

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
2. LEXICAL RESOURCES IN THE CLARIN INFRASTRUCTURE	2
2.1. LEXICA	3
2.1.1. VLO	3
2.1.2. AVAILABILITY	3
2.1.3. LANGUAGE	5
2.1.4. SIZE	5
2.1.5. LINGUISTIC INFORMATION	6
2.1.6. LICENCE	6
2.1.7. OVERVIEW OF THE LEXICA IN THE CLARIN INFRASTRUCTURE	7
2.2. DICTIONARIES	16
2.2.1. VLO	16
2.2.2. AVAILABILITY	16
2.2.3. LANGUAGE	18
2.2.4. SIZE	19
2.2.5. LINGUISTIC INFORMATION	19
2.2.6. LICENCE	20
2.2.7. OVERVIEW OF THE DICTIONARIES IN THE CLARIN INFRASTRUCTURE	20
2.3. CONCEPTUAL RESOURCES	30
2.3.1. VLO	30
2.3.2. AVAILABILITY	30
2.3.3. LANGUAGE	30
2.3.4. SIZE	31
2.3.5. LINGUISTIC INFORMATION	31
2.3.6. LICENCE	31
2.3.7. OVERVIEW OF THE CONCEPTUAL RESOURCES IN THE CLARIN INFRASTRUCTURE	32
2.4. GLOSSARIES	36
2.4.1. VLO	36
2.4.2. AVAILABILITY	36
2.4.3. LANGUAGE	37
2.4.4. SIZE	37
2.4.5. LINGUISTIC INFORMATION	37
2.4.6. LICENCE	38
2.4.7. OVERVIEW OF THE GLOSSARIES IN THE CLARIN INFRASTRUCTURE	38
2.5. WORDLISTS	42
2.5.1. VLO	42
2.5.2. AVAILABILITY	42
2.5.3. LANGUAGE	43
2.5.4. SIZE	44
2.5.5. LINGUISTIC INFORMATION	44
2.5.6. LICENCE	44
2.5.7. OVERVIEW OF THE WORDLISTS IN THE CLARIN INFRASTRUCTURE	45
3. CONCLUSIONS	52

1. 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:

- (i) by manually searching the national repositories with the following keywords: “lexicon”, “dictionary”, “glossary”, “wordlist”
- (ii) 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](#).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.

2. 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](#) (76 resources)
- [2] [Dictionaries](#) (94 resources)
- [3] [Conceptual resources](#) (29 resources)
- [4] [Glossaries](#) (33 resources)
- [5] [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.

¹ The categories of the resources in the spreadsheet were suggested by the contributors to the survey and may differ from this report.

2.1. Lexica

There are 76 lexica available in the CLARIN infrastructure. They are presented in detail in Table 1.

2.1.1. VLO

All the lexica in in Table 1 can be found in the VLO expect for the following 4:

- [1] [A machine-readable dictionary of Egyptian Arabic](#)
- [2] [A machine-readable Persian-English dictionary](#)
- [3] [A machine-readable glossary of Egyptian Arabic](#)
- [4] [A machine-readable Persian-English glossary of verbs](#)

2.1.2. Availability

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

- [1] [Beseda Corpus Lemmatisation Lexicon](#)
- [2] [CzEngClass 0.2](#)
- [3] [CzEngVallex](#)
- [4] [Database of Modern Icelandic Inflections](#)
- [5] [DeriNet 1.6](#)
- [6] [Emoji Sentiment Ranking 1.0](#)
- [7] [EngVallex](#)
- [8] [Lexicon of historical Slovene imp25k 1.1](#)
- [9] [Morphological lexicon Sloleks 2.0](#)
- [10] [Old Swedish morphology \(2017-10-16\)](#)
- [11] [Old Swedish morphology \(2017-10-16\)](#)
- [12] [Parole+ \(2017-10-16\)](#)
- [13] [PDT-Vallex](#)
- [14] [SALDO's morphology \(2017-10-16\)](#)
- [15] [Simple lexicon](#)
- [16] [The Database of Estonian Multi-Word Expressions](#)
- [17] [VALLEX 3.0](#)
- [18] [Emotional Annotations Dictionary](#)

The following 50 lexica are available for download only:

- [1] [A machine-readable dictionary of Egyptian Arabic](#)
- [2] [A machine-readable glossary of Egyptian Arabic](#)
- [3] [A machine-readable Persian-English dictionary](#)
- [4] [A machine-readable Persian-English glossary of verbs](#)
- [5] [Automatically constructed multiword lexicon hrMWElex v0.5](#)
- [6] [Automatically constructed multiword lexicon sIMWElex v0.5](#)
- [7] [Automatically constructed multiword lexicon srMWElex v0.5](#)
- [8] [Automatically stress labelled morphological lexicon Sloleks 1.2, version 1.1](#)
- [9] [Basilex Lexicon](#)
- [10] [Basiscript Lexicon](#)
- [11] [BioLexicon](#)
- [12] [CombiLex](#)
- [13] [Concreteness and imageability lexicon MEGA.HR-Crossling](#)
- [14] [Démonette](#)
- [15] [Dicovalence](#)

- [16] [Dutch Electronic Lexicon of Multiword Expressions](#)
- [17] [Extended dictionary of named entities NElexicon connected with Linked Open Data](#)
- [18] [ILSP PsychoLinguistic Resource](#)
- [19] [Italian Content Words v3](#)
- [20] [Italian Function Words v3](#)
- [21] [LEX-MWE-PT: Word Combination in Portuguese Language](#)
- [22] [LX-Rare Word Similarity Dataset](#)
- [23] [LX-SimLex-999](#)
- [24] [LX-WordSim-353](#)
- [25] [MarsaLex](#)
- [26] [MorFlex CZ](#)
- [27] [Morphalou](#)
- [28] [MULTEXT-East free lexicons 4.0](#)
- [29] [MULTEXT-East non-commercial lexicons 4.0](#)
- [30] [Multifunctional Computational Lexicon of Contemporary Portuguese](#)
- [31] [MWElexicon 1.1](#)
- [32] [OMBI Arabic-Dutch](#)
- [33] [OMBI Dutch-Arabic](#)
- [34] [OMBI Dutch-Danish](#)
- [35] [OMBI Dutch-Indonesian](#)
- [36] [OpeNER Sentiment Lexicon Italian - LMF](#)
- [37] [ParaDi 2.0](#)
- [38] [PAROLE Lexicon](#)
- [39] [QTLeap specialized lexicons](#)
- [40] [Reference Lexicon for Belgian-Dutch \(RBBN\)](#)
- [41] [Reference Lexicon for Dutch](#)
- [42] [Slovene sentiment lexicon JOB 1.0](#)
- [43] [Slovene sentiment lexicon KSS 1.1](#)
- [44] [STO morphology \(v2\) - LMF format](#)
- [45] [STO syntax \(v2\) - LMF format](#)
- [46] [VfrLPL](#)
- [47] [Walenty \(2018-06-29\)](#)
- [48] [Word embeddings CLARIN.SI-embed.hr 1.0](#)
- [49] [Word embeddings CLARIN.SI-embed.sl 1.0](#)
- [50] [Word embeddings CLARIN.SI-embed.sr 1.0](#)

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

- [1] [LX-Abbreviations](#)
- [2] [LX-DSemVectors](#)
- [3] [LX-StopWords](#)
- [4] [Maltese Speech Engine Lexicon](#)
- [5] [PAROLE Portuguese Lexicon](#)
- [6] [Porlex](#)
- [7] [Simple Portuguese Lexicon](#)

2.1.3. 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)

2.1.4. 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](#) (3.6 million entries), while the smallest is [Emoji Sentiment Ranking 1.0](#) (751 entries).

