wordnet/), which contains 506,815 synsets, while the smallest is [Open Ancient Greek WordNet 0.5](https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/ILC-56), which contains 7,447 synsets.

### Linguistic information

Only the following 13 resources provide explicit description on the linguistic information included in the description field:

1. [Cornetto-LMF](http://cornetto.clarin.inl.nl/index.html" \t "_parent) (semantic relationships and combinatorial information)

1. [DanNet, Danish Wordnet (v 2.2)](http://hdl.handle.net/20.500.12115/25" \t "_parent) (hypernymy, hyperonymy)

1. [Estonian Wordnet 2.1](https://metashare.ut.ee/repository/browse/estonian-wordnet-21/a2d6bb20988811e7a6e4005056b40024b1261fced7734054b5e1f9650181e4cf/" \t "_parent) (PoS-tags, synonymy, antonymy, hypernymy, hyponymy, meronymy)
2. [Finnish WordNet](http://urn.fi/urn:nbn:fi:lb-2014052714) (PoS-tags, synonymy, antonymy, hypernymy, hyponymy, meronymy)
3. [Geo-Net-PT 02](https://portulanclarin.net/repository/browse/geo-net-pt-02/1d9fe90acb7811e1a404080027e73ea241d3e2b7fb3c411481d3c23b787f8072/" \t "_parent) (qualia structure and lexical relations)

1. [GermaNet](http://www.sfs.uni-tuebingen.de/GermaNet/) (MSD-tags, lexical relations (hypernomy, hyponomy))
2. [Hontology](https://portulanclarin.net/repository/browse/hontology/a83c9d04cb7a11e1a404080027e73ea2359e10ea62b940109aabe03684aa5ea4/) (term correlations, synonymy)

1. [ItalWordNet v.2](https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/ILC-62) (equivalence relations between Italian synsets and closest concepts in an Inter-Lingual index)

1. [MWNPT-International WordNet of Portuguese](https://doi.org/21.11115/0000-000B-D509-B" \t "_parent) (PoS-tags, hyponymy, hypernymy)
2. [Ontology for the area of Nanoscience and Nanotechnology](https://portulanclarin.net/repository/browse/ontology-for-the-area-of-nanoscience-and-nanotechnology/34e4d1d6cf5e11e1a404080027e73ea2296e3e6e58204292a7dbc1b0aae74f85/" \t "_parent) (lexical semantic information)
3. [Polytropon EL Conceptual Lexicon](http://hdl.grnet.gr/11500/ATHENA-0000-0000-2862-5) (lemmas, MSD-tags, lexical relations)

1. [Prolex](https://hdl.handle.net/11403/prolex/v3.1" \t "_parent) (inflecte forms)
2. [Thesaurus of Modern Slovene 1.0](http://hdl.handle.net/11356/1166) (core and near synonymy)

### Licence

21 resources are available under CC-BY. One resource lacks information on licence:

1. [IWN-LOD](http://hdl.handle.net/20.500.11752/ILC-66)

### Overview of the conceptual resources in the CLARIN infrastructure

Table 3: CLARIN conceptual resources and their characteristics, sorted alphabetically by language

|  |  |  |
| --- | --- | --- |
| **Resource** | **Language** | **Description** |
| [Open Ancient Greek WordNet 0.5](https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/ILC-56)  **Size:** 7,447 synsets  **Licence:** CC-BY-SA 4.0 | Ancient Greek | This is a wordnet that is available for download through ILC4CLARIN. |
| [Ontology for the area of Nanoscience and Nanotechnology](https://portulanclarin.net/repository/browse/ontology-for-the-area-of-nanoscience-and-nanotechnology/34e4d1d6cf5e11e1a404080027e73ea2296e3e6e58204292a7dbc1b0aae74f85/)  **Size:** 511 entries  **Linguistic information:** semantic relations  **Licence:** MS NC-NoReD-ND | Brazilian Portuguese | This is an ontology of concepts related to nanoscience and nanotechnology available for download from CLARIN PORTULAN. |
| [DanNet, Danish Wordnet (v 2.2)](http://hdl.handle.net/20.500.12115/25)  **Size:** 65,000 entries  **Linguistic information:** hypernymy, hyperonymy  **Licence:** DanNet 1.0 Licence | Danish | This wordnet is available for download from the CLARIN-DK repository. The resource is available in the .[csv](http://hdl.handle.net/20.500.12115/24) and .[owl](http://hdl.handle.net/20.500.12115/25) formats. |
| [STO semantics (The Danish SIMPLE Lexicon) - LMF format](http://hdl.handle.net/20.500.12115/26)  **Size:** 12,609 entries  **Licence:** CC-BY 4.0 | Danish | This resource presents a unified, ontology-based semantic model – the so-called SIMPLE model – representing an extended Qualia Structure from the Danish SIMPLE Lexicon. The resource is available for download from CLARIN-DK. |
| [Cornetto-LMF](http://cornetto.clarin.inl.nl/index.html)  **Size:** 130,000 entries  **Linguistic information:** semantic relationships and combinatorial information  **Licence:** other | Dutch | This is a resource that combines two resources with different semantic organisations: the Dutch Wordnet with its synset organisation and the Dutch Reference Lexicon which includes definitions, usage constraints, selectional restrictions, syntactic behaviours, illustrative contexts, etc. The resource is available for online browsing through a dedicated webpage. |
| [Estonian Wordnet 2.1](https://metashare.ut.ee/repository/browse/estonian-wordnet-21/a2d6bb20988811e7a6e4005056b40024b1261fced7734054b5e1f9650181e4cf/)  **Size:** 115,318 keywords, 84,150 synets.  **Linguistic information:** PoS-tags, synonymy, antonymy, hypernymy, hyponymy, meronymy  **Licence:** CC-BY-SA | Estonian | This is a wordnet that is available for download from META-SHARE (CELR distribution). |
| [Finnish FrameNet](http://urn.fi/urn:nbn:fi:lb-2019040202)  **Size:** 866 frames  **Licence:** CC-BY 4.0 | Finnish | This is a wordnet that is available for download from FIN-CLARIN as well as for online browsing. |
| [Finnish WordNet](http://urn.fi/urn:nbn:fi:lb-2014052714)  **Size:** 170,000 synsets  **Linguistic information:** PoS-tags, synonymy, antonymy, hypernymy, hyponymy, meronymy  **Licence:** CC-BY 3.0 | Finnish, English | This is a wordnet that is available for download from FIN-CLARIN as well as for online browsing. |
| [The Sanat Version of the Finnish TransFrameNet](http://urn.fi/urn:nbn:fi:lb-2019040503)  **Size:** 866 frames  **Licence:** CC-BY 4.0 | Finnish, English | This is a framenet that is available for download from FIN-CLARIN as well as for online browsing. |
| [Prolex](https://hdl.handle.net/11403/prolex/v3.1)  **Size:** 72,572 lexical relations  **Linguistic information:** inflected forms  **Licence:** Licence Publique Générale Amoindrie GNU | French, English, Polish, Serbian | This is an ontology of place names available for download from Ortolang. |
| [GermaNet](http://www.sfs.uni-tuebingen.de/GermaNet/)  **Size:** 61,659 synsets  **Linguistic information:** MSD-tags, lexical relations (hypernymy, hyponymy)  **Licence:** CLARIN ACA | German | This wordnet is available for download from a dedicated webpage hosted by Uni Tübinger/CLARIN-D. |
| [Polytropon EL Conceptual Lexicon](http://hdl.grnet.gr/11500/ATHENA-0000-0000-2862-5" \t "_parent)  **Size:** 4,000 Multi Word units: 15,000 tokens  **Linguistic information:** lemmas, MSD-tags, Lexical relations  **Licence:** under negotiation | Greek | This is a lexicon of lexical-semantic relations (e.g. synonymy, antonymy); lexical relations (e.g. word families, allomorphs, syntactic variants); morphosyntactic features (PoS, gender, declension, etc.), which is not yet available for download. |
| [ItalWordNet Kyoto](https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/ILC-65)  **Size:** 49,514 synsets  **Licence:** CC-BY-NC-SA 4.0 | Italian | This is a wordnet available for download download from ILC4CLARIN. |
| [ItalWordNet v.2](https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/ILC-62)  **Size:** 49,350 synsets  **Linguistic information:** equivalence relations between Italian synsets and closest concepts in an Inter-Lingual index  **Licence:** CC-BY-NC-SA 4.0 | Italian | This is a wordnet available for download from ILC4CLARIN. |
| [IWN-LOD](http://hdl.handle.net/20.500.11752/ILC-66)  **Size:** 49,350 synsets | Italian | This is an RDF–Linguistic Open Data version of the ItalWordNet v.2. The resource is available for download through ILC4CLARIN. |
| [Maltese automatically produced distributional thesaurus](https://portulanclarin.net/repository/browse/maltese-automatically-produced-distributional-thesaurus/1390982c3aec11e2a2aa782bcb0741350903f12d7d514e119f8386683e41b6f3/)  **Size:** 36,034 entries  **Linguistic information:** lemmas  **Licence**: CC-BY-SA | Maltese | This is a thesaurus that is available for download from CLARIN PORTULAN. |
| [NE\_SUMO\_PLWN\_mapping](https://clarin-pl.eu/dspace/handle/11321/286)  **Size:** 120 terms  **Licence:** CC BY SA 4.0 | Polish | This conceptual resource provides a mapping between named entities types, SUMO categories and plWordNet synsets. The resource is available for download from the CLARIN-PL repository. |
| [PLWordNet to Sumo mapping](https://clarin-pl.eu/dspace/handle/11321/529)  **Size:** 175,635 synsets mapped to SUMO  **Licence:** CC-BY-NC-SA 3.0 | Polish | This conceptual resource provides a mapping of plWordNet onto the SUMO ontology. The resource is available for download from the CLARIN-PL repository. |
| [plWordNet 4.0](http://plwordnet.pwr.wroc.pl/wordnet/)  **Size:** 506,815 senses, 347,564 synsets  **Licence:** plwordnet-2 | Polish, English | This is a lexico-semantic network which reflects the lexical system of the Polish language with projection to the English language. Słowosieć. The resource is available for download and browsing. |
| [Geo-Net-PT 02](https://portulanclarin.net/repository/browse/geo-net-pt-02/1d9fe90acb7811e1a404080027e73ea241d3e2b7fb3c411481d3c23b787f8072/)  **Size:** 701,209 concepts  **Linguistic information:** qualia structure and lexical relations (hyponyms, synonyms)  **Licence:** CC-BY | Portuguese | This is an ontology of Portuguese geographic concepts. It is available for download from CLARIN PORTULAN. |
| [MWNPT-International WordNet of Portuguese](https://doi.org/21.11115/0000-000B-D509-B)  **Size:** 17,200 synsets  **Linguistic information:** hyponymy and hypernymy **Licence:** MS NC-NoReD-ND | Portuguese | This is a wordnet that is available for download and online browsing through CLARIN PORTULAN. |
| [Hontology](https://portulanclarin.net/repository/browse/hontology/a83c9d04cb7a11e1a404080027e73ea2359e10ea62b940109aabe03684aa5ea4/)  **Size:** 282 concepts  **Linguistic information:** terms correlation, rules (lexical patterns) and synonyms  **Licence:** CC-BY-NC-SA | Portuguese, English, Spanish, French | This is an ontology of concepts from the accommodation sector is available for download from CLARIN PORTULAN. |
| [Thesaurus of Modern Slovene 1.0](http://hdl.handle.net/11356/1166)  **Size:** 105,473 entries  **Linguistic information:** core and near synonymy  **Licence:** CC BY-SA 4.0 | Slovenian | This is a thesaurus of the modern Slovenian language that is available for download from CLARIN.SI |
| [Semantic lexicon of Slovene sloWNet 3.1](http://hdl.handle.net/11356/1026)  **Size:** 43,460 synsets **Linguistic information:** lexical semantic relations **Licence:** CC-BY-SA 4.0 | Slovenian, English | This is a wordnet available for download from CLARIN.SI and for online browsing through a dedicated environment. |
| [Bring (2015-05-08)](http://hdl.handle.net/10794/15)  **Size:** 148,815 entries  **Licence:** CC-BY 4.0 | Swedish | This is a digital version of Bring's thesaurus (1930) that is available for download from the SWE-CLARIN repository and can be queried online through KARP. |
| [Saldo](http://hdl.handle.net/10794/5)  **Size:** 131,020 entries  **Licence:** CC-BY 4.0 | Swedish | This is an extensive lexicon resource for modern Swedish written language. The resource can be download from the SWE-CLARIN repository and can be queried online through KARP. |
| [Swedish FrameNet (2017-10-16)](http://hdl.handle.net/10794/41)  **Size:** 1,195 entries  **Licence:** CC-BY 4.0 | Swedish | This is a Swedish conceptual resource that employs the FrameNet++ annotation. The resource can be download from the SWE-CLARIN repository and can be queried online through KARP. |
| [Swesaurus (2017-10-16)](http://hdl.handle.net/10794/44)  **Size:** 15,010 entries  **Licence:** CC-BY 4.0 | Swedish | This is a Swedish wordnet that can be downloaded from the SWE-CLARIN repository and can be queried online through KARP. |
| [WordNet-SALDO (2017-10-16)](http://hdl.handle.net/10794/47)  **Size:** 6,989 entries  **Licence:** CC-BY 4.0 | Swedish, English | This wordnet represents a link between SALDO senses and Core WordNet. The resource can be download from the SWE-CLARIN repository and can be queried online through KARP. |

