



Strumenti e Risorse per l'Analisi Lessicale dei Testi

L'Esempio di Manzoni

16 aprile 2025

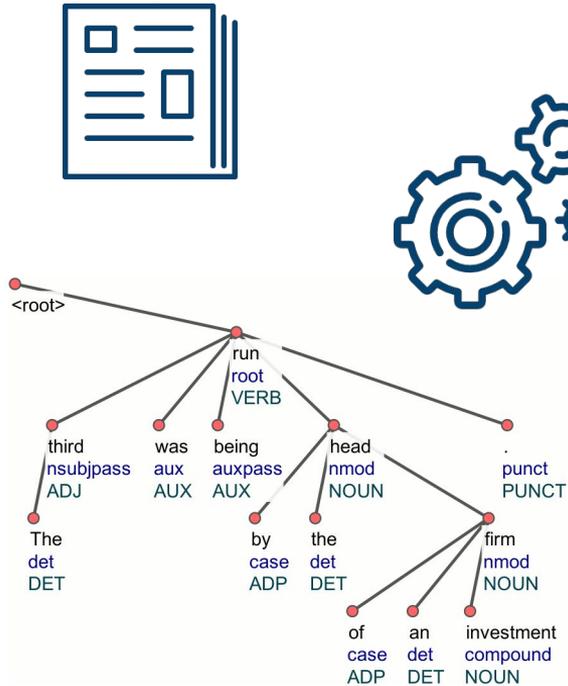
Giulia Pedonese, CNR-ILC
Rachele Sprugnoli, UniCatt



Indice

- Prima parte:
 - Cos'è l'Infrastruttura CLARIN e perché è utile alla didattica
- Seconda parte:
 - Distant, Close e Scalable Reading
 - Come utilizzare Voyant Tools
 - Un esempio pratico: l'analisi dei Promessi Sposi
 - Altre risorse del portale Manzoni
 - Un approccio diverso ai testi: la Sentiment Analysis

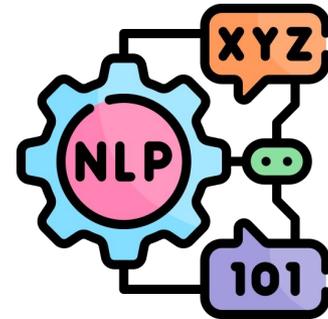
Le risorse linguistiche in CLARIN



Cosa sono le Risorse Linguistiche?

Per Risorse Linguistiche si intende **l'insieme di dati e metadati relativi al linguaggio in forma leggibile dalla macchina**. Esse vengono utilizzate per costruire, migliorare o valutare algoritmi o sistemi di linguaggio naturale (linguaggio umano) e vocale e, sempre più spesso, per l'apprendimento automatico. Sono inoltre utilizzati come risorse fondamentali ad esempio per gli studi linguistici e di traduzione, per l'editoria elettronica, per le transazioni internazionali, per gli specialisti di aree tematiche e per gli utenti finali

Fonte: [ELRA/ELDA](#)



Tipologie di Risorse Linguistiche

Le risorse linguistiche sono tipicamente suddivise in categorie a seconda del tipo di contenuto che includono:

Risorse testuali > corpora scritti e parlati

Risorse lessicali > lessici, dizionari e banche dati terminologiche

Strumenti NLP > lemmatizzatori, tagger PoS, parser ecc.