The following 2 lexica lack information on size:

- [1] [Basilex Lexicon](#)
- [2] [Basiscript Lexicon](#)

2.1.5. 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](#)
- [2] [Basiscript Lexicon](#)
- [3] [BioLexicon](#)
- [4] [CombiLex](#)
- [5] [DeriNet 1.6](#)
- [6] [Dutch Electronic Lexicon of Multiword Expressions](#)
- [7] [Emotional Annotations Dictionary](#)
- [8] [Extended dictionary of named entities NELexicon connected with Linked Open Data](#)
- [9] [Italian Content Words v3](#)
- [10] [Italian Function Words v3](#)
- [11] [MarsaLex](#)
- [12] [Name Component Lexicon](#)
- [13] [Old Swedish morphology \(2017-10-16\)](#)
- [14] [Old Swedish morphology \(2017-10-16\)](#)
- [15] [OMBI Arabic-Dutch](#)
- [16] [OMBI Dutch-Arabic](#)
- [17] [OMBI Dutch-Danish](#)
- [18] [OMBI Dutch-Indonesian](#)
- [19] [Parole+ \(2017-10-16\)](#)
- [20] [PAROLE-SIMPLE-CLIPS](#)
- [21] [QTLeap specialized lexicons](#)
- [22] [Reference Lexicon for Belgian-Dutch \(RBBN\)](#)
- [23] [SALDO's morphology \(2017-10-16\)](#)
- [24] [Simple lexicon](#)
- [25] [Simple Portuguese Lexicon](#)
- [26] [STO morphology \(v2\) - LMF format](#)
- [27] [STO syntax \(v2\) - LMF format](#)
- [28] [The Database of Estonian Multi-Word Expressions](#)
- [29] [Walenty \(2018-06-29\)](#)

2.1.6. Licence

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

- [1] [VfrLPL](#)

The most common licence types are as follows:

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

2.1.7. 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 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 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 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 . The resource is available for download from ARCHE.
A machine-readable glossary of Egyptian Arabic 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 project. The resource is available for download from ARCHE.
OMBI Dutch-Arabic 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 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 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	Croatian	This lexicon contains word embeddings extracted from the

<p>Size: 3,147,352 entries Linguistic information: PoS-tags, lemmas Licence: CC-BY 4.0</p>		<p>Croatian web corpus hrWaC and a 400-million-token-heavy collection of newspaper texts. The resource is available for download from CLARIN.SI.</p>
<p>DeriNet 1.6 Size: 1,027,832 entries Licence: CC-BY-NC-SA 3.0</p>	Czech	<p>This is a lexicon of derivational relations (both compounding and inflections). The resource is available for download and online browsing through LINDAT.</p>
<p>MorfFlex CZ Size: 124,259,099 lexical types Linguistic information: MSD-tags, derivational, semantic, NER information Licence: CC-BY-NC-SA 3.0</p>	Czech	<p>This is a morphological lexicon available for download from LINDAT.</p>
<p>ParaDi 2.0 Size: 1,621 entries Linguistic information: MSD-tags, syntactic/semantic features Licence: CC-BY 4.0</p>	Czech	<p>This is a lexicon of single-word paraphrases of Czech verbal multiword expressions. The resource is available for download from LINDAT.</p>
<p>PDT-Vallex Size: 7,121 entries, 11,933 frames Linguistic information: verb, adjective and noun valency Licence: CC-BY-NC-SA 4.0</p>	Czech	<p>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.</p>
<p>VALLEX 3.0 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</p>	Czech	<p>This is a valency lexicon available for download and online browsing through LINDAT.</p>
<p>CzEngClass 0.2 Size: 200 classes, 3,525 entries Linguistic information: valency and synonymy Licence: CC-BY-NC-SA 4.0</p>	Czech, English	<p>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.</p>
<p>CzEngVallex Size: 20,835 pairs (verb senses) Linguistic information: verb valency Licence: CC-BY-NC-SA 4.0</p>	Czech, English	<p>This is a valency lexicon linked to the parallel PCEDT corpus. The resource is available for download and online browsing through LINDAT.</p>
<p>STO morphology (v2) - LMF format Size: 87,209 entries Licence: CC BY-SA 4.0</p>	Danish	<p>This morphological lexicon is available for download from the CLARIN-DK repository. It is also available in the .csv format.</p>

STO syntax (v2) - LMF format 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	Dutch	This is a lexicon that comprises all lemmas from the 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 Licence: other	Dutch	This is a lexicon that comprises all lemmas from the 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 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 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 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) 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 Size: 50,000 lemmas Linguistic information: dialectal information Licence: other	Dutch	This is a corpus-based monolingual lexicon available for download the Dutch Language Institute (INT).
OMBI Arabic-Dutch 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).

<p>OMBI Dutch-Danish</p> <p>Size: 46,000 entries Licence: other</p>	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).
<p>OMBI Dutch-Indonesian</p> <p>Size: 50,000 entries Licence: other</p>	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).
<p>BioLexicon</p> <p>Size: over 2.2 million entries (over 3.3 million semantic relations) Licence: ELRA END USER</p>	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.
<p>EngVallex</p> <p>Size: 4,337 entries, 7,148 frames Linguistic information: verb valency Licence: CC-BY-NC-SA 4.0</p>	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.
<p>QTLeap specialized lexicons</p> <p>Size: 231,516 entries Licence: CC-BY</p>	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.
<p>The Database of Estonian Multi-Word Expressions</p> <p>Size: 12,500 words Licence: proprietary</p>	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.
<p>Démonette</p> <p>Size: 96,027 entries Linguistic information: MSD-tags (grace format), semantic types Licence: CC-BY 4.0</p>	French	This is a morphological lexicon available for download from ORTOLANG.
<p>Dicovalence</p> <p>Size: 8,000 entries Linguistic information: c- and s-selectional restrictions, Licence: Licence Publique Générale Amoudrie GNU</p>	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 Size: 595,000,000 inflected forms Licence: CC-BY 4.0	French	This is a morphological lexicon available for download from ORTOLANG.
Morphalou Size: 159,261 entries Linguistic information: spelling, phonetics, mood, tense, MSD-tags, spelling variant, feminine variation, pronominal Licence: Publique Générale Amondrie GNU	French	This is a morphological lexicon available for download from ORTOLANG.
VfrLPL 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 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 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 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 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 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 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	Macedonian, Persian, Polish, Russian, Serbian	These are morphological lexica available for download from the CLARIN.SI repository.

<p>Size: 2,288,228 entries Linguistic information: MSD-tags, lemmas Licence: CC-BY-NC 4.0</p>		
<p>Maltese Speech Engine Lexicon</p> <p>Size: 39,242 entries Linguistic information: PoS-tags, orthographic transcription, phonetic forms, syllables, stress position Licence: MS-BY-NC-SA</p>	Maltese	This is a speech lexicon that is useful for building speech-to-text systems. It is available for download from CLARIN PORTULAN.
<p>A machine-readable Persian-English dictionary</p> <p>Size: 1,892 entries Licence: CC-BY-NC-SA 3.0 Linguistic information: morphological information, usage examples</p>	Persian-English	This bilingual lexicon has been compiled for comparative as well as didactic purposes in the ongoing VICAV project. The resource is available for download from ARCHE.
<p>A machine-readable Persian-English glossary of verbs</p> <p>Size: 429 entries Linguistic information: basic morphological information Licence: CC-BY-NC-SA 3.0</p>	Persian-English	This lexicon of single-word verbs in Modern Persian is available for download from ARCHE.
<p>Emotional Annotations Dictionary</p> <p>Size: 178,514 elements Licence: CC-BY 4.0</p>	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.
<p>Extended dictionary of named entities NEXicon connected with Linked Open Data</p> <p>Size: 103,585 entries Licence: GNU LGPL 3.0</p>	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.
<p>MWElexicon 1.1</p> <p>Size: 56,500 lexical units Linguistic information: “syntactic behaviour” Licence: plWordNet</p>	Polish	This is a lexicon of multiword expressions available for download from CLARIN.PL.
<p>Walenty (2018-06-29)</p> <p>Size: 18,236 entries Licence: CC BY SA 4.0</p>	Polish	This is a lexicon of verb valency that is available for download from the CLARIN-PL repository.
<p>LEX-MWE-PT: Word Combination in Portuguese Language</p> <p>Size: 1,198 entries 12,753 multi word unit Linguistic information: lemmas</p>	Portuguese	This is a lexicon of multiword expressions available for download from CLARIN PORTULAN.