## Glossaries

There are 33 glossaries available in the CLARIN infrastructure, presented in detail in Table [4](#glossaries).

### VLO

All of the resources listed in Table [4](#glossaries) can be found in the VLO.

### Availability

The following 7 glossaries are available for download and online browsing:

1. [A Terminological Inventory for Biodiversity](https://portulanclarin.net/repository/browse/a-terminological-inventory-for-biodiversity/73b84c28fd1a11e6a2aa782bcb074135a9fd1289fd654abeb973c65280441a6d/)
2. [Dictionary of Estonian Place Names](https://metashare.ut.ee/repository/browse/dictionary-of-estonian-place-names/ade9bda617b211e8a6e4005056b40024e6e761d148c64fe0b4182285ccff16a7/)
3. [ESTERM (20190101)](https://metashare.ut.ee/repository/browse/esterm-v20190101/4d88ce202ada11e9a6e4005056b40024ff91181bfcce427fb3993820983bad6d/)
4. [The Dictionary of Education](https://metashare.ut.ee/repository/browse/the-dictionary-of-education/497c6772a68611e6a6e4005056b40024ac854587c87f4f2e95de5deb4b28fd7d/)
5. [The Dictionary of the Clothing Terminology of the Zilja Dialect in Canale Valley (Kanalska dolina – Val Canale – Kanaltal – Valcjanâl)](http://hdl.handle.net/11356/1217)
6. [The English-Estonian Maritime Dictionary](https://metashare.ut.ee/repository/browse/the-english-estonian-maritime-dictionary/c4f9aeb0a68511e6a6e4005056b400245fad4feafda9464e82b26c44e14a822a/)
7. [UHR's Termbase for Norwegian higher education institutions](http://hdl.handle.net/11509/122)

The following 21 glossaries are available for download only:

1. [Bibterms. Glossary of archive and library sciences](http://hdl.grnet.gr/11500/IONION-0000-0000-250E-8)
2. [English for business](http://hdl.handle.net/11509/116)
3. [French-Greek and Greek-French lexicon of football terms](http://hdl.grnet.gr/11500/UOA-0000-0000-2573-5)
4. [Glossary e-book "Open Linked Data and Applications"](http://hdl.grnet.gr/11500/IONION-0000-0000-25E7-2)
5. [Glossary of Ancient Greek dramaturgy](http://hdl.grnet.gr/11500/IONION-0000-0000-2510-4)
6. [Glossary of feminist theory](http://hdl.grnet.gr/11500/UOA-0000-0000-25EB-E)
7. [Glossary of Medical Terms](http://hdl.grnet.gr/11500/AUTH-0000-0000-2C49-E)
8. [Greek Crisis Multilingual Term Project](http://hdl.grnet.gr/11500/IONION-0000-0000-2511-3)
9. [Greek-French lexicon of football](http://hdl.grnet.gr/11500/UOA-0000-0000-25ED-C)
10. [INT IMPACT NE Lexicon](https://ivdnt.org/downloads/taalmaterialen/tstc-ne-lexicon-nederlands)
11. [MFSA\_Maltese\_Company\_Registry](https://portulanclarin.net/repository/browse/mfsa_maltese_company_registry/0a241662bc6911e1a404080027e73ea2c9eb952b442542538520035aeb59596d/)
12. [Multilingual glossary (Department of Foreign Languages, Translation and Interpretation)](http://hdl.grnet.gr/11500/IONION-0000-0000-3D44-0)
13. [Multilingual lexicon of toponyms](http://hdl.grnet.gr/11500/ATHENA-0000-0000-23C1-E)
14. [Netterms. Computer networking terminology](http://hdl.grnet.gr/11500/IONION-0000-0000-250F-7)
15. [Norwegian biodiversity terminology database](http://hdl.handle.net/11509/115)
16. [Orossimo Terminological Resource](http://hdl.grnet.gr/11500/ATHENA-0000-0000-4B49-B)
17. [PANACEA Environment Bilingual Glossary EL-EN (Greek-English)](http://hdl.grnet.gr/11500/ATHENA-0000-0000-23F6-3)
18. [PANACEA Labour Bilingual Glossary EL-EN (Greek-English)](http://hdl.grnet.gr/11500/ATHENA-0000-0000-23F5-4)
19. [Time-sensitive inventory of medical terminology](https://portulanclarin.net/repository/browse/time-sensitive-inventory-of-medical-terminology/1a171f22719e11e5a2aa782bcb074135530d6c33185d42539d22846d41ff5ec2/)
20. [Xenophobia](https://inventory.clarin.gr/resources/search/?q=xenophobia)
21. [Τρίγλωσσο Ορολογικό Λεξικό (ΤΟΛ)](http://hdl.grnet.gr/11500/ATHENA-0000-0000-5837-0)

The following 5 glossaries are available only for online browsing:

1. [MILITERM (20180801)](https://metashare.ut.ee/repository/browse/militerm-v20180801/9d5e4398277b11e8a6e4005056b40024e9e064c3685940ec950be7c93b4bec75/)
2. [Name Component Lexicon](http://urn.fi/urn:nbn:fi:lb-20140730188)
3. [The Family Name Database of the Institute of the Estonian Language](https://metashare.ut.ee/repository/browse/the-family-name-database-of-the-institute-of-the-estonian-language/f7aff296e7bd11e6a6e4005056b40024c99cce2897d14ea3891285eea548c474/)
4. [The Helsinki Term Bank for the Arts and Sciences](http://urn.fi/urn:nbn:fi:lb-20140730152)
5. [The Place Name Database of the Institute of the Estonian Language (KNAB)](https://metashare.ut.ee/repository/browse/the-place-name-database-of-the-institute-of-the-estonian-language-knab/563f609af3d111e4a6e4005056b4002492057892e4f04a83afb5d759c60b87f9/)

### Language

10 glossaries are monolingual and represent the following 6 languages:

1. Dutch (1 resource)
2. English (2 resources)
3. Estonian (3 resources)
4. Greek (2 resources)
5. Slovenian (1 resource)
6. Swedish (1 resource)

23 glossaries are multilingual and represent the following 14 language combinations:

1. English, Greek (10 resources)
2. Estonian, English
3. Estonian, English, Finnish, French, German
4. Estonian, English, Finnish, French, German, Russian
5. Estonian, English, German, Finnish, Russian
6. Finnish, multilingual
7. Greek, English, Turkish
8. Greek, French
9. Greek, Italian, Turkish, English, German, French, Spanish, Castilian
10. Maltese, English
11. Norwegian Bokmal, English
12. Norwegian Bokmål, English, Norwegian Nynorsk
13. Norwegian Bokmål, Norwegian Nynorsk, Latin
14. Polish, English, Romanian, Hungarian, Bulgarian, Polish, Greek, German, Serbian, Slovak, Croatian

### Size

In terms of entries, the largest resource is the glossary [A Terminological Inventory for Biodiversity](https://portulanclarin.net/repository/browse/a-terminological-inventory-for-biodiversity/73b84c28fd1a11e6a2aa782bcb074135a9fd1289fd654abeb973c65280441a6d/), containing 288,562 entries, while the smallest is [Netterms. Computer networking terminology](http://hdl.grnet.gr/11500/IONION-0000-0000-250F-7" \t "_parent), containing 205 entries.

The following glossaries lacks information on size:

1. [Multilingual glossary (Department of Foreign Languages, Translation and Interpretation)](http://hdl.grnet.gr/11500/IONION-0000-0000-3D44-0" \t "_parent)
2. [Name Component Lexicon](http://urn.fi/urn:nbn:fi:lb-20140730188)
3. [The Family Name Database of the Institute of the Estonian Language](https://metashare.ut.ee/repository/browse/the-family-name-database-of-the-institute-of-the-estonian-language/f7aff296e7bd11e6a6e4005056b40024c99cce2897d14ea3891285eea548c474/)

### Linguistic information

Only the following 8 resources provide explicit information on the linguistic information included, though note that some labels are vague (e.g. [Multilingual lexicon of toponyms](http://hdl.grnet.gr/11500/ATHENA-0000-0000-23C1-E) lists the annotation layer “semantics – semantic class”, but does not further specify what semantic class pertains to).

1. [A Terminological Inventory for Biodiversity](https://portulanclarin.net/repository/browse/a-terminological-inventory-for-biodiversity/73b84c28fd1a11e6a2aa782bcb074135a9fd1289fd654abeb973c65280441a6d/" \t "_parent) (semantic relatedness)

1. [Greek-French lexicon of football](http://hdl.grnet.gr/11500/UOA-0000-0000-25ED-C" \t "_parent) (lexical semantic features)
2. [Multilingual glossary (Department of Foreign Languages, Translation and Interpretation)](http://hdl.grnet.gr/11500/IONION-0000-0000-3D44-0) (lemmas)

1. [Multilingual lexicon of toponyms](http://hdl.grnet.gr/11500/ATHENA-0000-0000-23C1-E" \t "_parent) (lemmas, semantics – semantic class, translation equivalents)

1. [Orossimo Terminological Resource](http://hdl.grnet.gr/11500/ATHENA-0000-0000-4B49-B" \t "_parent) (lexical semantic relations, usage notes)

1. [PANACEA Environment Bilingual Glossary EL-EN (Greek-English)](http://hdl.grnet.gr/11500/ATHENA-0000-0000-23F6-3) (lemmas, translation equivalent)

1. [Time-sensitive inventory of medical terminology](https://portulanclarin.net/repository/browse/time-sensitive-inventory-of-medical-terminology/1a171f22719e11e5a2aa782bcb074135530d6c33185d42539d22846d41ff5ec2/" \t "_parent) (Semantics - relations [synonyms, holonymy, hyperonymy...])
2. [The Dictionary of the Clothing Terminology of the Zilja Dialect in Canale Valley (Kanalska dolina – Val Canale – Kanaltal – Valcjanâl)](http://hdl.handle.net/11356/1217) (MSD-tags, phonological information)

### Licence

The most common licence types are as follows:

1. CC-BY (23 resources)
2. CLARIN ACA (5 resources)
3. CLARIN PUB (2 resources)

The following resource lacks information on licence:

1. [Λεξικό Γλωσσολογικών όρων: Γερμανικά – Ελληνικά - Αγγλικά (lexicon of linguistic terms: DE-EL-EN)](http://hdl.grnet.gr/11500/AUTH-0000-0000-28CB-F)

### Overview of the glossaries in the CLARIN infrastructure

Table 4: CLARIN glossaries and their characteristics, sorted alphabetically by language