Le risorse linguistiche non includono solo i dati, ma anche la loro descrizione, o metadati, che arricchiscono i dati con informazioni aggiuntive, come la divisione strutturale in libri o tratti linguistici come i tag PoS e le funzioni sintattiche

CLARIN Resource Families

Corpora

- Computer-Mediated Communication Corpora
- Corpora of Academic Texts
- Historical Corpora
- L2 Learner Corpora
- Legal Corpora
- Literary Corpora
- Manually Annotated Corpora
- Multimodal Corpora
- Newspaper Corpora
- Oral History Corpora
- Parallel Corpora
- Parliamentary Corpora
- Reference Corpora
- Sign Language Resources
- Spoken Corpora

Lexical Resources

- Language Models
- Lexica
- Dictionaries
- Conceptual Resources
- Glossaries
- Wordlists

Tools

- Corpus Query Tools
- Normalisation
- Named Entity Recognition
- Part-of-Speech Tagging and Lemmatisation
- Tools for Sentiment Analysis

Dove trovare le Risorse Linguistiche? Gli Archivi Digitali

“Definiamo un repository - o deposito o archivio digitale - come un servizio gestito da organizzazioni di ricerca, in cui i materiali di ricerca sono conservati, gestiti e resi accessibili”. Fonte: [DataCite](#)

Per visualizzare i diversi tipi di risorse linguistiche e comprendere il modo in cui vengono descritte in un repository d'esempio, si può visitare ad la pagina <https://ilc4clarin.ilc.cnr.it/>



L'infrastruttura di Ricerca CLARIN

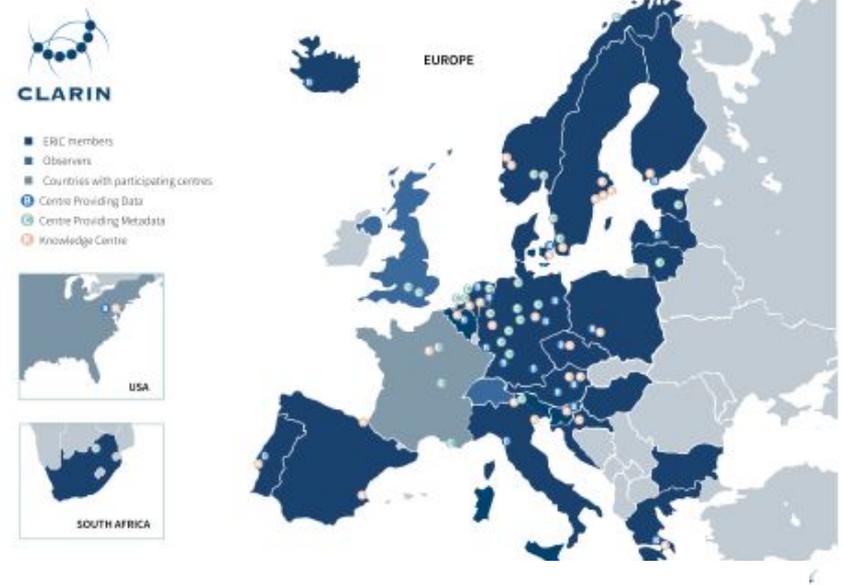
CLARIN è acronimo di **Common Language Resources and Technology Infrastructure**



CLARIN è un'infrastruttura digitale distribuita che fornisce un accesso facile e sostenibile a un'ampia gamma di dati e strumenti linguistici a sostegno della ricerca nelle scienze umane e sociali e non solo. CLARIN fornisce l'accesso a dati linguistici digitali multimodali (testo, audio, video) e a strumenti avanzati con cui esplorare, analizzare o combinare questi insiemi di dati

CLARIN Oggi

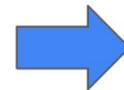
- Una rete distribuita di 72 centri ([Centre Registry](#))
- 24 membri: AT, BE, BG, CY, CZ, DK, EE, ES, FI, GR, HR, HU, IS, IT, LT, LV, NL, NO, PL, PT, SE, SI, ZA
- 2 osservatori: CH, UK
- Terza parte: Università Carnegie Mellon, USA



CLARIN ERIC



Consorzi Nazionali



Centri CLARIN

I Consorzi Nazionali: CLARIN-IT

CLARIN-IT è il nodo italiano di CLARIN.