Licence: MS NC-NoReD-ND		
LX-Abbreviations 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 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 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 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 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 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 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 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 Size: 27,374 words	Portuguese	This is a lexicon that provides psycholinguistic and cognitive information that is useful to select stimulus materials for experiments

<p>Linguistic information: orthographic and phonological/phonetic transcriptions, phonetic, MSD-tags, and frequency information Licence: MS NC-NoReD-ND</p>		and/or training vocabularies. The resource is available for download from CLARIN PORTULAN.
<p>Simple Portuguese Lexicon Size: 10,438 entries Linguistic information: qualia structure, semantic relations (hyponymy, synonymy, etc.) Licence: MS-BY-NC-SA</p>	Portuguese	This semantic lexicon is available for download from CLARIN PORTULAN.
<p>Automatically constructed multiword lexicon srMWElex v0.5 Size: 22,290 entries Linguistic information: MWEs Licence: CC-BY 4.0</p>	Serbian	This is a lexicon of multiword expressions available for download from CLARIN.SI.
<p>Word embeddings CLARIN.SI-embed.sr 1.0 Size: 1,480,566 entries Linguistic information: PoS-tags, lemmas Licence: CC-BY 4.0</p>	Serbian	The lexicon contains word embeddings from the srWaC web corpus . The resource is available for download from CLARIN.SI.
<p>Automatically constructed multiword lexicon slMWElex v0.5 Size: 47,579 entries Linguistic information: MWEs Licence: CC-BY 4.0</p>	Slovenian	This is a lexicon of multiword expressions available for download from CLARIN.SI.
<p>Automatically stress labelled morphological lexicon Sloleks 1.2, version 1.1 Size: 100,805 entries; 2,774,745 words Linguistic information: wordforms, PoS-tags, lemmas, frequency, prosody Licence: CC-BY-NC-SA 4.0</p>	Slovenian	This is an extended version of the morphological lexicon Sloleks 1.2 with added information about the stress of each word form. The resource is available for download from CLARIN.SI.
<p>Beseda Corpus Lemmatisation Lexicon Size: 3,228,127 entries Linguistic information: wordforms, PoS-tags, lemmas, frequency Licence: CC-BY 4.0</p>	Slovenian	This lexicon contains inflected open class words from the Dictionary of Standard Slovenian that are augmented by wordforms, their part of speech tags and their lemmas used during the PoS tagging and lemmatization of the Beseda corpus . The resource is available for download from CLARIN.SI and for online browsing.
<p>Lexicon of historical Slovene imp25k 1.1</p>	Slovenian	This is a morphological lexicon available for download from

<p>Size: 28,034 entries Linguistic information: MSD-tags, lemmas, etymological glosses Licence: CC-BY 4.0</p>		CLARIN.SI and for online browsing through a dedicated environment.
<p>Morphological lexicon Sloleks 2.0</p> <p>Size: 100,805 entries Linguistic information: wordforms, PoS-tags, lemmas, frequency, phonology Licence: CC-BY-NC-SA 4.0</p>	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.
<p>Slovene sentiment lexicon JOB 1.0</p> <p>Size: 25,524 entries Linguistic information: sentiment tags Licence: CC-BY-S15A 4.0</p>	Slovenian	This is a lexicon of sentiment labels available for download from the CLARIN.SI repository.
<p>Slovene sentiment lexicon KSS 1.1</p> <p>Size: 90,620 lexica Linguistic information: lemmas, sentiment tags Licence: CC-BY 4.0</p>	Slovenian	This is a lexicon of sentiment labels available for download from the CLARIN.SI repository.
<p>Word embeddings CLARIN.SI-embed.sl 1.0</p> <p>Size: 4,560,444 entries Linguistic information: PoS-tags, lemmas Licence: CC-BY 4.0</p>	Slovenian	This is a lexicon of word embeddings that is available for download from CLARIN.SI.
<p>Old Swedish morphology (2017-10-16)</p> <p>Size: 41,958 entries Licence: CC-BY 4.0</p>	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.
<p>Parole+ (2017-10-16)</p> <p>Size: 24,523 entries Licence: CC-BY 4.0</p>	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.
<p>SALDO's morphology (2017-10-16)</p> <p>Size: 128,036 entries Licence: CC-BY 4.0</p>	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.
<p>Simple lexicon</p> <p>Size: 11,624 entries Licence: CC-BY 4.0</p>	Swedish	This is a semantic lexicon that is available for download from the SWE-CLARIN repository and can be queried online through KARP.

2.2. Dictionaries

There are 94 dictionaries available in the CLARIN infrastructure. They are presented in detail in Table 2.

2.2.1. VLO

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

- [1] [A digital dictionary of Dagaare](#)
- [2] [A digital dictionary of Damascus Arabic](#)
- [3] [A digital dictionary of Modern Standard Arabic](#)
- [4] [A Digital Dictionary of Tunis Arabic](#)
- [5] [Dictionary of Modern Latvian \(MLVV\)](#)
- [6] [English-Latvian Dictionary](#)
- [7] [Historical Dictionary of Latvian \(16th-17th Centuries\) \(LVVV\)](#)
- [8] [Latvian-Latgalian Dictionary](#)
- [9] [Latvian-Russian Dictionary](#)
- [10] [Lithuanian-Latvian Dictionary](#)
- [11] [Lithuanian-Latvian-Latgalian Dictionary \(LLL\)](#)
- [12] [Mühlenbach Endzelin Latvian Dictionary \(MEV\)](#)
- [13] [Rendering Dictionary of Personal Names](#)
- [14] [Russian-Latvian Dictionary](#)

2.2.2. Availability

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

- [1] [Comprehensive Estonian-French Dictionary](#)
- [2] [Dictionary of New Slovenian Words](#)
- [3] [Dictionary of Slovenian Particles](#)
- [4] [Dictionary of Slovenian Phrasemes](#)
- [5] [Dictionary of the Slovenian Language in the Works of Janez Svetokriški](#)
- [6] [First Estonian Slang Dictionary](#)
- [7] [ISLEX](#)
- [8] [Slovenian-German Dictionary of Maks Pleteršnik \(1894-1895\)](#)
- [9] [A Digital Dictionary of Tunis Arabic](#)
- [10] [A digital dictionary of Dagaare](#)
- [11] [Dalin's Dictionary \(2017-10-16\)](#)
- [12] [Hellquist's Swedish etymology \(2017-10-16\)](#)
- [13] [Dictionary of Twitterese Janes-Dict 1.0](#)
- [14] [Japanese-Slovene learner's dictionary jaSlo 3.1](#)

The following 17 dictionaries are available for download only:

- [1] [A digital dictionary of Damascus Arabic](#)
- [2] [A digital dictionary of Modern Standard Arabic](#)
- [3] [Dictionary of Contemporary Dutch \(ANW\)](#)
- [4] [Dictionary of Lesser Used Slovenian Words](#)
- [5] [English-Greek Dictionary of False Friends](#)
- [6] [French-Greek lexicon of idioms](#)
- [7] [Jysk Ordbog \(a-h\), 17. April, 2011, xml-edition](#)
- [8] [Kven-Norwegian Bokmål dictionary](#)
- [9] [Multilingual dictionaries](#)
- [10] [North Saami-Finnish dictionary](#)
- [11] [Norwegian Bokmål-North Saami dictionary](#)
- [12] [Norwegian Bokmål-South Saami dictionary](#)

- [13] [Norwegian-Vietnamese Dictionary](#)
- [14] [Pronunciation Dictionary](#)
- [15] [Reverse dictionary of Slovenian language](#)
- [16] [Swedish-South Saami dictionary](#)
- [17] [Words of the 16th-Century Slovenian Literary Language](#)

The following 61 dictionaries are available only for online browsing:

- [1] [Dictionaries of the CLARIN-PL Lexical Platform](#)
- [2] [Dictionary of Contemporary Finnish](#)
- [3] [Dictionary of Finnish Dialects](#)
- [4] [Dictionary of Middle Dutch \(MNW\)](#)
- [5] [Dictionary of Modern Latvian \(MLVV\)](#)
- [6] [Dictionary of Old Dutch \(ONW\)](#)
- [7] [Dictionary of Old Literary Finnish](#)
- [8] [Dictionary of Standard Latvian Language \(LLVV\)](#)
- [9] [Digital Dictionary of the German Language \(DWDS\)](#)
- [10] [Early Middle Dutch Dictionary \(VMNW\)](#)
- [11] [English Hask Collocation Database](#)
- [12] [English-Estonian Machine Translation Dictionary](#)
- [13] [English-Latvian Dictionary](#)
- [14] [Estonian-Latvian Dictionary](#)
- [15] [Estonian-Russian Learner's Orthographic Dictionary](#)
- [16] [Etymological Database of the Saami Languages](#)
- [17] [Finnish-Estonian Dictionary](#)
- [18] [Historical Dictionary of Latvian \(16th-17th Centuries\) \(LVVV\)](#)
- [19] [iWNT](#)
- [20] [Karelian Dictionary](#)
- [21] [Latvian-English Dictionary](#)
- [22] [Latvian-Latgalian Dictionary](#)
- [23] [Latvian-Lithuanian Dictionary](#)
- [24] [Latvian-Russian Dictionary](#)
- [25] [Lithuanian-Latvian Dictionary](#)
- [26] [Lithuanian-Latvian-Latgalian Dictionary \(LLL\)](#)
- [27] [Livonian-Estonian-Latvian Dictionary](#)
- [28] [Low-German Loanwords in the Estonian Language](#)
- [29] [Mühlenbach Endzelin Latvian Dictionary \(MEV\)](#)
- [30] [Online Dictionary of Standard Estonian ÕS 2018](#)
- [31] [Online Lexicon of Veps Language](#)
- [32] [Polish Hask Collocation Database](#)
- [33] [Rendering Dictionary of Personal Names](#)
- [34] [Russian-Latvian Dictionary](#)
- [35] [Tezaurs.lv](#)
- [36] [The Advisory Dictionary of Document Language](#)
- [37] [The Basic Estonian Dictionary](#)
- [38] [The Dialectological Dictionary of Estonian](#)
- [39] [The Dictionary of Estonian Antonyms](#)
- [40] [The Dictionary of Estonian Dialects \(2013-...\)](#)
- [41] [The Dictionary of Estonian Phraseology](#)
- [42] [The Dictionary of Estonian Synonyms](#)
- [43] [The Dictionary of Estonian Word Families - Index](#)
- [44] [The Dictionary of Foreign Words](#)
- [45] [The Dictionary of Old Võru Dialect](#)
- [46] [The Dictionary of the Eastern Dialect](#)

