

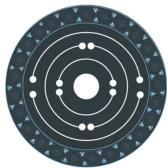
CLARIN e la scienza aperta



- Promozione della condivisione e riutilizzo dei dati attraverso dei registri sostenibili
- Tutti i dataset sono disponibili in accesso libero per scopi di ricerca
- Aderenza ai principi dei dati FAIR
 - Findable, Accessible, Interoperable, Re-usable (reperibile, accessibile, interoperabile, ri-utilizzabile)
 - Interoperabilità attraverso un sistema di metadati comune
- Promozione di una scienza responsabile
- Supporto alla diversità linguistica
 - I dati riguardano più di 1500 lingue
 - Strumenti per diverse lingue
 - Risorse Linguistiche disponibili in varie modalità
- Rafforzamento del supporto per professionisti e ricercatori nelle scienze sociali e materie umanistiche (> 500.000 in Europa)

[CLARIN: Towards FAIR and Responsible Data Science Using Language Resources.](#)" In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, May 2018, 3259-3264.

Progetti infrastrutturali



PARTHENOS

Pooling Activities, Resources and Tools
for Heritage E-research Networking,
Optimization and Synergies



Domande?

Esercizi 1 e 2

Pausa

CLARIN-IT - Consorzio

CLARIN-IT (dal 2015)



About Governance Consortium Centres Join Access Events Initiatives News **Home**

the Italian Common Language Resources and Technology Infrastructure

English ▾



ONLINE EVENTS

EUPORIA 2021 Webinar - Encoding a Critical Apparatus



07/12/2020 - [Registration](#) to follow the Webinar via Zoom

CLARIN-IT Consorzio



- Il Dipartimento di scienze della formazione, scienze imane e della comunicazione interculturale presso l'Università degli Studi di Siena
- Il Dipartimento di Filologia e Critica Letteraria presso l'Università degli Studi di Siena
- L'associazione Eurac Research (Bolzano)
- La fondazione Bruno Kessler (Trento)
- La Soprintendenza Archivistica e Bibliografica della Toscana (Firenze)
- Il Dipartimento di Ingegneria Elettrica e delle tecnologie dell'informazione e il centro interdipartimentale di ricerca "URBAN/ECO" de l'Università di Napoli Federico II
- L'Università Cattolica del Sacro Cuore (Milano) ha avviato la procedura per diventare membro.



CLARIN
Common Language Resources and
Technology Infrastructure



GARR cloud

Data registration



SSO



ERCC



CLARIN-IT Tematiche di ricerca



- **Risorse e Strumenti per la Lingua Italiana**

- Creazione di nuove risorse attraverso l'arricchimento dei corpora esistente
- Creazione di lessici da Linked Open Data
- Corpora specializzati per la comunicazione computer-mediated.
- strumenti di elaborazione e analisi del linguaggio naturale, offerti come [servizi web](#) and integrati in [Weblicht](#).

- **Risorse per Italiano regionali e corpora multilingue**

- Corpora di apprendenti (learner corpus) per il tedesco, l'italiano e il ceco.

- **Archivi Orali**

- Grafo, Archivio Caterina Bueno

- **Classici digitalizzati**

- Risorse per [greco antico](#) e [latino](#) (versione LOD del dizionario TEI-dict Perseus Liddell-Scott Jones)
- Latinità durante il medioevo italiano
- Edizioni digitali di antichi frammenti

ILC4CLARIN



[Chi siamo](#) [Organizzazione](#) [Repository](#) [Servizi](#) [Eventi CLARIN](#)



A large, complex network graph in the background, composed of numerous teal-colored nodes (dots) connected by thin teal lines (edges), representing a global research infrastructure.

ILC4CLARIN REPOSITORY

Easy to be found | Easy to be cited

ILC4CLARIN: risorse



CLARIN-IT Repository About CLARIN

ItalWordNet v.2

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

Roventini, Adriana; Marinelli, hosted at Institute for Computer Science
<http://hdl.handle.net/11336/1000>

BIBTEX CMDI

ALIM Share Home Authors Item identifier Project URL Demo URL Date issued Type Size Language Description

IT EN

ARCHIVIOVi.Vo.
Conservazione e diffusione
degli archivi orali e audiovisivi

REGIONE TOSCANA - Archivioorale Audiovisivo e Bibliografico della Toscana UNIVERSITÀ DI SIENA CASENTINO

Il CLARIN è un'infrastruttura europea per la conservazione, la diffusione e l'utilizzo dei dati di linguistica digitale.