L'Italia è diventata il 16° membro a pieno titolo di CLARIN ERIC nel 2015. Il membro fondatore del Consorzio nazionale è l'Istituto di Linguistica Computazionale "Antonio Zampolli" del Consiglio Nazionale delle Ricerche. CLARIN è coinvolto nel progetto H2IOSC attraverso il suo consorzio nazionale italiano CLARIN-IT



A screenshot of the CLARIN-IT website. At the top right, the word "REPOSITORY" is written in bold, with the tagline "Facile da trovare | Facile da citare" below it. Below this are four circular icons: a red upload icon labeled "DEPOSITA", a green magnifying glass icon labeled "CERCA", a blue document icon labeled "CITA", and a yellow question mark icon labeled "FAQ". The main content area includes the CLARIN-IT logo, a navigation menu (About, Governance, Consorzio, Carri, Login), and a section titled "L'infrastruttura Comune Italiana per le Risorse". Below this is a banner image of the Italian flag with the CLARIN-IT logo and the text "L'Italia Membro Titolare di CLARIN ERIC dal 1° Ottobre 2015". To the right, there are two news snippets: "CONFERENZA CLARIN" and "ULTIME NOTIZIE". At the bottom right, the "ERIC FORUM" logo is visible.

CLARIN-IT

Il Consorzio CLARIN italiano comprende i seguenti centri:

- [ILC4CLARIN](#) istituito presso il CNR Istituto di Linguistica Computazionale “Antonio Zampolli” di Pisa, è un centro di fornitura servizi ([Type B](#))
- [EURAC Research CLARIN Centre](#) (ERCC) presso l’EURAC Research Association di Bolzano, è un centro di fornitura metadati ([Type C](#))
- [The Knowledge Centre for Computer-Mediated Communication and Social Media Corpora](#) (CKCMC), è un centro CLARIN [di Tipo K](#) che mette a disposizione le proprie competenze e la propria conoscenza su aspetti del dominio disciplinare di CLARIN
- [The CLARIN Knowledge Centre for Digital and Public Textual Scholarship](#) (DiPText), anch’esso un centro di tipo K distribuito fra due partner: CNR-ILC e l’Università Ca’ Foscari di Venezia

I servizi di CLARIN

- [Servizi di deposito](#) per garantire che le risorse linguistiche possano essere archiviate e rese disponibili alla comunità in modo affidabile e per aiutare i ricercatori a conservare le loro risorse in modo sostenibile
- [Virtual Language Observatory](#) fornisce un'interfaccia di facile utilizzo, che consente un processo di ricerca e scoperta di risorse provenienti da un'ampia varietà di domini
- [Federated Content Search](#) un motore di ricerca che si collega alle raccolte di dati locali disponibili nei centri.
- [Language Resource Switchboard](#) aiuta gli utenti a trovare un'applicazione web per elaborare i propri dati linguistici
- [Virtual Collection Registry](#) fornisce un registro in cui gli utenti possono creare e pubblicare le loro collezioni virtuali

Come accedere ai servizi CLARIN

Tutti gli utenti possono esplorare liberamente i servizi centrali di CLARIN per cercare risorse e competenze linguistiche. A causa di restrizioni di licenza, alcune risorse sono disponibili solo per gli utenti accademici e il login è richiesto utilizzando le credenziali istituzionali o le credenziali CLARIN

Gli utenti accademici di tutti i Paesi partecipanti possono accedere e utilizzare le risorse linguistiche disponibili nei centri dati CLARIN con un accesso single sign-on attraverso la [CLARIN Service Provider Federation](#) utilizzando le proprie credenziali istituzionali

CLARIN Service Provider Federation

 Sign in via the CLARIN Service Provider Federation



Select your home organisation below. This is usually the organisation where you work or study. Signing in here will allow you to access certain CLARIN resources and services which are only available to users who have logged in. If you cannot find your organisation in the list below, please select the clarin.eu website account and use your CLARIN website credentials. If you don't have such credentials you can register an account [here](#). For questions please contact spf@clarin.eu.

Warning: It appears as if you visited this page directly, this will not work. Please login [via the service](#) you are trying to access.