- [47] [The Dictionary of the Hiiu Dialect](#)
- [48] [The Dictionary of the Kihnu Dialect](#)
- [49] [The Dictionary of the Mulgi Dialect](#)
- [50] [The Dictionary of the Seto Dialect](#)
- [51] [The Estonian Etymological Dictionary](#)
- [52] [The Estonian-Latvian-Estonian Dictionary](#)
- [53] [The Estonian-Mari Dictionary](#)
- [54] [The Estonian-Norwegian-Estonian Dictionary](#)
- [55] [The Estonian-Russian Dictionary](#)
- [56] [The Estonian-Udmurt Dictionary](#)
- [57] [The Explanatory Dictionary of Estonian Language](#)
- [58] [The Livonian-German Dictionary](#)
- [59] [The Online Dictionary of Estonian Sign Language - Estonian](#)
- [60] [The Russian-Estonian Dictionary](#)
- [61] [Wurdboek fan de Frsyke taal](#)

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](#) is broken).

- [1] [Basic English-Maltese Dictionary](#)

2.2.3. *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

2.2.4. *Size*

The following 3 dictionaries lack information on size:

- [1] [Multilingual dictionaries](#)
- [2] [Online Lexicon of Veps Language](#)
- [3] [The Dictionary of Estonian Dialects \(2013-...\)](#)

2.2.5. *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](#) (PoS-tags, nominal and verbal valency)
- [2] [Digital Dictionary of the German Language \(DwDs\)](#) (morphosyntactic information, pronunciation, etymology, lexical relations, examples of use, typical word connections)
- [3] [ISLEX](#) (phonology and MSD-tags)
- [4] [Japanese-Slovene learner's dictionary jaSlo 3.1](#) (headwords in all Japanese scripts, PoS-tags, difficulty level, Slovene glosses, Slovenian translation equivalents, examples of use)
- [5] [Multilingual dictionaries](#) (MSD-tags, semantic information, multi-word units)
- [6] [Norwegian-Vietnamese Dictionary](#) (phonetic transcription, PoS-tags, translation and example sentences)
- [7] [Online Dictionary of Standard Estonian ÕS 2018](#) (standardised orthography, phonology, MSD-tags)
- [8] [Pronunciation Dictionary](#) (IPA and SAMPA phonetic transcription)
- [9] [The Dialectological Dictionary of Estonian](#) (geographical distribution)
- [10] [The Dictionary of Estonian Dialects \(2013-...\)](#) (geographical distribution; phonetic/phonological markup)
- [11] [The Dictionary of Foreign Words](#) (orthography, MSD-tags, semantics and word etymology)
- [12] [The Dictionary of the Seto Dialect](#) (dialect orthography and simplified spelling)
- [13] [The Estonian Etymological Dictionary](#) (word etymology)
- [14] [Words of the 16th-Century Slovenian Literary Language](#) (PoS tags, phonological features)

2.2.6. 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 \(ANW\)](#)
- [2] [Dictionary of Middle Dutch \(MNW\)](#)
- [3] [Dictionary of Modern Latvian \(MLVV\)](#)
- [4] [Dictionary of Standard Latvian Language \(LLVV\)](#)
- [5] [Digital Dictionary of the German Language \(DWDS\)](#)
- [6] [English-Latvian Dictionary](#)
- [7] [Estonian-Latvian Dictionary](#)
- [8] [Historical Dictionary of Latvian \(16th-17th Centuries\) \(LVVV\)](#)
- [9] [Latvian-English Dictionary](#)
- [10] [Latvian-Latgalian Dictionary](#)
- [11] [Latvian-Lithuanian Dictionary](#)
- [12] [Latvian-Russian Dictionary](#)
- [13] [Lithuanian-Latvian Dictionary](#)
- [14] [Lithuanian-Latvian-Latgalian Dictionary \(LLL\)](#)
- [15] [Livonian-Estonian-Latvian Dictionary](#)
- [16] [Mühlenbach Endzelin Latvian Dictionary \(MEV\)](#)
- [17] [Rendering Dictionary of Personal Names](#)
- [18] [Russian-Latvian Dictionary](#)
- [19] [Tezaurs.lv](#)

2.2.7. 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 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 <i>Lehrbuch des Syrisch-Arabischen</i> . The resource is available for download from ARCHE.
A digital dictionary of Modern Standard Arabic 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 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 20 th century.

		The resource is available for download from ARCHE.
A digital dictionary of Dagaare 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 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) 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) 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) 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) 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 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 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 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 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 Size: 1,150 keywords	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.

Licence: CLARIN ACA		
Online Dictionary of Standard Estonian ÕS 2018 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 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 Size: 50,000 keywords Licence: CC-BY	Estonian	This 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 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 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-...) Linguistic information: geographical distribution, phonetic/phonological information Licence: CLARIN ACA	Estonian	This 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 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 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 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 Size: 33,000 keywords Linguistic information: orthography, inflection, semantics and etymology of the	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.

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		
The Dictionary of Old Võru Dialect 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 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 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 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 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 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 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 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	Estonian, English	This is a constantly incremented English-Estonian online dictionary, which is helpful in

<p>Size: 90,000 entries Licence: BSD - Style</p>		both human and machine translation. The resource is available for online browsing.
<p>The Online Dictionary of Estonian Sign Language - Estonian</p> <p>Size: 4,000 keywords Licence: CLARIN ACA</p>	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.
<p>Finnish-Estonian Dictionary</p> <p>Size: 90,000 keywords Licence: CC-BY</p>	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.
<p>Comprehensive Estonian-French Dictionary</p> <p>Size: 15,300 keywords Licence: undefined</p>	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.
<p>Estonian-Latvian Dictionary</p> <p>Size: 25,000 entries</p>	Estonian, Latvian	This is a bilingual dictionary that is available for online browsing.
<p>The Estonian-Latvian-Estonian Dictionary</p> <p>Size: 47,000 keywords Licence: CLARIN ACA</p>	Estonian, Latvian	This bilingual dictionary is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website.
<p>The Estonian-Mari Dictionary</p> <p>Size: 9,800 keywords Licence: CLARIN ACA</p>	Estonian, Mari	This bilingual dictionary is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website.
<p>The Estonian-Norwegian-Estonian Dictionary</p> <p>Size: 19,000 Estonian keywords, 21,000 Norwegian keywords Licence: CLARIN ACA</p>	Estonian, Norwegian	This bilingual dictionary is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website.
<p>Estonian-Russian Learner's Orthographic Dictionary</p> <p>Size: 23,000 keywords Licence: CLARIN ACA</p>	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.
<p>The Estonian-Russian Dictionary</p> <p>Size: 66,423 keywords Licence: CC-BY</p>	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.
<p>The Russian-Estonian Dictionary</p> <p>Size: 74,000 keywords Licence: CC-BY</p>	Estonian, Russian	This bilingual dictionary is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website.
<p>The Estonian-Udmurt Dictionary</p> <p>Size: 10,000 keywords Licence: CLARIN ACA</p>	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 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 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 Size: accruing Licence: CC-BY 4.0	Finnish	This is a dictionary of Old Literary Finnish. The resource is available for online browsing through the Language Bank of Finland.
Wurdboek fan de Frsyke taal 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) 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: <ol style="list-style-type: none">1. Dictionary of contemporary German2. DWDS-Dictionary3. Etymological dictionary of German4. The German Dictionary by Jacob Grimm and Wilhelm Grimm5. The Great Dictionary of the German Language in 10 Volumes6. OpenThesaurus This collection is available for online browsing from a dedicated environment associated with CLARIN-D.
Multilingual dictionaries 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 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 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 Size: 48,700 entries Linguistic information: Icelandic headwords with pronunciation (sound files) and inflection; translations into	Icelandic, Danish, Norwegian (Bokmål), Norwegian (Nynorsk),	This multilingual dictionary is available for download from CLARIN-IS as well as for online browsing.

