



CLARIN ERIC and CLARIN-IT

Monica Monachini - *National Coordinator CLARIN-IT*

Francesca Frontini - *CLARIN Board of Directors*

Istituto di Linguistica Computazionale - ILC CNR

08/06/2021

Scienza aperta e gestione dei dati per le scienze umane e del patrimonio culturale: corso di formazione a distanza

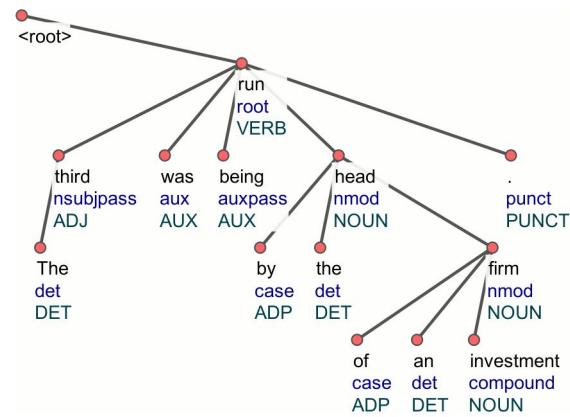
- CLARIN ERIC - di cosa si tratta?
- Cosa offre CLARIN-IT?
- Alcuni esercizi per mostrarvi le varie possibilità e i servizi che offre CLARIN
 - link in chat al documento
 - <https://docs.google.com/document/d/1XZoacpC1wd0YISKIJSIFKTYNxo6AmYG>

CLARIN ERIC



CLARIN

Common Language Resources and Technology Infrastructure



CLARIN in due parole



- detiene lo status di **ESFRI ERIC** dal 2012 e **Landmark** dal 2016
- attraverso modalità intuitiva e sostenibile, fornisce agli studiosi delle **scienze umane e scienze sociali** accesso a:
 - dati linguistici digitalizzati (in formato scritto, parlato, video o multimediali)
 - **strumenti avanzati** per scoprire, esplorare, annotare, analizzare o confrontare/combinare tali dati a prescindere di dove si trovino
 - attraverso un sistema di controllo di accesso **single sign-on**
- costituisce un ecosistema ideale per la condivisione di conoscenze
- è parte integrante dell' **European Open Science Cloud**
 - See clarin.eu/eosc

CLARIN: dati e comunità



- Newspaper archives
 - Literary texts
 - Parliamentary records
 - Literary texts
 - Historical letters
 - Broadcast archives
 - Oral History data
 - Social Media data
 - L2 Learner Resources
 - Survey data
 - ...
- Digital humanities
 - Linguistics and Philology
 - Translation and Lexicography
 - Literary Studies
 - History
 - Political and Social Sciences
 - Media Studies
 - Culture, Folklore, Anthropology
 - Speech therapy
 - General Public
 - ...

For the CLARIN Resource Families initiative, see:
<https://www.clarin.eu/resource-families>

CLARIN oggi



- **21 membri:** (AT, BG, CY, CZ, DE, DK, EE, FI, GR, HR, HU, IS, IT, LT, LV, NL, NO, PL, PT, SE, SI)
- 3 observers: FR, UK, ZA
- > 68 centri (25 centri certificati CTS)



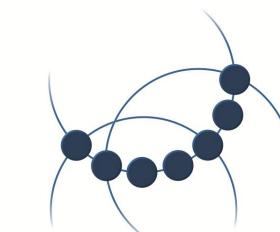
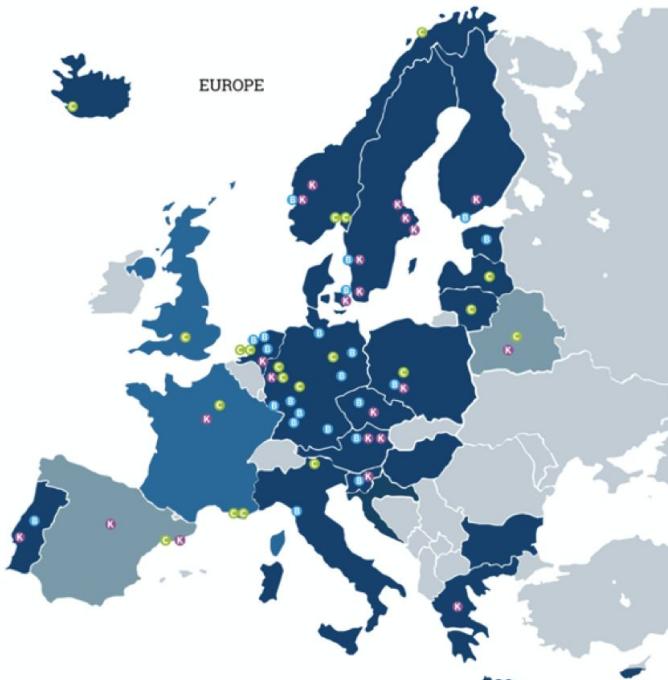
accesso federato



raccolta centralizzata di metadati



servizi concatenati



CLARIN

Common Language Resources and
Technology Infrastructure



Read more about CLARIN-UK

Data & Tools

[More](#)