Home organisation list



All countries



IRCCS ISMETT - Palermo



Italy

ISMETT

University of Palermo



Italy



UNIVERSITÀ
DEGLI STUDI
DI PALERMO

Trovare risorse linguistiche pubblicate

Virtual Language Observatory

- Un catalogo che raccoglie metadati sulle risorse linguistiche disponibili in archivi distribuiti
- Non contiene risorse linguistiche, ma aiuta a localizzarle tramite identificatori persistenti
- Anche se una risorsa ha un accesso limitato, i metadati sono sempre liberamente accessibili
- Utilizza la ricerca a faccette per restringere le ricerche

NB: uso di [licenze](#)



CLARIN Virtual Language Observatory

VLO / Faceted search / Search results

treasure island

Showing 3 results for treasure island

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

Language

Collection

Resource type

File type

▼

Temporal Coverage

Availability

Search options

Treasure Island / Robert Louis Stevenson

(Part of OTA Core Collection)

No description

English

The search results include 1 record with the same title.

Landing page for this record

Mind the Gap

(Part of DARIAH-DE Repository)

Mind The Gap was founded in 1988 by Tim Wheeler and Susan Brown and has since become one of Europe's leading disability theatre companies, with operations in both the UK and internationally. The primary aim of Mind the Gap's extensive team is to create equal opportunities for actors with learning disabilities within the...

English

La isla del tesoro - Spanish

(Part of DoBeS archive : Movima)

The Movima Archive Movima is a genetically unclassified language spoken in the so-called Moxos region in the savannahs of the Bolivian Amazon area. It is still spoken by more than 1,000 people (1,452 in 1996). Most speakers are over 50 years old and bilingual in Spanish. There are only very few children who

Treasure Island / Robert Louis Stevenson

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

BIBTEX CMDI

Stevenson, Robert Louis, 1850-1894, 1994, *Treasure Island / Robert Louis Stevenson*, Literary and Linguistic Data Service, <http://hdl.handle.net/20.500.14106/2055>.

Share:   

Oxford Text Archive

Authors Stevenson, Robert Louis, 1850-1894

Date of publication 1882

Type Text

Language(s) English

OTA identifier ota:2055

Collection(s) OTA Core Collection

Show full item record

This item is **Publicly Available** and licensed under:
Attribution-NonCommercial-ShareAlike 3.0 Unported (CC BY-NC-SA 3.0)

Files for this item

Name	Size	Format	Description
treas10-2055.txt	370.51 KB	Text file	Version of the work in plain text format



VCR

2 141



Browse

> All of the Repository

My Account

Login

Statistics

Statistics **BETA**

General Information

Deposit

Cite

Oxford University users

Submission Lifecycle

FAQ

About

Help Desk

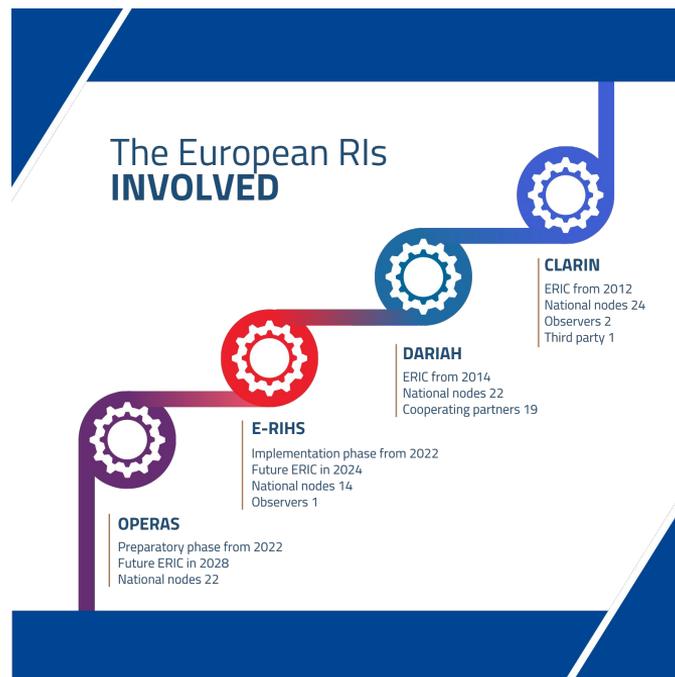
Privacy policy



Humanities and cultural Heritage Italian Open Science Cloud

Per approfondire la formazione di
CLARIN-IT in H2IOSC:

[https://www.h2iosc.cnr.it/training
-infrastructure/](https://www.h2iosc.cnr.it/training-infrastructure/)



SECONDA PARTE

Quali file usare

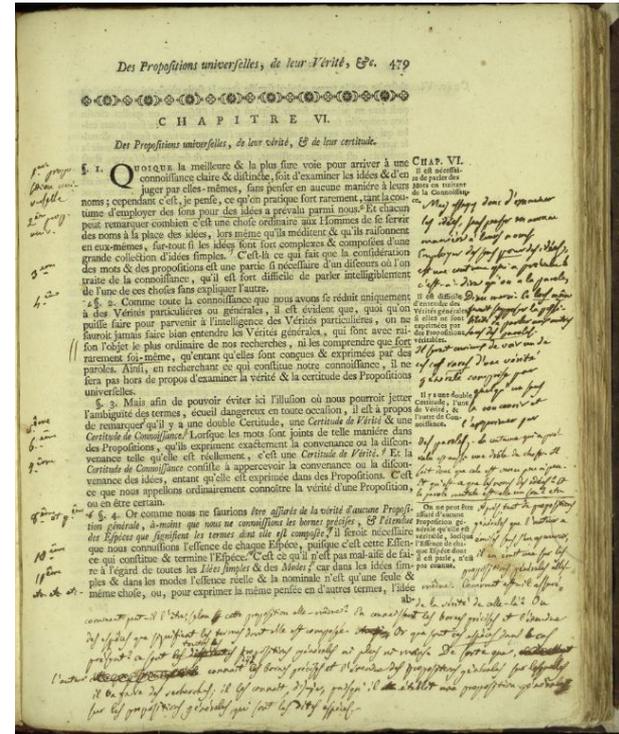
- *Promessi Sposi 1840* - singoli capitoli
- Le tre redazioni:
 - *Fermo e Lucia* - testo intero
 - *Promessi Sposi 1827* - testo intero
 - *Promessi Sposi 1840* - testo intero
- Lista di stopwords

<https://tinyurl.com/5yubvvx2>

Close Reading

“The principal object of close reading is to **unpack the text**. Close readers linger over words, verbal images, elements of style, sentences, argument patterns, and entire paragraphs and larger discursive units within the text to explore their significance on multiple levels.”

Jasinski, *Sourcebook on Rhetoric*, 2001



Scalable Reading

*"I am not suggesting a wholesale shelving of close reading and highly interpretive "readings" of literature. Quite the opposite, I am suggesting a **blended approach**. [...] It is exactly this sort of unification, of the macro and micro scales, that promises a new, enhanced, and better understanding of the literary record. **The two scales of analysis work in tandem and inform each other. Human interpretation** of the "data," whether it be mined at the macro or micro scale, **remains essential**. Although the methods of inquiry, of evidence gathering, are different, they are **not antithetical**, and they share the same ultimate goal of informing our understanding of the literary record, be it writ large or small. The most fundamental and important difference in the two approaches is that the macroanalytic approach reveals details about texts that are, practically speaking, unavailable to close readers of the texts."*

Jockers, *Macroanalysis. Digital Methods and Literary History*, 2013

Voyant Tools

- Voyant Tools è un ambiente web per la lettura e l'analisi di testi:
 - aggrega vari strumenti di analisi lessicale
 - consente di usare vari formati di input: txt, pdf, html, xml
 - può essere integrato su altri siti
 - è interattivo
 - permette lo scalable reading
 - è (per lo più) indipendente dalla lingua