Danish, Norwegian (Bokmål and Nynorsk), Swedish, Faroese Licence: CC-BY-NC-ND	Swedish, Faroese	
Karelian Dictionary 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 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) Size: 45,877 entries	Latvian	The resource is available for online browsing through the tezaurs.lv repository.
Dictionary of Standard Latvian Language (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) Size: 1,800 entries	Latvian	This is a dictionary of Latvian from the period between the 16 th and 17 th centuries with a reference to the “Historical Corpus of the Latvian Language”. The resource is available for online browsing.
Rendering Dictionary of Personal Names 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 Size: 309,016 entries	Latvian	This is a comprehensive corpus-based dictionary of the Latvian language. The resource is available for online browsing.
Mühlenbach Endzelin Latvian Dictionary (MEV) Size: 120,000 entries	Latvian (German)	This is a bilingual dictionary that is available for online browsing through the tezaurs.lv repository.
English-Latvian Dictionary Size: 70,000 entries	Latvian, English	This is a bilingual dictionary that is available for online browsing through the tezaurs.lv repository.
Latvian-English Dictionary Size: 80,000 entries	Latvian, English	This is a bilingual dictionary that is available for online browsing through the tezaurs.lv repository.
Latvian-Latgalian Dictionary Size: 25,700 entries	Latvian, Latgalian	This is a bilingual dictionary that is available for online browsing through the tezaurs.lv repository.
Latvian-Lithuanian Dictionary Size: 43,000 entries	Latvian, Lithuanian	This is a bilingual dictionary that is available for online browsing through the tezaurs.lv repository.
Latvian-Russian Dictionary Size: 44,000 entries	Latvian, Russian	This is a bilingual dictionary that is available for online browsing through the tezaurs.lv repository.
Lithuanian-Latvian Dictionary Size: 60,000 entries	Lithuanian, Latvian	This is a bilingual dictionary that is available for online browsing through the tezaurs.lv repository.

Lithuanian-Latvian-Latgalian Dictionary (LLL) 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	Livonian, Estonian, Latvian	This is a trilingual dictionary that is available for online browsing through the tezaurs.lv repository.
The Livonian-German Dictionary 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 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 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 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 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 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 Size: 512,973 entries Licence: Other - Institute of Computer Science PAS	Polish	<p>This is a collection of the following 4 dictionaries, which are available through the same online environment:</p> <ol style="list-style-type: none"> 1. Grammatical Dictionary of Polish (334,845 entries) 2. Vilnius Dictionary (108,513 entries) 3. The Polish Electronic Dictionary of the 16th c. (30,700 entries) 4. The Polish Electronic Dictionary of the 17th and 18th c. (38,915 entries) <p>All the 4 dictionaries are available through the same online lexical platform developed by CLARIN-PL.</p>
Polish Hask Collocation Database 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	Russian, Latvian	This is a bilingual dictionary that is available for online browsing through the tezausur.lv repository.
Etymological Database of the Saami Languages 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 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 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 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 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 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 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 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 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 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)	Slovenian, German	This is a historical dictionary compiling the works of Maks Pleteršnik. The resource is

<p>Size: 103,185 entries Licence: CC-BY 4.0</p>		available for download from CLARIN.SI and for online browsing.
<p>Japanese-Slovene learner's dictionary jaSlo 3.1</p> <p>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</p>	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.
<p>Dalín's Dictionary (2017-10-16)</p> <p>Size: 62,975 entries Licence: CC-BY 4.0</p>	Swedish	Dalín's Dictionary of 19th century Swedish. The resource can be download from the SWE-CLARIN repository and can be queried online through KARP.
<p>Hellquist's Swedish etymology (2017-10-16)</p> <p>Size: 12,376 entries Licence: CC-BY 4.0</p>	Swedish	This is an etymological dictionary that is available for download from the SWE-CLARIN repository and can be queried online through KARP.
<p>Swedish-South Saami dictionary</p> <p>Size: 3,351 entries Licence: CC-BY 3.0</p>	Swedish, Southern Sami	This is a bilingual dictionary that is available for download from CLARINO.
<p>Online Lexicon of Veps Language</p> <p>Licence: EUPL</p>	Veps	This is a dictionary of the Veps Language available for online browsing through FIN-CLARIN.

2.3. Conceptual resources

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

2.3.1. VLO

All of the conceptual resources listed in Table 3 can be found in the VLO.

2.3.2. Availability

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

- [1] [Bring \(2015-05-08\)](#)
- [2] [Cornetto-LMF](#)
- [3] [Finnish FrameNet](#)
- [4] [Finnish WordNet](#)
- [5] [MWNPT-International WordNet of Portuguese](#)
- [6] [Saldo](#)
- [7] [Semantic lexicon of Slovene sloWNet 3.1](#)
- [8] [Swedish FrameNet \(2017-10-16\)](#)
- [9] [Swesaurus \(2017-10-16\)](#)
- [10] [The Sanat Version of the Finnish TransFrameNet](#)
- [11] [WordNet-SALDO \(2017-10-16\)](#)
- [12] [plWordNet 4.0](#)

The following 16 resources are available for download only:

- [1] [Estonian Wordnet 2.1](#)
- [2] [Geo-Net-PT 02](#)
- [3] [Hontology](#)
- [4] [ItalWordNet Kyoto](#)
- [5] [ItalWordNet v.2](#)
- [6] [IWN-LOD](#)
- [7] [Maltese automatically produced distributional thesaurus](#)
- [8] [Ontology for the area of Nanoscience and Nanotechnology](#)
- [9] [Open Ancient Greek WordNet 0.5](#)
- [10] [Thesaurus of Modern Slovene 1.0](#)
- [11] [DanNet, Danish Wordnet \(v 2.2\)](#)
- [12] [STO semantics \(The Danish SIMPLE Lexicon\) - LMF format](#)
- [13] [Prolex](#)
- [14] [GermaNet](#)
- [15] [NE_SUMO_PLWN_mapping](#)
- [16] [PLWordNet to Sumo mapping](#)

The following resource is unavailable for download or online browsing:

- [1] [Polytropon EL Conceptual Lexicon](#)

2.3.3. 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
- [4] Polish, English
- [5] Portuguese, English, Spanish, French
- [6] Swedish, English

2.3.4. Size

All the resources in this category contain information on size. Among wordnets, the largest is [plWordNet 4.0](#), which contains 506,815 synsets, while the smallest is [Open Ancient Greek WordNet 0.5](#), which contains 7,447 synsets.

2.3.5. Linguistic information

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

- [1] [Cornetto-LMF](#) (semantic relationships and combinatorial information)
- [2] [DanNet, Danish Wordnet \(v 2.2\)](#) (hypernymy, hyperonymy)
- [3] [Estonian Wordnet 2.1](#) (PoS-tags, synonymy, antonymy, hypernymy, hyponymy, meronymy)
- [4] [Finnish WordNet](#) (PoS-tags, synonymy, antonymy, hypernymy, hyponymy, meronymy)
- [5] [Geo-Net-PT 02](#) (qualia structure and lexical relations)
- [6] [GermaNet](#) (MSD-tags, lexical relations (hypernymy, hyponymy))
- [7] [Hontology](#) (term correlations, synonymy)
- [8] [ItalWordNet v.2](#) (equivalence relations between Italian synsets and closest concepts in an Inter-Lingual index)
- [9] [MWNPT-International WordNet of Portuguese](#) (PoS-tags, hyponymy, hypernymy)
- [10] [Ontology for the area of Nanoscience and Nanotechnology \(lexical semantic information\)](#)
- [11] [Polytropon EL Conceptual Lexicon](#) (lemmas, MSD-tags, lexical relations)
- [12] [Prolex](#) (inflecte forms)
- [13] [Thesaurus of Modern Slovene 1.0 \(core and near synonymy\)](#)

2.3.6. Licence

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

- [1] [IWN-LOD](#)

2.3.7. 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 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 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) 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 and .owl formats.
STO semantics (The Danish SIMPLE Lexicon) - LMF format 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 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 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 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 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 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 Size: 72,572 lexical relations Linguistic information: inflected forms Licence: Licence Publique Générale Amondrie GNU	French, English, Polish, Serbian	This is an ontology of place names available for download from Ortolang.
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 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 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 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 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 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 Size: 120 terms Licence: CC BY SA 4.0	Polish	This conceptual resource provides a mapping between named entities types, SUMO categories and pWordNet synsets. The resource is available for download from the CLARIN-PL repository.
PLWordNet to Sumo mapping Size: 175,635 synsets mapped to SUMO Licence: CC-BY-NC-SA 3.0	Polish	This conceptual resource provides a mapping of pWordNet onto the SUMO ontology. The resource is available for download from the CLARIN-PL repository.