|  |  |  |  |
| --- | --- | --- | --- |
| **Resource** | **Language** | | **Description** |
| [INT IMPACT NE Lexicon](https://ivdnt.org/downloads/taalmaterialen/tstc-ne-lexicon-nederlands)  **Size:** 200,000 lemmas  **Licence:** other | Dutch | | This is a lexicon that features historical names and variants for the period between 1750 and 1945. The resource is available for download from the Dutch Language Institute (INT). |
| [A Terminological Inventory for Biodiversity](https://portulanclarin.net/repository/browse/a-terminological-inventory-for-biodiversity/73b84c28fd1a11e6a2aa782bcb074135a9fd1289fd654abeb973c65280441a6d/)  **Size:** 288,562 Entries  **Linguistic information:** semantic relatedness  **Licence:** CC-BY | English | | This is a glossary of animal species available for download from CLARIN PORTULAN and online browsing. |
| [Time-sensitive inventory of medical terminology](https://portulanclarin.net/repository/browse/time-sensitive-inventory-of-medical-terminology/1a171f22719e11e5a2aa782bcb074135530d6c33185d42539d22846d41ff5ec2/)  **Size:** 175,000 entries  **Linguistic information:** Semantics - relations (synonyms, holonymy, hyperonymy, ...)  **Licence**: CC-BY | | English | This glossary is available for download from CLARIN PORTULAN. |
| [Glossary e-book "Open Linked Data and Applications"](http://hdl.grnet.gr/11500/IONION-0000-0000-25E7-2)  **Size:** 100 semantic units  **Licence**: CC-BY | | English, Greek | This glossary is available for download from clarin:el. |
| [Glossary of Medical Terms](http://hdl.grnet.gr/11500/AUTH-0000-0000-2C49-E)  **Size:** 3,750 words  **Licence**: CC-BY | | English, Greek | This is a glossary of medical terms. The resource is available for download from clarin:el. |
| [Multilingual glossary (Department of Foreign Languages, Translation and Interpretation)](http://hdl.grnet.gr/11500/IONION-0000-0000-3D44-0)  **Linguistic information:** lemmas  **Licence:** CC-BY-NC | English, Greek | | This is a general-purpose glossary that is available for download from clarin:el. |
| [Netterms. Computer networking terminology](http://hdl.grnet.gr/11500/IONION-0000-0000-250F-7)  **Size:** 205 entries  **Licence:** CC-BY | English, Greek | | This is glossary of terms related to computer networking available for download from clarin:el. |
| [Dictionary of Estonian Place Names](https://metashare.ut.ee/repository/browse/dictionary-of-estonian-place-names/ade9bda617b211e8a6e4005056b40024e6e761d148c64fe0b4182285ccff16a7/)  **Size:** 6,214 keywords  **Licence:** CC-BY | Estonian | | This etymological glossary contains all official village names of Estonia and selected group of other Estonian place names. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Family Name Database of the Institute of the Estonian Language](https://metashare.ut.ee/repository/browse/the-family-name-database-of-the-institute-of-the-estonian-language/f7aff296e7bd11e6a6e4005056b40024c99cce2897d14ea3891285eea548c474/)  **Licence:** CLARIN ACA | Estonian | | This glossary is available for online browsing. |
| [The Place Name Database of the Institute of the Estonian Language (KNAB)](https://metashare.ut.ee/repository/browse/the-place-name-database-of-the-institute-of-the-estonian-language-knab/563f609af3d111e4a6e4005056b4002492057892e4f04a83afb5d759c60b87f9/)  **Size:** 428,344 elements; 145,261 elements  **Licence:** CLARIN ACA | Estonian | | This glossary is available for online browsing. |
| [The English-Estonian Maritime Dictionary](https://metashare.ut.ee/repository/browse/the-english-estonian-maritime-dictionary/c4f9aeb0a68511e6a6e4005056b400245fad4feafda9464e82b26c44e14a822a/)  **Size:** 30,000 keywords  **Licence:** CLARIN ACA | | Estonian, English | This glossary contains nautical terms. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [MILITERM (20180801)](https://metashare.ut.ee/repository/browse/militerm-v20180801/9d5e4398277b11e8a6e4005056b40024e9e064c3685940ec950be7c93b4bec75/)  **Licence**: CC-BY | | Estonian, English, Finnish, French, German | MILITERM is a multilingual termbase of military, security and defence terms, enabling search by Estonian, English, French and German keywords. The Estonian and English terms are provided with definitions. The resource is available for online browsing. |
| [ESTERM (20190101)](https://metashare.ut.ee/repository/browse/esterm-v20190101/4d88ce202ada11e9a6e4005056b40024ff91181bfcce427fb3993820983bad6d/)  **Size:** 53,327 concepts  **Licence:** CC-BY | | Estonian, English, Finnish, French, German, Russian | This is a termbase that includes terms mainly from the legal acts of the Republic of Estonia and the European Union. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Dictionary of Education](https://metashare.ut.ee/repository/browse/the-dictionary-of-education/497c6772a68611e6a6e4005056b40024ac854587c87f4f2e95de5deb4b28fd7d/)  **Size:** 4,850 keywords  **Licence:** CLARIN ACA | | Estonian, English, German, Finnish, Russian | This is a comprehensive glossary of education terminology. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website. |
| [The Helsinki Term Bank for the Arts and Sciences](http://urn.fi/urn:nbn:fi:lb-20140730152)  **Size:** accruing  **Licence:** CC-BY-SA | | Finnish, multilingual | This is a terminological database available for online browsing through for online browsing through FIN-CLARIN. |
| [French-greek and greek-french lexicon of football terms](http://hdl.grnet.gr/11500/UOA-0000-0000-2573-5)  **Size:** 843 entries  **Licence:** CC-BY – ND | French-Greek | | This bilingual glossary of football terminology is available for download from clarin:el. |
| [Glossary of Ancient Greek dramaturgy](http://hdl.grnet.gr/11500/IONION-0000-0000-2510-4)  **Size:** 228 Lexical Types  **Licence**: CC-BY | | Greek | This is a glossary of terms related to Ancient Greek Dramaturgy. The resource is available for download from clarin:el. |
| [Xenophobia](https://inventory.clarin.gr/resources/search/?q=xenophobia)  **Size:** 12 knowledge bases  **Licence:** CC-BY | Greek | | This is a collection of 12 knowledge bases on various forms of xenophobia, available for download from clarin:el. |
| [Bibterms. Glossary of archive and library sciences](http://hdl.grnet.gr/11500/IONION-0000-0000-250E-8)  **Size:** 1,223 Lexical Types  **Licence**: CC-BY | | Greek, English | This is a glossary of archival library science terms. The resource is available for download from clarin:el. |
| [Glossary of feminist theory](http://hdl.grnet.gr/11500/UOA-0000-0000-25EB-E)  **Size:** 100 entries  **Licence**: CC-BY-NC | | Greek, English | This is a glossary of terms related to Feminist theory/criticism, and gender theory. The resource is available for download from clarin:el. |
| [Orossimo Terminological Resource](http://hdl.grnet.gr/11500/ATHENA-0000-0000-4B49-B)  **Size:** 22,308 terms  **Linguistic information:** lexical semantic relations, usage notes  **Licence**: CC-BY-NC-SA 4.0 | | Greek, English | This is a glossary of terms from various domains e.g. History, Computer Science, Biology etc. The resource is available for download from clarin:el. |
| [PANACEA Environment Bilingual Glossary EL-EN (Greek-English)](http://hdl.grnet.gr/11500/ATHENA-0000-0000-23F6-3)  **Size:** 18,582 entries  **Linguistic information:** lemmas, translation equivalent  **Licence:** CC-BY | | Greek, English | This glossary is available for download from clarin:el. |
| [Τρίγλωσσο Ορολογικό Λεξικό (ΤΟΛ)](http://hdl.grnet.gr/11500/ATHENA-0000-0000-5837-0)  **Size:** 5,224 Entries  **Licence:** CC-BY-NC–ND | | Greek, English, Turkish | This is a trilingual term glossary for foreign students who want to improve their language skills in Greek and familiarize themselves with information technology. The resource is available for download from clarin:el. |
| [Greek-French lexicon of football](http://hdl.grnet.gr/11500/UOA-0000-0000-25ED-C)  **Size:** 870 entries  **Linguistic information:** lexical semantic features  **Licence:** CC-BY-NC | Greek, French | | This is a glossary of football terms available for download from clarin:el. |
| [Greek Crisis Multilingual Term Project](http://hdl.grnet.gr/11500/IONION-0000-0000-2511-3)  **Size:** 820 concepts  **Licence**: CC-BY | | Greek, Italian, Turkish, English, German, French, Spanish, Castilian | This is a glossary of terms related to the Greek government-debt crisis. The resource is available for download from clarin:el. |
| [MFSA\_Maltese\_Company\_Registry](https://portulanclarin.net/repository/browse/mfsa_maltese_company_registry/0a241662bc6911e1a404080027e73ea2c9eb952b442542538520035aeb59596d/)  **Size:** 45,993 entries  **Licence:** CC-BY-NC | | Maltese, English | This glossary is available for download from CLARIN PORTULAN. |
| [English for business](http://hdl.handle.net/11509/116)  **Size:** 498 entries  **Licence:** CLARIN ACA | Norwegian Bokmål, English | | This glossary includes terms from the domains of economics and business administration. The resource is available for download from CLARINO. |
| [UHR's Termbase for Norwegian higher education institutions](http://hdl.handle.net/11509/122)  **Size:** 1,786 entries  **Licence:** CLARIN\_PUB-BY | | Norwegian Bokmål, English, Norwegian Nynorsk | This is a database of terms available for download from CLARINO and for online browsing. |
| [Norwegian biodiversity terminology database](http://hdl.handle.net/11509/115)  **Size:** 297 entries  **Licence:** CLARIN\_PUB-BY | | Norwegian Bokmål, Norwegian Nynorsk, Latin | This glossary includes terms from the domains of biology. The resource is available for download from CLARINO. |
| [Multilingual lexicon of toponyms](http://hdl.grnet.gr/11500/ATHENA-0000-0000-23C1-E)  **Size:** 281,375 entries  **Linguistic information:** lemmas, semantics, translation equivalents  **Licence:** CC-BY-SA | Polish + translations into English, Romanian, Hungarian, Bulgarian, Polish, Greek, German, Serbian, Slovak, Croatian | | This is a glossary of toponyms available for download from clarin:el. |
| [The Dictionary of the Clothing Terminology of the Zilja Dialect in Canale Valley (Kanalska dolina – Val Canale – Kanaltal – Valcjanâl)](http://hdl.handle.net/11356/1217)  **Size:** 594 entries  **Linguistic information:** MSD-tags, phonological information  **Licence:** CC-BY 4.0 | Slovenian | | This is a term dictionary of Clothing Terminology of the Zilja Dialect spoken in the Val Canale Valley. Entries in the dictionary are equipped with [sound clips](http://hdl.handle.net/11356/1220) and [illustrative photographs](http://hdl.handle.net/11356/1221) of traditional clothing items. The resource is available for download from CLARIN.SI and for online browsing. |
| [Name Component Lexicon](http://urn.fi/urn:nbn:fi:lb-20140730188)  **Licence:** EUPL | Swedish | | This is a glossary of the most common elements out of which Swedish place names are built. The resource is available online through FIN-CLARIN. |

## Wordlists

There are 56 wordlists available in the CLARIN infrastructure, described in detail in Table [5](#Wordlists).