Voyant Tools

- Voyant Tools è un ambiente **web** per la lettura e l'analisi di testi:
 - aggrega vari strumenti di **analisi lessicale**
 - consente di usare vari formati di input: txt, pdf, html, xml
 - può essere integrato su altri siti
 - è interattivo
 - permette lo scalable reading
 - è (per lo più) indipendente dalla lingua



Due modalità d'uso in ambito didattico:

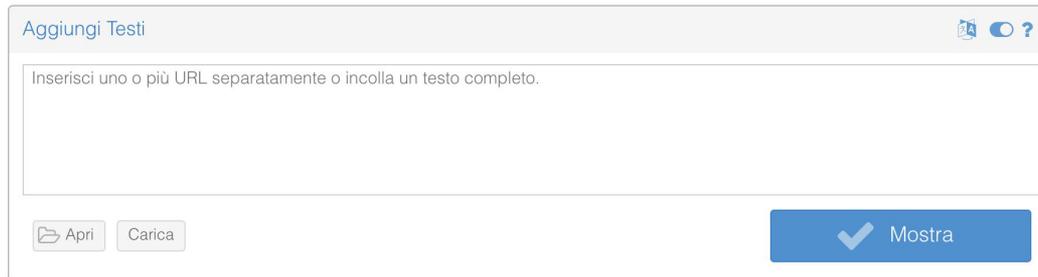
- 1) a **scuola**, nel laboratorio informatico
- 2) a **casa** → in fase di preparazione delle lezioni

Dove trovare Voyant

- Sito ufficiale: <https://voyant-tools.org/>
- Mirror (siti alternativi):
 - <https://voyant-tools.huma-num.fr>
 - <https://voyant.lincsproject.ca>
 - <https://service.sadilar.org/voyant/>
- Versione installabile (non richiede l'accesso a Internet): <https://github.com/voyanttools/VoyantServer>
- CLARIN switchboard: <https://switchboard.clarin.eu/>

Quale tipo di file caricare

- Formati accettati: TXT, HTML, XML, TEI, PDF, RTF, MS Word, tabelle in fogli di calcolo
- PDF: contiene un OCR (*Optical Character Recognition*), il risultato può variare
- HTML: rimuove i principali tag HTML
 - Accetta in input anche cartelle compresse (.zip) contenenti file nei suddetti formati
- **ATTENZIONE:** il modo in cui vengono caricati i file influenza il tipo di analisi successiva
 - eg.: testo unico in un solo file *versus* divisione in più file: un file solo con tutti i Promessi Sposi oppure un file per ogni capitolo?