<p>plWordNet 4.0</p> <p>Size: 506,815 senses, 347,564 synsets Licence: plwordnet-2</p>	Polish, English	This is a lexico-semantic network which reflects the lexical system of the Polish language with projection to the English language. Słowsieć. The resource is available for download and browsing.
<p>Geo-Net-PT 02</p> <p>Size: 701,209 concepts Linguistic information: qualia structure and lexical relations (hyponyms, synonyms) Licence: CC-BY</p>	Portuguese	This is an ontology of Portuguese geographic concepts. It is available for download from CLARIN PORTULAN.
<p>MWNPT-International WordNet of Portuguese</p> <p>Size: 17,200 synsets Linguistic information: hyponymy and hypernymy Licence: MS NC-NoReD-ND</p>	Portuguese	This is a wordnet that is available for download and online browsing through CLARIN PORTULAN.
<p>Hontology</p> <p>Size: 282 concepts Linguistic information: terms correlation, rules (lexical patterns) and synonyms Licence: CC-BY-NC-SA</p>	Portuguese, English, Spanish, French	This is an ontology of concepts from the accommodation sector is available for download from CLARIN PORTULAN.
<p>Thesaurus of Modern Slovene 1.0</p> <p>Size: 105,473 entries Linguistic information: core and near synonymy Licence: CC BY-SA 4.0</p>	Slovenian	This is a thesaurus of the modern Slovenian language that is available for download from CLARIN.SI
<p>Semantic lexicon of Slovene sloWNet 3.1</p> <p>Size: 43,460 synsets Linguistic information: lexical semantic relations Licence: CC-BY-SA 4.0</p>	Slovenian, English	This is a wordnet available for download from CLARIN.SI and for online browsing through a dedicated environment.
<p>Bring (2015-05-08)</p> <p>Size: 148,815 entries Licence: CC-BY 4.0</p>	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.
<p>Saldo</p> <p>Size: 131,020 entries Licence: CC-BY 4.0</p>	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.
<p>Swedish FrameNet (2017-10-16)</p> <p>Size: 1,195 entries Licence: CC-BY 4.0</p>	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.
<p>Swesaurus (2017-10-16)</p> <p>Size: 15,010 entries</p>	Swedish	This is a Swedish wordnet that can be downloaded from the SWE-CLARIN

Licence: CC-BY 4.0		repository and can be queried online through KARP.
WordNet-SALDO (2017-10-16) 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.

2.4. Glossaries

There are 33 glossaries available in the CLARIN infrastructure, presented in detail in Table 4.

2.4.1. VLO

All of the resources listed in Table 4 can be found in the VLO.

2.4.2. Availability

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

- [1] [A Terminological Inventory for Biodiversity](#)
- [2] [Dictionary of Estonian Place Names](#)
- [3] [ESTERM \(20190101\)](#)
- [4] [The Dictionary of Education](#)
- [5] [The Dictionary of the Clothing Terminology of the Zilja Dialect in Canale Valley \(Kanalska dolina – Val Canale – Kanaltal – Valcjanâl\)](#)
- [6] [The English-Estonian Maritime Dictionary](#)
- [7] [UHR's Termbase for Norwegian higher education institutions](#)

The following 21 glossaries are available for download only:

- [1] [Bibterms. Glossary of archive and library sciences](#)
- [2] [English for business](#)
- [3] [French-Greek and Greek-French lexicon of football terms](#)
- [4] [Glossary e-book "Open Linked Data and Applications"](#)
- [5] [Glossary of Ancient Greek dramaturgy](#)
- [6] [Glossary of feminist theory](#)
- [7] [Glossary of Medical Terms](#)
- [8] [Greek Crisis Multilingual Term Project](#)
- [9] [Greek-French lexicon of football](#)
- [10] [INT IMPACT NE Lexicon](#)
- [11] [MFSA_Maltese_Company_Registry](#)
- [12] [Multilingual glossary \(Department of Foreign Languages, Translation and Interpretation\)](#)
- [13] [Multilingual lexicon of toponyms](#)
- [14] [Netterms. Computer networking terminology](#)
- [15] [Norwegian biodiversity terminology database](#)
- [16] [Orossimo Terminological Resource](#)
- [17] [PANACEA Environment Bilingual Glossary EL-EN \(Greek-English\)](#)
- [18] [PANACEA Labour Bilingual Glossary EL-EN \(Greek-English\)](#)
- [19] [Time-sensitive inventory of medical terminology](#)
- [20] [Xenophobia](#)
- [21] [Τρίγλωσσο Ορολογικό Λεξικό \(ΤΟΛ\)](#)

The following 5 glossaries are available only for online browsing:

- [1] [MILITERM \(20180801\)](#)
- [2] [Name Component Lexicon](#)
- [3] [The Family Name Database of the Institute of the Estonian Language](#)
- [4] [The Helsinki Term Bank for the Arts and Sciences](#)
- [5] [The Place Name Database of the Institute of the Estonian Language \(KNAB\)](#)

2.4.3. 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

2.4.4. Size

In terms of entries, the largest resource is the glossary [A Terminological Inventory for Biodiversity](#), containing 288,562 entries, while the smallest is [Netterms. Computer networking terminology](#), containing 205 entries.

The following glossaries lack information on size:

- [1] [Multilingual glossary \(Department of Foreign Languages, Translation and Interpretation\)](#)
- [2] [Name Component Lexicon](#)
- [3] [The Family Name Database of the Institute of the Estonian Language](#)

2.4.5. 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](#) lists the annotation layer “semantics – semantic class”, but does not further specify what semantic class pertains to).

- [1] [A Terminological Inventory for Biodiversity](#) (semantic relatedness)
- [2] [Greek-French lexicon of football](#) (lexical semantic features)
- [3] [Multilingual glossary \(Department of Foreign Languages, Translation and Interpretation\)](#) (lemmas)
- [4] [Multilingual lexicon of toponyms](#) (lemmas, semantics – semantic class, translation equivalents)
- [5] [Orossimo Terminological Resource](#) (lexical semantic relations, usage notes)
- [6] [PANACEA Environment Bilingual Glossary EL-EN \(Greek-English\)](#) (lemmas, translation equivalent)
- [7] [Time-sensitive inventory of medical terminology](#) (Semantics - relations [synonyms, holonymy, hyperonymy...])
- [8] [The Dictionary of the Clothing Terminology of the Zilja Dialect in Canale Valley \(Kanalska dolina – Val Canale – Kanaltal – Valcjanâl\)](#) (MSD-tags, phonological information)

2.4.6. 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 \(lexicon of linguistic terms: DE-EL-EN\)](#)

2.4.7. 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 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 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 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" Size: 100 semantic units Licence: CC-BY	English, Greek	This glossary is available for download from clarin:el.
Glossary of Medical Terms 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) 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 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	Estonian	This etymological glossary contains all official village names of Estonia and selected group of other

<p>Size: 6,214 keywords Licence: CC-BY</p>		Estonian place names. The resource is available for download from META-SHARE (CELR distribution) and for online browsing through a dedicated website.
<p>The Family Name Database of the Institute of the Estonian Language</p> <p>Licence: CLARIN ACA</p>	Estonian	This glossary is available for online browsing.
<p>The Place Name Database of the Institute of the Estonian Language (KNAB)</p> <p>Size: 428,344 elements; 145,261 elements Licence: CLARIN ACA</p>	Estonian	This glossary is available for online browsing.
<p>The English-Estonian Maritime Dictionary</p> <p>Size: 30,000 keywords Licence: CLARIN ACA</p>	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.
<p>MILITERM (20180801)</p> <p>Licence: CC-BY</p>	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.
<p>ESTERM (20190101)</p> <p>Size: 53,327 concepts Licence: CC-BY</p>	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.
<p>The Dictionary of Education</p> <p>Size: 4,850 keywords Licence: CLARIN ACA</p>	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.
<p>The Helsinki Term Bank for the Arts and Sciences</p> <p>Size: accruing Licence: CC-BY-SA</p>	Finnish, multilingual	This is a terminological database available for online browsing through for online browsing through FIN-CLARIN.
<p>French-greek and greek-french lexicon of football terms</p> <p>Size: 843 entries Licence: CC-BY – ND</p>	French-Greek	This bilingual glossary of football terminology is available for download from clarin:el.
<p>Glossary of Ancient Greek dramaturgy</p> <p>Size: 228 Lexical Types Licence: CC-BY</p>	Greek	This is a glossary of terms related to Ancient Greek Dramaturgy. The resource is available for download from clarin:el.
<p>Xenophobia</p> <p>Size: 12 knowledge bases Licence: CC-BY</p>	Greek	This is a collection of 12 knowledge bases on various forms of xenophobia, available for download from clarin:el.