### VLO

All the wordlists in Table [5](#Wordlists) can be found in the VLO except for:

1. [Deutsche Wortschatz](http://corpora.informatik.uni-leipzig.de/de?corpusId=deu_newscrawl_2011)

### Availability

The following 5 wordlists are available for download and online browsing:

1. [Estonian Frequency Dictionary (ver. 2.0)](https://metashare.ut.ee/repository/browse/estonian-frequency-dictionary-ver-20/07a87ad00f6f11e9a6e4005056b400241184c428448348b5a5654740a32e57bc/)
2. [Idioms from the NEO lexicon DB](http://hdl.handle.net/10794/27)
3. [Kelly (2017-10-16)](http://hdl.handle.net/10794/28)
4. [Swedish words, LEXIN](http://hdl.handle.net/10794/40)
5. [Vocation list (2015-01-10)](http://hdl.handle.net/10794/48)

The following 43 wordlists are available for download only:

1. [Dictionary for the CST Lemmatizer](https://repository.clarin.dk/repository/xmlui/handle/20.500.12115/19)
2. [Kres corpus n-grams 2.0](http://hdl.handle.net/11356/1193)
3. [Gos corpus n-grams 2.0](https://www.clarin.si/repository/xmlui/handle/11356/1195)
4. [IMP corpus n-grams 2.0](http://hdl.handle.net/11356/1194)
5. [Janes corpus n-grams 1.0](http://hdl.handle.net/11356/1192)
6. [Psycholinguistic Descriptives](http://urn.fi/urn:nbn:fi:lb-2018081601)
7. [Finnish Verbal Colorative Constructions](http://urn.fi/urn:nbn:fi:lb-2017090401)
8. [INT Historical Word List](https://ivdnt.org/downloads/taalmaterialen/tstc-int-historische-woordenlijst)
9. [JRC-Names - a multilingual named entity resource](http://hdl.grnet.gr/11500/ATHENA-0000-0000-258E-7)
10. [KELLY word-list Greek](http://hdl.grnet.gr/11500/ATHENA-0000-0000-25C1-C)
11. [Keywords and n-grams from a textbook corpus](http://hdl.handle.net/11356/1215)
12. Labial vibrants in Mangbetu: Archival form
13. [LCM-PL](http://hdl.handle.net/11321/585)
14. Lexical functions of Spanish verb-noun collocations
15. [Maltese Fiction Wordlist](https://portulanclarin.net/repository/browse/maltese-fiction-wordlist/dd3d5a0849dc11e2a2aa782bcb074135d9cb4fe73e884aa3b771fd0a7a11a2b7/)
16. [Maltese Wordlist](https://portulanclarin.net/repository/browse/maltese-wordlist/1eb37910a35a11e1a404080027e73ea2dbdb15cf3d18412b943f9db788f20d5c/)
17. [Modern Finnish Word List](http://urn.fi/urn:nbn:fi:lb-20140730169)
18. [Neologisms Online v3](http://tst.inl.nl/producten/neon_v3/versie1.1/displaywidx.php)
19. [Ngbugu digital wordlist: Archival form](https://hdl.handle.net/11403/sldr000833/v1)
20. [Relative frequencies of part-of-speech n-grams in native and translated Finnish literary prose](http://urn.fi/urn:nbn:fi:lb-2018120301)
21. [The Finnish N-grams 1820-2000 of the Newspaper and Periodical Corpus of the National Library of Finland](http://urn.fi/urn:nbn:fi:lb-2014073038)
22. [The Swedish N-grams 1770-1940 of the Newspaper and Periodical Corpus of the National Library of Finland](http://urn.fi/urn:nbn:fi:lb-2014091902)
23. [Topics of library and information science](http://hdl.grnet.gr/11500/IONION-0000-0000-25BD-2)
24. [Λεξικό Γλωσσολογικών όρων: Γερμανικά – Ελληνικά - Αγγλικά (lexicon of linguistic terms: DE-EL-EN)](http://hdl.grnet.gr/11500/AUTH-0000-0000-28CB-F)
25. [Vocabulaire d'archéologie](https://hdl.handle.net/11403/vocabulaire-darcheologie/v1)
26. Vocabulaire d'art et archéologie
27. [Vocabulaire de géographie de l'Amérique du Nord](https://hdl.handle.net/11403/vocabulaire-de-geographie-de-lamerique-du-nord/v1)
28. [Vocabulaire de Nutrition artificielle](https://hdl.handle.net/11403/vocabulaire-de-nutrition-artificielle/v1)
29. Vocabulaire de Pathologies humaines
30. Vocabulaire de philosophie

1. [Vocabulaire de préhistoire et protohistoire](https://hdl.handle.net/11403/prehistoire-protohistoire/v1.1" \t "_parent)
2. [Vocabulaire de Psychopathologie](https://hdl.handle.net/11403/vocabulaire-de-psychopathologie/v1" \t "_parent)
3. [Vocabulaire de Sciences de l'éducation](https://hdl.handle.net/11403/vocabulaire-de-sciences-de-leducation/v1)
4. [Vocabulaire de Sciences du langage](https://hdl.handle.net/11403/vocabulaire-de-sciences-du-langage/v1)
5. [Vocabulaire de sociologie](https://hdl.handle.net/11403/vocabulaire-de-sociologie/v1)
6. [Vocabulaire de Transferts de chaleur](https://hdl.handle.net/11403/transferts-de-chaleur/v1.2)
7. [Vocabulaire de Transfusion sanguine](https://hdl.handle.net/11403/vocabulaire-de-transfusion-sanguine/v1)
8. Vocabulaire d'ethnologie
9. [Vocabulaire d'histoire des sciences et des techniques](https://hdl.handle.net/11403/vocabulaire-dhistoire-et-sciences-de-la-litterature/v1)
10. [Vocabulaire d'histoire et sciences de la littérature](https://hdl.handle.net/11403/vocabulaire-dhistoire-et-sciences-de-la-litterature/v1)
11. [Vocabulaire d'Histoire et sciences des religions](https://hdl.handle.net/11403/vocabulaire-histoire-et-sciences-des-religions/v1)
12. [Vocabulaire de sciences de la Terre](https://hdl.handle.net/11403/vocabulaire-de-sciences-de-la-terre/v1)
13. Vocabulaire d'Electronique et électro-optique

The following 8 wordlists are available only for online browsing:

1. [Frequencies of Early Modern Finnish Words](http://urn.fi/urn:nbn:fi:lb-20140730139)
2. [Frequencies of Old Literary Finnish Words](http://urn.fi/urn:nbn:fi:lb-20140730166)
3. [Frequency Lexicon of the Finnish Newspaper Language](http://urn.fi/urn:nbn:fi:lb-201405272)
4. [Frequency List of Written Finnish Word Forms](http://urn.fi/urn:nbn:fi:lb-20140730146)
5. [Names of Countries](https://metashare.ut.ee/repository/browse/names-of-countries/d0097290e7bc11e6a6e4005056b400246df1676946ee4140a9aedaf00e007aea/)
6. [The Conceptual File of Estonian Lexis of the Institute of Estonian Language](https://metashare.ut.ee/repository/browse/the-conceptual-file-of-estonian-lexis-of-the-institute-of-the-estonian-language/9003e1b6e7ae11e6a6e4005056b40024e93d428bb76f4299a9fffd6a2e13cec6/)
7. [Deutsche Wortschatz](http://corpora.informatik.uni-leipzig.de/de?corpusId=deu_newscrawl_2011)

### Language

29 wordlists are monolingual and represent the following 10 languages:

1. Dutch (2 resources)
2. Estonian (3 resources)
3. Finnish (9 resources)
4. German (1 resource)
5. Greek (1 resources)
6. Maltese (2 resources)
7. Ngbugu (1 resource)
8. Slovenian (5 resource)
9. Spanish (1 resource)
10. Swedish (4 resources)

26 wordlists are multilingual and represent the following 9 language combinations:

1. English, Greek (3 resources)
2. French, English (17 resources)
3. French, English, Spanish (2 resources)
4. Mangbetu, French, English
5. German, Greek, English
6. Greek, French (1 resources)
7. Greek, Italian, Turkish, English, German, French, Spanish, Castilian
8. Slovenian, Swedish, Bulgarian, English, Greek, Estonian, Spanish, Castilian, Czech, German, Danish, French, Finnish, Italian, Hungarian, Latvian, Lithuanian, Maltese, Dutch, Flemish, Portuguese, Polish, Slovak, Romanian
9. Swedish, Albanian, Bosnian, English, Finnish, Modern Greek, Croatian, Iranian Persian, Russian, Serbian, Somali, Spanish, Turkish

### Size

The largest resource is [INT Historical Word List](https://ivdnt.org/downloads/taalmaterialen/tstc-int-historische-woordenlijst), containing 500,000 words, while the smallest is [Topics of library and information science](http://hdl.grnet.gr/11500/IONION-0000-0000-25BD-2), which contains 732 words.

The following 7 wordlist lack information on size:

1. [Dictionary for the CST Lemmatizer](https://repository.clarin.dk/repository/xmlui/handle/20.500.12115/19)
2. [JRC-Names - a multilingual named entity resource](http://hdl.grnet.gr/11500/ATHENA-0000-0000-258E-7)
3. [Labial vibrants in Mangbetu: Archival form](https://hdl.handle.net/11403/sldr000829/v1)
4. [Names of Countries](https://metashare.ut.ee/repository/browse/names-of-countries/d0097290e7bc11e6a6e4005056b400246df1676946ee4140a9aedaf00e007aea/)
5. [The Conceptual File of Estonian Lexis of the Institute of Estonian Language](https://metashare.ut.ee/repository/browse/the-conceptual-file-of-estonian-lexis-of-the-institute-of-the-estonian-language/9003e1b6e7ae11e6a6e4005056b40024e93d428bb76f4299a9fffd6a2e13cec6/)

1. [Vocabulaire de Pathologies humaines](https://hdl.handle.net/11403/vocabulaire-de-pathologies-humaines/v1)

1. [Vocabulaire de Transferts de chaleur](https://hdl.handle.net/11403/transferts-de-chaleur/v1.2" \t "_parent)

### Linguistic information

The following 26 resources provide explicit information on the linguistic information included:

1. [Gos corpus n-grams 2.0](https://www.clarin.si/repository/xmlui/handle/11356/1195) (frequency)
2. [IMP corpus n-grams 2.0](http://hdl.handle.net/11356/1194) (frequency)
3. [Janes corpus n-grams 1.0](http://hdl.handle.net/11356/1192) (frequency)
4. [JRC-Names - a multilingual named entity resource](http://hdl.grnet.gr/11500/ATHENA-0000-0000-258E-7) (spelling varieties of names)
5. [Kres corpus n-grams 2.0](http://hdl.handle.net/11356/1193) (frequency)
6. [Maltese Fiction Wordlist](https://portulanclarin.net/repository/browse/maltese-fiction-wordlist/dd3d5a0849dc11e2a2aa782bcb074135d9cb4fe73e884aa3b771fd0a7a11a2b7/) (frequency)
7. [Modern Finnish Word List](http://urn.fi/urn:nbn:fi:lb-20140730169) (MSD-tags, lemmas)
8. [Vocabulaire d'archéologie](https://hdl.handle.net/11403/vocabulaire-darcheologie/v1) (paradigmatic lexical relations)
9. Vocabulaire d'art et archéologie (paradigmatic lexical relations)
10. [Vocabulaire de géographie de l'Amérique du Nord](https://hdl.handle.net/11403/vocabulaire-de-geographie-de-lamerique-du-nord/v1) (paradigmatic lexical relations)
11. [Vocabulaire de Nutrition artificielle](https://hdl.handle.net/11403/vocabulaire-de-nutrition-artificielle/v1) (paradigmatic lexical relations)
12. Vocabulaire de Pathologies humaines(paradigmatic lexical relations)
13. Vocabulaire de philosophie (paradigmatic lexical relations)
14. [Vocabulaire de préhistoire et protohistoire](https://hdl.handle.net/11403/prehistoire-protohistoire/v1.1) (paradigmatic lexical relations)
15. [Vocabulaire de Psychopathologie](https://hdl.handle.net/11403/vocabulaire-de-psychopathologie/v1) (paradigmatic lexical relations)
16. [Vocabulaire de sciences de la Terre](https://hdl.handle.net/11403/vocabulaire-de-sciences-de-la-terre/v1) (paradigmatic lexical relations)
17. [Vocabulaire de Sciences de l'éducation](https://hdl.handle.net/11403/vocabulaire-de-sciences-de-leducation/v1) (paradigmatic lexical relations)
18. [Vocabulaire de Sciences du langage](https://hdl.handle.net/11403/vocabulaire-de-sciences-du-langage/v1) (paradigmatic lexical relations)
19. [Vocabulaire de sociologie](https://hdl.handle.net/11403/vocabulaire-de-sociologie/v1) (paradigmatic lexical relations)
20. [Vocabulaire de Transferts de chaleur](https://hdl.handle.net/11403/transferts-de-chaleur/v1.2) (paradigmatic lexical relations)
21. [Vocabulaire de Transfusion sanguine](https://hdl.handle.net/11403/vocabulaire-de-transfusion-sanguine/v1) (paradigmatic lexical relations)
22. Vocabulaire d'Electronique et électro-optique (paradigmatic lexical relations)
23. Vocabulaire d'ethnologie (paradigmatic lexical relations)
24. [Vocabulaire d'histoire des sciences et des techniques](https://hdl.handle.net/11403/vocabulaire-dhistoire-et-sciences-de-la-litterature/v1) (paradigmatic lexical relations)
25. [Vocabulaire d'histoire et sciences de la littérature](https://hdl.handle.net/11403/vocabulaire-dhistoire-et-sciences-de-la-litterature/v1) (paradigmatic lexical relations)
26. [Vocabulaire d'Histoire et sciences des religions](https://hdl.handle.net/11403/vocabulaire-histoire-et-sciences-des-religions/v1) (paradigmatic lexical relations)

### Licence

The most common licence types are as follows:

1. CC-BY (49 wordlists)
2. CLARIN ACA (2 wordlists)

The following 4 resources lack information on licence:

1. [Lexical functions of Spanish verb-noun collocations](https://hdl.handle.net/11403/ortolang-000920/v1)
2. [Neologisms Online v3](http://tst.inl.nl/producten/neon_v3/versie1.1/displaywidx.php)

1. [Ngbugu digital wordlist: Archival form](https://hdl.handle.net/11403/sldr000833/v1" \t "_parent)

1. [Deutsche Wortschatz](http://corpora.informatik.uni-leipzig.de/de?corpusId=deu_newscrawl_2011)

### Overview of the wordlists in the CLARIN infrastructure

Table 5: CLARIN wordlists and their characteristics, sorted alphabetically by language