CorCenCC

Corpus
Cenedlaethol
Cymraeg
Cyfoes – the
National Corpus
of
Contemporary
Welsh

Latest News

[More](#)

New
members of
the
CLARIN_UK
consortium

18 November
2020
Welcome
aboard!

Events

[More](#)

Lancaster
Symposium
on Innovation
in Corpus
Linguistics
2021

Wednesday 23
June, Online



Read more about CLARIN-UK

Data & Tools

[More](#)

CorCenCC
Corpus
Cenedlaethol
Cymraeg
Cyfoes – the
National Corpus
of
Contemporary
Welsh

#LancsBox

BNC

CKLD

CLAWS

CLiC

CorCenCC

CQPweb

ELAR

GATE

GATE Cloud

Hansard at
Huddersfield

[More](#)

Events

[More](#)

Lancaster
Symposium
on Innovation
in Corpus
Linguistics
2021

Wednesday 23
June, Online



Corpus
Linguistics

CLARIN Resource families



Corpora

- Computer-mediated communication corpora
- Corpora of academic texts
- Historical corpora
- L2 learner corpora
- Literary corpora
- Manually annotated corpora
- Multimodal corpora
- Newspaper corpora
- Parallel corpora
- Parliamentary corpora
- Reference corpora
- Spoken corpora

Lexical Resources

- Lexica
- Dictionaries
- Conceptual Resources
- Glossaries
- Wordlists

Tools

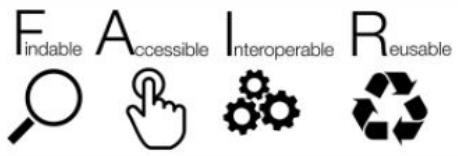
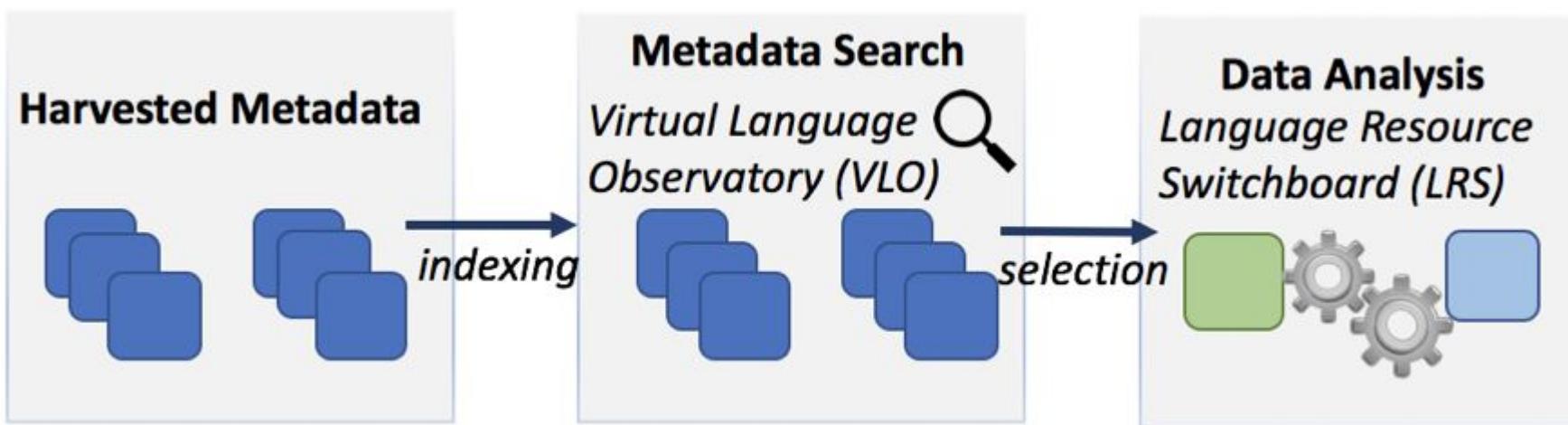
- Normalization
- Named entity recognition
- Part-of-speech tagging and lemmatization
- Tools for sentiment analysis

Spoken corpora in the CLARIN infrastructure

Corpora with transcriptions and audio recordings

Corpus	Language	Description	Availability
Arabic Speech Corpus	Arabic	<p>The corpus is available for download from a dedicated webpage.</p> <p>For a relevant publication, see Halabi (2016).</p>	Download
DIALEKT v1: dialectal corpus with multi-tier transcription	Czech	<p>This corpus contains traditional dialectological material, mostly unprepared monologue-type speech.</p> <p>The corpus is available download (upon request) and through the concordancer KonText.</p> <p>For a related publication, see Komrsková et al. (2018).</p>	Concordancer Download

L'infrastruttura tecnica



clarin.eu/fair



vlo.clarin.eu



switchboard.clarin.eu