<p>Bibterms. Glossary of archive and library sciences</p> <p>Size: 1,223 Lexical Types Licence: CC-BY</p>	Greek, English	This is a glossary of archival library science terms. The resource is available for download from clarin:el.
<p>Glossary of feminist theory</p> <p>Size: 100 entries Licence: CC-BY-NC</p>	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.
<p>Orossimo Terminological Resource</p> <p>Size: 22,308 terms Linguistic information: lexical semantic relations, usage notes Licence: CC-BY-NC-SA 4.0</p>	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.
<p>PANACEA Environment Bilingual Glossary EL-EN (Greek-English)</p> <p>Size: 18,582 entries Linguistic information: lemmas, translation equivalent Licence: CC-BY</p>	Greek, English	This glossary is available for download from clarin:el.
<p>Τρίγλωσσο Ορολογικό Λεξικό (TOΛ)</p> <p>Size: 5,224 Entries Licence: CC-BY-NC-ND</p>	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.
<p>Greek-French lexicon of football</p> <p>Size: 870 entries Linguistic information: lexical semantic features Licence: CC-BY-NC</p>	Greek, French	This is a glossary of football terms available for download from clarin:el.
<p>Greek Crisis Multilingual Term Project</p> <p>Size: 820 concepts Licence: CC-BY</p>	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.
<p>MFSA_Maltese_Company_Registry</p> <p>Size: 45,993 entries Licence: CC-BY-NC</p>	Maltese, English	This glossary is available for download from CLARIN PORTULAN.
<p>English for business</p> <p>Size: 498 entries Licence: CLARIN ACA</p>	Norwegian Bokmål, English	This glossary includes terms from the domains of economics and business administration. The resource is available for download from CLARINO.
<p>UHR's Termbase for Norwegian higher education institutions</p> <p>Size: 1,786 entries Licence: CLARIN_PUB-BY</p>	Norwegian Bokmål, English, Norwegian Nynorsk	This is a database of terms available for download from CLARINO and for online browsing.

<p>Norwegian biodiversity terminology database</p> <p>Size: 297 entries Licence: CLARIN_PUB-BY</p>	<p>Norwegian Bokmål, Norwegian Nynorsk, Latin</p>	<p>This glossary includes terms from the domains of biology. The resource is available for download from CLARINO.</p>
<p>Multilingual lexicon of toponyms</p> <p>Size: 281,375 entries Linguistic information: lemmas, semantics, translation equivalents Licence: CC-BY-SA</p>	<p>Polish + translations into English, Romanian, Hungarian, Bulgarian, Polish, Greek, German, Serbian, Slovak, Croatian</p>	<p>This is a glossary of toponyms available for download from clarin:el.</p>
<p>The Dictionary of the Clothing Terminology of the Zilja Dialect in Canale Valley (Kanalska dolina – Val Canale – Kanaltal – Valcjanâl)</p> <p>Size: 594 entries Linguistic information: MSD-tags, phonological information Licence: CC-BY 4.0</p>	<p>Slovenian</p>	<p>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 and illustrative photographs of traditional clothing items. The resource is available for download from CLARIN.SI and for online browsing.</p>
<p>Name Component Lexicon</p> <p>Licence: EUPL</p>	<p>Swedish</p>	<p>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.</p>

2.5. Wordlists

There are 56 wordlists available in the CLARIN infrastructure, described in detail in Table 5.

2.5.1. VLO

All the wordlists in Table 5 can be found in the VLO except for:

- [1] [Deutsche Wortschatz](#)

2.5.2. Availability

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

- [1] [Estonian Frequency Dictionary \(ver. 2.0\)](#)
- [2] [Idioms from the NEO lexicon DB](#)
- [3] [Kelly \(2017-10-16\)](#)
- [4] [Swedish words, LEXIN](#)
- [5] [Vocation list \(2015-01-10\)](#)

The following 43 wordlists are available for download only:

- [1] [Dictionary for the CST Lemmatizer](#)
- [2] [Kres corpus n-grams 2.0](#)
- [3] [Gos corpus n-grams 2.0](#)
- [4] [IMP corpus n-grams 2.0](#)
- [5] [Janes corpus n-grams 1.0](#)
- [6] [Psycholinguistic Descriptives](#)
- [7] [Finnish Verbal Colorative Constructions](#)
- [8] [INT Historical Word List](#)
- [9] [JRC-Names - a multilingual named entity resource](#)
- [10] [KELLY word-list Greek](#)
- [11] [Keywords and n-grams from a textbook corpus](#)
- [12] [Labial vibrants in Mangbetu: Archival form](#)
- [13] [LCM-PL](#)
- [14] [Lexical functions of Spanish verb-noun collocations](#)
- [15] [Maltese Fiction Wordlist](#)
- [16] [Maltese Wordlist](#)
- [17] [Modern Finnish Word List](#)
- [18] [Neologisms Online v3](#)
- [19] [Ngbugu digital wordlist: Archival form](#)
- [20] [Relative frequencies of part-of-speech n-grams in native and translated Finnish literary prose](#)
- [21] [The Finnish N-grams 1820-2000 of the Newspaper and Periodical Corpus of the National Library of Finland](#)
- [22] [The Swedish N-grams 1770-1940 of the Newspaper and Periodical Corpus of the National Library of Finland](#)
- [23] [Topics of library and information science](#)
- [24] [Λεξικό Γλωσσολογικών όρων: Γερμανικά – Ελληνικά - Αγγλικά \(lexicon of linguistic terms: DE-EL-EN\)](#)
- [25] [Vocabulaire d'archéologie](#)
- [26] [Vocabulaire d'art et archéologie](#)
- [27] [Vocabulaire de géographie de l'Amérique du Nord](#)
- [28] [Vocabulaire de Nutrition artificielle](#)
- [29] [Vocabulaire de Pathologies humaines](#)
- [30] [Vocabulaire de philosophie](#)
- [31] [Vocabulaire de préhistoire et protohistoire](#)
- [32] [Vocabulaire de Psychopathologie](#)

- [33] [Vocabulaire de Sciences de l'éducation](#)
- [34] [Vocabulaire de Sciences du langage](#)
- [35] [Vocabulaire de sociologie](#)
- [36] [Vocabulaire de Transferts de chaleur](#)
- [37] [Vocabulaire de Transfusion sanguine](#)
- [38] [Vocabulaire d'ethnologie](#)
- [39] [Vocabulaire d'histoire des sciences et des techniques](#)
- [40] [Vocabulaire d'histoire et sciences de la littérature](#)
- [41] [Vocabulaire d'Histoire et sciences des religions](#)
- [42] [Vocabulaire de sciences de la Terre](#)
- [43] [Vocabulaire d'Electronique et électro-optique](#)

The following 8 wordlists are available only for online browsing:

- [2] [Frequencies of Early Modern Finnish Words](#)
- [3] [Frequencies of Old Literary Finnish Words](#)
- [4] [Frequency Lexicon of the Finnish Newspaper Language](#)
- [5] [Frequency List of Written Finnish Word Forms](#)
- [6] [Names of Countries](#)
- [7] [The Conceptual File of Estonian Lexis of the Institute of Estonian Language](#)
- [8] [Deutsche Wortschatz](#)

2.5.3. *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

2.5.4. Size

The largest resource is [INT Historical Word List](#), containing 500,000 words, while the smallest is [Topics of library and information science](#), which contains 732 words.