|  |  |  |  |
| --- | --- | --- | --- |
| **Resource** | **Language** | | **Description** |
| [Dictionary for the CST Lemmatizer](https://repository.clarin.dk/repository/xmlui/handle/20.500.12115/19" \t "_parent)  **Licence:** CLARIN-ACA-NC | | Danish | This wordlist was used for the [CST lemmatizer](https://cst.dk/online/lemmatiser/uk/) as a supplement to the lemmatisation rules. It is useful both for tagged and untagged input. The resource is available for download from CLARIN-DK. |
| [INT Historical Word List](https://ivdnt.org/downloads/taalmaterialen/tstc-int-historische-woordenlijst)  **Size:** 500,000 word forms  **Licence: o**ther | | Dutch | This wordlist includes historical lexemes for the period between 1550 and 1970. The resource is available for download from the Dutch Language Institute (INT). |
| [Neologisms Online v3](http://tst.inl.nl/producten/neon_v3/versie1.1/displaywidx.php)  **Size:** 19,000 words and expressions | | Dutch | This wordlist of neologisms is available for online browsing the Dutch Language Institute (INT). |
| [Topics of library and information science](http://hdl.grnet.gr/11500/IONION-0000-0000-25BD-2)  **Size:** 732 words  **Licence**: CC-BY-NC | | English, Greek | This is a word list of terms from the domain of library and information science. The resource is available for download from clarin:el. |
| [Estonian Frequency Dictionary (ver. 2.0)](https://metashare.ut.ee/repository/browse/estonian-frequency-dictionary-ver-20/07a87ad00f6f11e9a6e4005056b400241184c428448348b5a5654740a32e57bc/)  **Size:** 997,934 word forms  **Licence**: CLARIN PUB | | Estonian | This is a frequency list available for download from META-SHARE (CELR distribution) and for online browsing. |
| [Names of Countries](https://metashare.ut.ee/repository/browse/names-of-countries/d0097290e7bc11e6a6e4005056b400246df1676946ee4140a9aedaf00e007aea/)  **Licence:** CLARIN ACA | | Estonian | This is a wordlist that is based on the Estonian orthography of foreign place names. The resource is available for online browsing. |
| [The Conceptual File of Estonian Lexis of the Institute of Estonian Language](https://metashare.ut.ee/repository/browse/the-conceptual-file-of-estonian-lexis-of-the-institute-of-the-estonian-language/9003e1b6e7ae11e6a6e4005056b40024e93d428bb76f4299a9fffd6a2e13cec6/" \t "_parent)  **Licence:** CLARIN ACA | | Estonian | This is a controlled vocabulary of several more-and-less related concepts (e.g., gardening, haymaking, weather, fishing, religion). The resource is available for online browsing. |
| [Finnish Verbal Colorative Constructions](http://urn.fi/urn:nbn:fi:lb-2017090401)  **Size:** 61,617 words  **Licence:** CC-BY | | Finnish | This is a wordlist that contains Finnish verbal “colorative” (i.e., stylistically marked) constructions­. The resource is available for download through FIN-CLARIN. |
| [Frequencies of Early Modern Finnish Words](http://urn.fi/urn:nbn:fi:lb-20140730139)  **Size:** 4,862,190 words  **Licence:** EUPL | | Finnish | This is a frequency lexicon that consists of words from the Old Literary Finnish text corpus. The resource is available online through FIN-CLARIN. |
| [Frequencies of Old Literary Finnish Words](http://urn.fi/urn:nbn:fi:lb-20140730166)  **Size:** 3,425,382 words  **Licence:** EUPL | | Finnish | This is a frequency lexicon that is constituted of words from Old Literary Finnish text corpus. The resource is available online through FIN-CLARIN. |
| [Frequency Lexicon of the Finnish Newspaper Language](http://urn.fi/urn:nbn:fi:lb-201405272)  **Size:** 9,996 words  **Licence:** CC-BY NC ND 1.0 | | Finnish | This is a frequency lexicon available online through FIN-CLARIN. |
| [Frequency List of Written Finnish Word Forms](http://urn.fi/urn:nbn:fi:lb-20140730146)  **Size:** 17,604 lemmas; 1,339,787 word forms  **Licence:** EUPL | | Finnish | This is a frequency lexicon of Finnish word forms that appear in the Finnish Parole text corpus. The resource is available online through FIN-CLARIN. |
| [Modern Finnish Word List](http://urn.fi/urn:nbn:fi:lb-20140730169)  **Size:** 94,110 entries  **Linguistic information:** MSD-tags, lemmas  **Licence:** GNU LGPL; EUPL v.1.1; CC-BY SA 3.0 | | Finnish | This is a wordlist of contemporary general vocabulary that is available for download through FIN-CLARIN. |
| [Psycholinguistic Descriptives](http://urn.fi/urn:nbn:fi:lb-2018081601)  **Size:** 2.5 billion words  **Licence:** CC-BY 4.0 | | Finnish | This is a frequency wordlist (accompanied by a query tool) of acquiring commonly used psycholinguistic descriptives for Finnish words, and word surface form frequencies, lemma frequencies, syllable frequencies and letter n-gram frequencies. The resource is available for download from FIN-CLARIN. |
| [Relative frequencies of part-of-speech n-grams in native and translated Finnish literary prose](http://urn.fi/urn:nbn:fi:lb-2018120301)  **Licence:** CC-BY 4.0 | | Finnish | This is a frequency list of N-grams appearing in the corpus Classics of English and American Literature, English-Finnish parallel corpus and the corpus of Translated Finnish. The resource is available for download from FIN-CLARIN. |
| [The Finnish N-grams 1820-2000 of the Newspaper and Periodical Corpus of the National Library of Finland](http://urn.fi/urn:nbn:fi:lb-2014073038)  **Licence:** CC-BY 4.0 | | Finnish | This is a frequency list that contains sets of unigrams, bigrams and trigrams extracted from a newspaper corpus. The resource is available for download from FIN-CLARIN. |
| [Vocabulaire d'archéologie](https://hdl.handle.net/11403/vocabulaire-darcheologie/v1)  **Size:** 4,431 entries  **Linguistic information:** preferred form, synonym shape  **Licence:** CC-BY 4.0 | | French, English | This is a controlled vocabulary of expressions from the domain of archaeology available for download from ORTOLANG. |
| [Vocabulaire d'art et archéologie](https://hdl.handle.net/11403/art-et-archeologie/v1.1)  **Size:** 1,960 entries  **Linguistic information:** preferred forms, synonym shape  **Licence:** CC-BY 4.0 | | French, English | This is a controlled vocabulary of expressions from fine arts and archaeology available for download from Ortolang. |
| [Vocabulaire de géographie de l'Amérique du Nord](https://hdl.handle.net/11403/vocabulaire-de-geographie-de-lamerique-du-nord/v1)  **Size:** 4,232 entries  **Linguistic information:** hierarchical relationship, preferred form, synonym shape  **Licence:** CC-BY 4.0 | | French, English | This is a thesaurus of terms from the geography of North America available for download from ORTOLANG. |
| [Vocabulaire de Nutrition artificielle](https://hdl.handle.net/11403/vocabulaire-de-nutrition-artificielle/v1)  **Size:** 2,500 entries  **Linguistic information:** associative relationship, hierarchical relationship, preferred form, synonym shape  **Licence:** CC-BY 4.0 | | French, English | This is a controlled vocabulary of expressions from the domain of nutrition available for download from ORTOLANG. |
| [Vocabulaire de Pathologies humaines](https://hdl.handle.net/11403/vocabulaire-de-pathologies-humaines/v1)  **Linguistic information:** hierarchical relationship, preferred form, synonym shape  **Licence:** CC-BY 4.0 | | French, English | This is a controlled vocabulary of expressions from the domain of medicine (pathological diseases) available for download from ORTOLANG. |
| [Vocabulaire de philosophie](https://hdl.handle.net/11403/philosophie/v1.1)  **Size:** 4,435 entries  **Linguistic information:** preferred form, synonym shape  **Licence:** CC-BY 4.0 | | French, English | This is a controlled vocabulary of expressions of philosophical terms available for download from Ortolang. |
| [Vocabulaire de préhistoire et protohistoire](https://hdl.handle.net/11403/prehistoire-protohistoire/v1.1)  **Size:** 3,093 entries  **Linguistic information:** hierarchical relationship, preferred forms, synonym shape  **Licence:** CC-BY 4.0 | | French, English | This is a controlled vocabulary of expressions historical terms available for download from Ortolang. |
| [Vocabulaire de Psychopathologie](https://hdl.handle.net/11403/vocabulaire-de-psychopathologie/v1)  **Size:** 575 terms  **Linguistic information:** associative relationship, hierarchical relationship, preferred form, synonym shape  **Licence:** CC-BY 4.0 | | French, English | This is a controlled vocabulary of expressions from the domain of psychopathology available for download from Ortolang. |
| [Vocabulaire de Sciences de l'éducation](https://hdl.handle.net/11403/vocabulaire-de-sciences-de-leducation/v1)  **Size:** 2,681 entries  **Linguistic information:** preferred form, synonym shape  **Licence:** CC-BY 4.0 | | French, English | This is a controlled vocabulary of expressions from the domain of education available for download from ORTOLANG. |
| [Vocabulaire de Sciences du langage](https://hdl.handle.net/11403/vocabulaire-de-sciences-du-langage/v1)  **Size:** 6,142 entries  **Linguistic information:** hierarchical relationship, preferred form, synonym shape  **Licence:** CC-BY 4.0 | | French, English | This is a controlled vocabulary of expressions from the domain of linguistics available for download from ORTOLANG. |
| [Vocabulaire de sociologie](https://hdl.handle.net/11403/vocabulaire-de-sociologie/v1)  **Size:** 5,277 entries  **Linguistic information:** preferred form, synonym shape  **Licence:** CC-BY 4.0 | | French, English | This is a controlled vocabulary of expressions from the domain of sociology available for download from ORTOLANG. |
| [Vocabulaire de Transferts de chaleur](https://hdl.handle.net/11403/transferts-de-chaleur/v1.2)  **Linguistic information:** associative relationship, Hierarchical relationship, Preferred form, Synonym shape  **Licence:** CC-BY 4.0 | | French, English | This is a controlled vocabulary of expressions of thermodynamic terms available for download from Ortolang. |
| [Vocabulaire de Transfusion sanguine](https://hdl.handle.net/11403/vocabulaire-de-transfusion-sanguine/v1)  **Size:** 2,000 entries  **Linguistic information:** associative relationship, hierarchical relationship, preferred form, synonym shape  **Licence:** CC-BY 4.0 | | French, English | This is a controlled vocabulary of expressions from the domain of medicine (related to blood transfusion) available for download from ORTOLANG. |
| [Vocabulaire d'ethnologie](https://hdl.handle.net/11403/vocabulaire-dethnologie/v1)  **Size:** 9,517 entries  **Linguistic information:** preferred form, synonym shape  **Licence:** CC-BY 4.0 | | French, English | This is a controlled vocabulary of expressions from the domain of ethnology available for download from ORTOLANG. |
| [Vocabulaire d'histoire des sciences et des techniques](https://hdl.handle.net/11403/vocabulaire-dhistoire-et-sciences-de-la-litterature/v1)  **Size:** 3,766 entries  **Linguistic information:**  contextual information, usage examples,  preferred form, synonym shape  **Licence:** CC-BY 4.0 | | French, English | This is a controlled vocabulary of expressions from the domain technical sciences available for download from ORTOLANG. |
| [Vocabulaire d'histoire et sciences de la littérature](https://hdl.handle.net/11403/vocabulaire-dhistoire-et-sciences-de-la-litterature/v1)  **Size:** 11,065 entries  **Linguistic information:** preferred form, synonym shape  **Licence:** CC-BY 4.0 | | French, English | This is a controlled vocabulary of expressions from the domain of literary studies available for download from ORTOLANG. |
| [Vocabulaire d'Histoire et sciences des religions](https://hdl.handle.net/11403/vocabulaire-histoire-et-sciences-des-religions/v1)  **Size:** 4,581 entries  **Linguistic information:** preferred form, synonym shape  **Licence:** CC-BY 4.0 | | French, English | This is a controlled vocabulary of expressions from the domain of philosophy and religion available for download from ORTOLANG. |
| [Vocabulaire de sciences de la Terre](https://hdl.handle.net/11403/vocabulaire-de-sciences-de-la-terre/v1)  **Size:** 19,707 entries  **Linguistic information:** hierarchical relationship, preferred form, synonym shape  **Licence:** CC-BY 4.0 | | French, English, Spanish | This is a controlled vocabulary of expressions from the domain of geology available for download from ORTOLANG. |
| [Vocabulaire d'Electronique et électro-optique](https://hdl.handle.net/11403/vocabulaire-delectronique-et-electro-optique/v1)  **Size:** 4,456 entries  **Linguistic information:** associative relationship, hierarchical relationship, preferred form, synonym shape  **Licence:** CC-BY 4.0 | | French, English, Spanish | This is a controlled vocabulary of expressions from the domain of electronics available for download from Ortolang. |
| [Deutsche Wortschatz](http://corpora.informatik.uni-leipzig.de/de?corpusId=deu_newscrawl_2011)  **Size:** 5.8 million types  **Linguistic information:** synonymy, examples of use | | German | This resource provides a list of annotated words taken from the deu\_newscrawl\_2011 corpus. The resource is available for online browsing through CLARIN-D/University of Leipzig. |
| [Λεξικό Γλωσσολογικών όρων: Γερμανικά – Ελληνικά - Αγγλικά (lexicon of linguistic terms: DE-EL-EN)](http://hdl.grnet.gr/11500/AUTH-0000-0000-28CB-F)  **Size:** 2,000 words | | German, Greek, English | This is a wordlist of linguistic terms that is available for download from clarin:el. |
| [KELLY word-list Greek](http://hdl.grnet.gr/11500/ATHENA-0000-0000-25C1-C)  **Size:** 7,385 Entries  **Licence**: CC-BY-NC | | Greek | This wordlist is useful for learning and teaching Greek as a foreign/second language; the words are classified according to the language levels of CEFR. The resource is available for download from clarin:el. |
| [Maltese Fiction Wordlist](https://portulanclarin.net/repository/browse/maltese-fiction-wordlist/dd3d5a0849dc11e2a2aa782bcb074135d9cb4fe73e884aa3b771fd0a7a11a2b7/)  **Size:** 41,251 tokens  **Linguistic information:** frequency  **Licence:** MS-NC-No ReD | | Maltese | This is a wordlist from 32 fictional books available for download from CLARIN PORTULAN. |
| [Maltese Wordlist](https://portulanclarin.net/repository/browse/maltese-wordlist/1eb37910a35a11e1a404080027e73ea2dbdb15cf3d18412b943f9db788f20d5c/)  **Size:** 824,839 words  **Licence:** LGPL | | Maltese | This is a wordlist used for spell-checking. It is available for download from CLARIN PORTULAN. |
| [Labial vibrants in Mangbetu: Archival form](https://hdl.handle.net/11403/sldr000829/v1)  **Licence:** CC-BY | | Mangbetu, French, English | This is a wordlist of lexical items that exemplify occurrences of bilabial trills and the labiodental flaps. The resource is available for download from ORTOLANG. |
| [Ngbugu digital wordlist: Archival form](https://hdl.handle.net/11403/sldr000833/v1)  **Size:** 204 words | | Ngbugu | This is a wordlist used in language documentation, phonetics and lexicography. The resource is available for download from Ortolang. |
| [LCM-PL](http://hdl.handle.net/11321/585)  **Size:** 10,000 entries  **Licence:** CC BY SA 3.0 | | Polish | This is a wordlist that list the frequencies and abstract levels of verbs. The resource is available for download from the CLARIN-PL repository. |
| [Gos corpus n-grams 2.0](https://www.clarin.si/repository/xmlui/handle/11356/1195)  **Size:** 2,598,153 n-grams  **Linguistic information:** frequency  **Licence:** CC-BY-SA 4.0 | | Slovenian | This is a list of n-grams extracted from the [Gos corpus of spoken Slovene](http://eng.slovenscina.eu/korpusi/gos). The resource is available for download from CLARIN.SI |
| [IMP corpus n-grams 2.0](http://hdl.handle.net/11356/1194)  **Size:** 34,668,696 n-grams  **Linguistic information:** frequency  **Licence:** CC-BY-SA 4.0 | | Slovenian | This is a list of n-grams extracted from the [IMP corpus of historical Slovene](http://nl.ijs.si/imp/). The resource is available for download from CLARIN.SI. |
| [Janes corpus n-grams 1.0](http://hdl.handle.net/11356/1192)  **Size:** 351,029,703 n-grams  **Linguistic information:** frequency  **Licence:** CC-BY-SA 4.0 | | Slovenian | This is a list of n-grams extracted from [the Janes corpus of Slovenian user-generated content version 1.0](http://nl.ijs.si/janes/). The resource is available for download from CLARIN.SI |
| [Keywords and n-grams from a textbook corpus](http://hdl.handle.net/11356/1215)  **Size:** 5,977 words  **Licence:** CC-BY 4.0 | Slovenian | | This is a wordlist of keywords and n-grams extracted from a corpus of textbooks for Slovenian elementary and secondary schools available for download from CLARIN.SI. |
| [Kres corpus n-grams 2.0](http://hdl.handle.net/11356/1193)  **Size:** 211,104,769 n-grams  **Linguistic information:** frequency  **Licence:** CC-BY-SA 4.0 | | Slovenian | This is a list of n-grams extracted from the [Kres corpus of written Slovenian](http://eng.slovenscina.eu/korpusi/kres). The resource is available for download from CLARIN.SI |
| [JRC-Names - a multilingual named entity resource](http://hdl.grnet.gr/11500/ATHENA-0000-0000-258E-7)  **Linguistic information:** spelling varieties of names  **Licence**: Open for Reuse with Restrictions | | Slovenian, Swedish, Bulgarian, English, Greek, Estonian, Spanish, Castilian, Czech, German, Danish, French, Finnish, Italian, Hungarian, Latvian, Lithuanian, Maltese, Dutch, Flemish, Portuguese, Polish, Slovak, Romanian | This is a wordlist of named entities (person and organisation names). The resource is available for download from clarin:el. |
| [Lexical functions of Spanish verb-noun collocations](https://hdl.handle.net/11403/ortolang-000920/v1)  **Size:** 1,000 verb-noun pairs  **Linguistic information:** lexicological classifications (free-word combinations, errors), semantic information | | Spanish | This is a wordlist that consists of the most frequent 1000 verb-noun pairs extracted automatically from the Spanish Web Corpus classifications (collocation vs. free-word-combo). The resource is available for download from Ortolang. |
| [Idioms from the NEO lexicon DB](http://hdl.handle.net/10794/27)  **Size:** 4,928 entries  **Licence:** CC-BY 4.0 | | Swedish | This is a wordlist of idioms with explanations extracted from the database for the dictionary Nationalencyklopediens ordbok. The resource can be download from the SWE-CLARIN repository and can be queried online through KARP. |
| [Kelly (2017-10-16)](http://hdl.handle.net/10794/28)  **Size:** 10,510 entries  **Licence:** CC-BY 4.0 | | Swedish | This is a list of keywords for Language Learning for Young and adults alike. The resource can be download from the SWE-CLARIN repository and can be queried online through KARP. |
| [The Swedish N-grams 1770-1940 of the Newspaper and Periodical Corpus of the National Library of Finland](http://urn.fi/urn:nbn:fi:lb-2014091902)  **Licence:** CC-BY 4.0 | | Swedish | This frequency list contains sets of unigrams, bigrams and trigrams extracted from a corpus compiled by the University of Helsinki from the digitized newspapers from the National Library of Finland. The resource is available for download from FIN-CLARIN. |
| [Vocation list (2015-01-10)](http://hdl.handle.net/10794/48)  **Size:** 13,833 entries  **Licence:** CC-BY 4.0 | | Swedish | This is a wordlist of vocations in Swedish. The resource can be download from the SWE-CLARIN repository and can be queried online through KARP. |
| [Swedish words, LEXIN](http://hdl.handle.net/10794/40)  **Size:** 29,111 entries  **Licence:** CC-BY 4.0 | | Swedish, Albanian, Bosnian, English, Finnish, Modern Greek, Croatian, Iranian Persian, Russian, Serbian, Somali, Spanish, Turkish | This is a word list to be used by immigrants to Sweden. The resource can be download from the SWE-CLARIN repository and can be queried online through KARP. |