The screenshot shows the ItalWordNet v.2 repository interface. On the left, there's a sidebar with various navigation links like 'ALIM', 'Share', 'Home', 'Authors', etc. The main content area features the 'ARCHIVIOVi.Vo.' service, which is described as 'Conservazione e diffusione degli archivi orali e audiovisivi'. Below this, logos for 'REGIONE TOSCANA', 'Archivioorale Audiovisivo e Bibliografico della Toscana', 'UNIVERSITÀ DI SIENA', and 'CASENTINO' are displayed. A large image of a person working on a computer is visible in the background. At the bottom, there's some handwritten text in a medieval script.

ILC4CLARIN: deposito



Chi siamo Organizzazione Repository Servizi Eventi CLARIN



CLARIN-IT: i servizi



Editing lessicale



LexO

Editor Web collaborativo utilizzato per la creazione e la gestione di risorse lessicali e terminologiche (multilingue) come risorse di dati collegati.

[Approfondisci](#)

Annotazione/analisi del testo



Freeling IT

Eseguibile di freeling solo per la lingua italiana, up to tagged.

[Prova questo strumento](#)



DeSR

Parser shift/reduce di dipendenze per la lingua italiana.

[Prova questo strumento](#)



LTFW (Linguistic Tools For Weblicht)

Porting Java del tokenizzatore basato su perl sviluppato all'interno del progetto OpeNER.

[Prova questo strumento](#)

Merging



LMF Merger

Merger LMF di due lessici italiani.

[Approfondisci](#)



LMF ML Merger

Merger LMF multilivello.

[Approfondisci](#)



Multiword Extractor: Average Frequency Filter

Estrattore di parole multiple che applica un filtro di frequenza media su MWE.

[Prova questo strumento](#)



Multiword Extractor: Merge Overlapping Strings

Estrattore di parole multiple che applica un filtro che unisce stringhe sovrapposte.

[Prova questo strumento](#)



Multiword Extractor: Remove Substrings

Estrattore di parole multiple che applica un filtro che rimuove le sottostringhe.

[Prova questo strumento](#)



Multiword Extractor: Reorder

Estrattore di parole multiple che applica un filtro che riordina i dati ExtractorMW per coppia_frequenza.

[Prova questo strumento](#)

Servizi lessicali



ItalWordNet

Parola: Invia Mostra tutte le relazioni

casa, Nome

- [1] - edificio o parte di esso in cui si abita;
([abitazione](#) [1], [casa](#) [1], [dimora](#) [2], [magione](#) [1], [ostello](#) [2], [tetto](#) [4])
- [2] - edificio con particolari funzioni.
([casa](#) [2])
- [3] - dinastia, casa regnante; l'insieme dei sovrani, appartenenti a una stessa famiglia, che si succedono sul trono
([casa](#) [3], [casa regnante](#) [1], [dinastia](#) [1])
"la casa di Savoia non ha regnato a lungo sull'Italia"
- [4] - la famiglia alla quale si appartiene
([casa](#) [4])
"sta pensando di mettere su casa"
- [5] - ditta, compagnia, in particolare nel settore dell'editoria e della moda
([casa](#) [5])
"casa editrice"
- [6] - casella nel gioco degli scacchi
([casa](#) [6])

Servizi TAL



URL del servizio: [freeling_it \(WSDL\)](#)

[Come usare il servizio](#)

Prova questo servizio attraverso il form sottostante.

Run service

Inputs

input

Choose file No file chosen

as URL
 direct data or local file

language

it.cfg ▾

Report

mandatory
optional

multiword yes no

ner --Use Default-- ▾

output_format --Use Default-- ▾

Esercizio 2

- Cercate il deposito/repository ILC4CLARIN
 - Quale è la lingua più rappresentata in termini numero di voci?
 - Che tipo di dati sono disponibili in Arabo?
 - Come si cita CophiWordNet?
- Provate ad accedere con il vostro profilo istituzionale attraverso la funzione Login in alto a destra