The following 7 wordlist lack information on size:

- [1] [Dictionary for the CST Lemmatizer](#)
- [2] [JRC-Names - a multilingual named entity resource](#)
- [3] [Labial vibrants in Mangbetu: Archival form](#)
- [4] [Names of Countries](#)
- [5] [The Conceptual File of Estonian Lexis of the Institute of Estonian Language](#)
- [6] [Vocabulaire de Pathologies humaines](#)
- [7] [Vocabulaire de Transferts de chaleur](#)

2.5.5. Linguistic information

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

- [1] [Gos corpus n-grams 2.0](#) (frequency)
- [2] [IMP corpus n-grams 2.0](#) (frequency)
- [3] [Janes corpus n-grams 1.0](#) (frequency)
- [4] [JRC-Names - a multilingual named entity resource](#) (spelling varieties of names)
- [5] [Kres corpus n-grams 2.0](#) (frequency)
- [6] [Maltese Fiction Wordlist](#) (frequency)
- [7] [Modern Finnish Word List](#) (MSD-tags, lemmas)
- [8] [Vocabulaire d'archéologie](#) (paradigmatic lexical relations)
- [9] [Vocabulaire d'art et archéologie](#) (paradigmatic lexical relations)
- [10] [Vocabulaire de géographie de l'Amérique du Nord](#) (paradigmatic lexical relations)
- [11] [Vocabulaire de Nutrition artificielle](#) (paradigmatic lexical relations)
- [12] [Vocabulaire de Pathologies humaines](#)(paradigmatic lexical relations)
- [13] [Vocabulaire de philosophie](#) (paradigmatic lexical relations)
- [14] [Vocabulaire de préhistoire et protohistoire](#) (paradigmatic lexical relations)
- [15] [Vocabulaire de Psychopathologie](#) (paradigmatic lexical relations)
- [16] [Vocabulaire de sciences de la Terre](#) (paradigmatic lexical relations)
- [17] [Vocabulaire de Sciences de l'éducation](#) (paradigmatic lexical relations)
- [18] [Vocabulaire de Sciences du langage](#) (paradigmatic lexical relations)
- [19] [Vocabulaire de sociologie](#) (paradigmatic lexical relations)
- [20] [Vocabulaire de Transferts de chaleur](#) (paradigmatic lexical relations)
- [21] [Vocabulaire de Transfusion sanguine](#) (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](#) (paradigmatic lexical relations)
- [25] [Vocabulaire d'histoire et sciences de la littérature](#) (paradigmatic lexical relations)
- [26] [Vocabulaire d'Histoire et sciences des religions](#) (paradigmatic lexical relations)

2.5.6. 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](#)
- [2] [Neologisms Online v3](#)
- [3] [Ngbugu digital wordlist: Archival form](#)
- [4] [Deutsche Wortschatz](#)

2.5.7. 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 Licence: CLARIN-ACA-NC	Danish	This wordlist was used for the CST lemmatizer 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 Size: 500,000 word forms Licence: other	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 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 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) 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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.

<p>Vocabulaire de Nutrition artificielle</p> <p>Size: 2,500 entries Linguistic information: associative relationship, hierarchical relationship, preferred form, synonym shape Licence: CC-BY 4.0</p>	French, English	This is a controlled vocabulary of expressions from the domain of nutrition available for download from ORTOLANG.
<p>Vocabulaire de Pathologies humaines</p> <p>Linguistic information: hierarchical relationship, preferred form, synonym shape Licence: CC-BY 4.0</p>	French, English	This is a controlled vocabulary of expressions from the domain of medicine (pathological diseases) available for download from ORTOLANG.
<p>Vocabulaire de philosophie</p> <p>Size: 4,435 entries Linguistic information: preferred form, synonym shape Licence: CC-BY 4.0</p>	French, English	This is a controlled vocabulary of expressions of philosophical terms available for download from Ortolang.
<p>Vocabulaire de préhistoire et protohistoire</p> <p>Size: 3,093 entries Linguistic information: hierarchical relationship, preferred forms, synonym shape Licence: CC-BY 4.0</p>	French, English	This is a controlled vocabulary of expressions historical terms available for download from Ortolang.
<p>Vocabulaire de Psychopathologie</p> <p>Size: 575 terms Linguistic information: associative relationship, hierarchical relationship, preferred form, synonym shape Licence: CC-BY 4.0</p>	French, English	This is a controlled vocabulary of expressions from the domain of psychopathology available for download from Ortolang.
<p>Vocabulaire de Sciences de l'éducation</p> <p>Size: 2,681 entries Linguistic information: preferred form, synonym shape Licence: CC-BY 4.0</p>	French, English	This is a controlled vocabulary of expressions from the domain of education available for download from ORTOLANG.
<p>Vocabulaire de Sciences du langage</p> <p>Size: 6,142 entries Linguistic information: hierarchical relationship, preferred form, synonym shape Licence: CC-BY 4.0</p>	French, English	This is a controlled vocabulary of expressions from the domain of linguistics available for download from ORTOLANG.
<p>Vocabulaire de sociologie</p> <p>Size: 5,277 entries</p>	French, English	This is a controlled vocabulary of expressions from the domain of sociology available for download from ORTOLANG.

<p>Linguistic information: preferred form, synonym shape Licence: CC-BY 4.0</p>		
<p>Vocabulaire de Transferts de chaleur</p> <p>Linguistic information: associative relationship, Hierarchical relationship, Preferred form, Synonym shape Licence: CC-BY 4.0</p>	French, English	This is a controlled vocabulary of expressions of thermodynamic terms available for download from Ortolang.
<p>Vocabulaire de Transfusion sanguine</p> <p>Size: 2,000 entries Linguistic information: associative relationship, hierarchical relationship, preferred form, synonym shape Licence: CC-BY 4.0</p>	French, English	This is a controlled vocabulary of expressions from the domain of medicine (related to blood transfusion) available for download from ORTOLANG.
<p>Vocabulaire d'ethnologie</p> <p>Size: 9,517 entries Linguistic information: preferred form, synonym shape Licence: CC-BY 4.0</p>	French, English	This is a controlled vocabulary of expressions from the domain of ethnology available for download from ORTOLANG.
<p>Vocabulaire d'histoire des sciences et des techniques</p> <p>Size: 3,766 entries Linguistic information: contextual information, usage examples, preferred form, synonym shape Licence: CC-BY 4.0</p>	French, English	This is a controlled vocabulary of expressions from the domain technical sciences available for download from ORTOLANG.
<p>Vocabulaire d'histoire et sciences de la littérature</p> <p>Size: 11,065 entries Linguistic information: preferred form, synonym shape Licence: CC-BY 4.0</p>	French, English	This is a controlled vocabulary of expressions from the domain of literary studies available for download from ORTOLANG.
<p>Vocabulaire d'Histoire et sciences des religions</p> <p>Size: 4,581 entries Linguistic information: preferred form, synonym shape Licence: CC-BY 4.0</p>	French, English	This is a controlled vocabulary of expressions from the domain of philosophy and religion available for download from ORTOLANG.
<p>Vocabulaire de sciences de la Terre</p> <p>Size: 19,707 entries Linguistic information: hierarchical relationship, preferred form, synonym shape</p>	French, English, Spanish	This is a controlled vocabulary of expressions from the domain of geology available for download from ORTOLANG.

Licence: CC-BY 4.0		
Vocabulaire d'Electronique et électro-optique 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 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) 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 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 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 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 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 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 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 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 . The resource is available for download from CLARIN.SI
IMP corpus n-grams 2.0 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 . The resource is available for download from CLARIN.SI.

<p>Janes corpus n-grams 1.0</p> <p>Size: 351,029,703 n-grams Linguistic information: frequency Licence: CC-BY-SA 4.0</p>	Slovenian	This is a list of n-grams extracted from the Janes corpus of Slovenian user-generated content version 1.0 . The resource is available for download from CLARIN.SI
<p>Keywords and n-grams from a textbook corpus</p> <p>Size: 5,977 words Licence: CC-BY 4.0</p>	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.
<p>Kres corpus n-grams 2.0</p> <p>Size: 211,104,769 n-grams Linguistic information: frequency Licence: CC-BY-SA 4.0</p>	Slovenian	This is a list of n-grams extracted from the Kres corpus of written Slovenian . The resource is available for download from CLARIN.SI
<p>JRC-Names - a multilingual named entity resource</p> <p>Linguistic information: spelling varieties of names Licence: Open for Reuse with Restrictions</p>	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.
<p>Lexical functions of Spanish verb-noun collocations</p> <p>Size: 1,000 verb-noun pairs Linguistic information: lexicological classifications (free-word combinations, errors), semantic information</p>	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.
<p>Idioms from the NEO lexicon DB</p> <p>Size: 4,928 entries Licence: CC-BY 4.0</p>	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.
<p>Kelly (2017-10-16)</p> <p>Size: 10,510 entries Licence: CC-BY 4.0</p>	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.
<p>The Swedish N-grams 1770-1940 of the Newspaper and Periodical Corpus of the National Library of Finland</p>	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.

Licence: CC-BY 4.0		
Vocation list (2015-01-10) 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 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.

3. 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, Ngunu, 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](#) 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](#), 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.