# Conclusions

One of the most important findings of the present survey is that the CLARIN infrastructure offers not only a large number of lexical resources (290) but also that they cover a lot of languages. The monolingual resources cover 26 languages (Arabic, Croatian, Czech, Dagaare, Danish, Dutch, English, Estonian, Finnish, French, Frisian, German, Greek, Icelandic, Italian, Latvian, Maltese, Ngbugu, Polish, Portuguese, Serbian, Slovenian, Swedish), whereas the multilingual corpora account for more than 50 diverse language combinations, often including non-Indo-European languages.

The metadata of the lexical resources in the CLARIN infrastructure are generally well-described. Information on size is provided for nearly all the resources – 278 (96%) out of 290. Information on licence type is also readily included and is available for 262 (90%) out of 290 resources. It is noteworthy that 18 out of 28 resources with missing information on licence are Latvian dictionaries, which are not made available through a certified repository.

The type of linguistic information included in the resources e.g., morphological features of a dictionary entry; the types of lexical relations included in wordnets) is, by contrast, generally seldom provided, at least in an explicit way as part of the resource description. Only about 35% out of the 290 resource provide such information in an explicit manner, i.e., as part of the metadata, though in some cases this information can be inferred from e.g., the title of the resource. Sometimes, the information that *is* provided is vague – e.g., [Multilingual lexicon of toponyms](http://hdl.grnet.gr/11500/ATHENA-0000-0000-23C1-E) lists the annotation layer “semantics – semantic class”, but does not further specify what semantic class pertains to.

From a lexicographic standpoint, there seems to be a lack of consensus on the typology of the lexical resources in the CLARIN infrastructure. For example, [Topics of library and information science](http://hdl.grnet.gr/11500/IONION-0000-0000-25BD-2), where the description field in the clarin:el repository lists it as a glossary, but the metadata list it as a wordlist.

A similar issue is related to the conventions used for describing the size of the corpora. First, the descriptors *lexical entry*, *keyword*, *lemma* and *word* all seem to be in use for the same type of lexical resources. Second, certain resources use unclear descriptors, e.g., *lexical types*. Finally, certain resources do not list the size of the resource in terms of lexical entries but in terms of byte size, e.g., MBs or GBs (not included in this overview).

Finally, a small number (19 out of 290) of the surveyed resources cannot be found in the VLO, including 8 resources from ARCHE (cf. the VLO sections under *Lexica*and *Dictionaries*), which is a CLARIN-B Centre.

1. The categories of the resources in the spreadsheet were suggested by the contributors to the survey and may differ from this report. [↑](#footnote-ref-1)