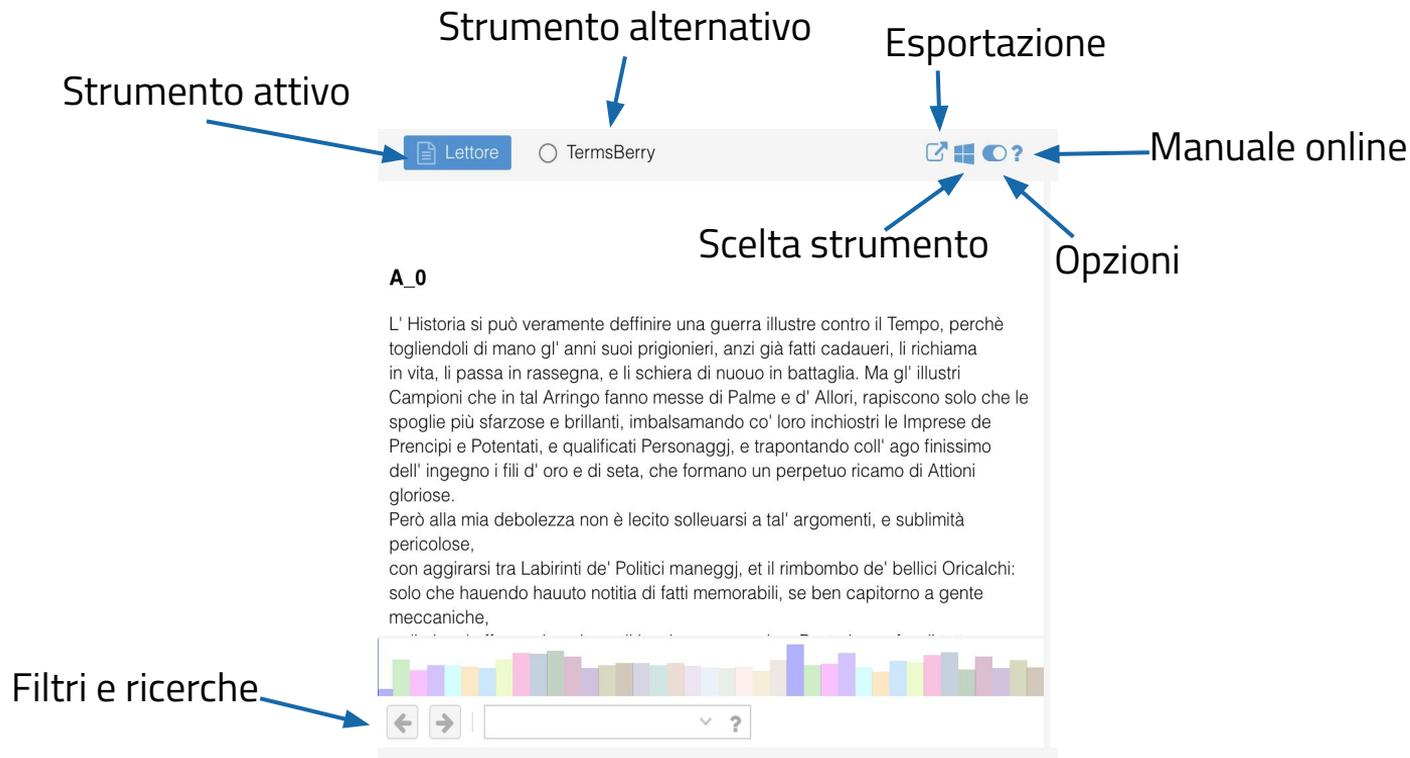
Aggiungi Testi

Inserisci uno o più URL separatamente o incolla un testo completo.

Apri Carica Mostra

The screenshot shows a web interface for adding texts. It features a title bar 'Aggiungi Testi' with a search icon, a toggle switch, and a help icon. Below the title bar is a large text input field with the placeholder text 'Inserisci uno o più URL separatamente o incolla un testo completo.'. At the bottom of the interface, there are three buttons: 'Apri' (with a folder icon), 'Carica' (with a document icon), and 'Mostra' (with a checkmark icon).

Struttura di una skin



Funzioni principali - 1

- Cambiare skin



- READER: **lettore** del testo, permette il close reading
- CIRRUS: visualizzatore **frequenza** dei termini
- BUBBLES: visualizzatore **frequenza** dei termini
- TERMS: analisi della **frequenza** dei termini
- TRENDS: andamento delle **frequenza** dei termini
- BUBBLELINES: **frequenza** e **distribuzione** dei termini
- MICROSEARCH: **frequenza** e **distribuzione** dei termini
- CONTEXT: **contesti** di occorrenza dei termini

Funzioni principali - 2

- Cambiare skin



- PHRASES: sequenze di parole che co-occorrono
- COLLOCATES: termini che appaiono vicino ad altri termini
- CORRELATIONS: termini la cui frequenza varia in sintonia
- MANDALA: relazioni tra termini e documenti
- SUMMARY: **informazioni** sul corpus
- DOCUMENTS: **informazioni** sui singoli documenti
- TOPICS: estrazione degli **argomenti** principali (topic modeling)



<https://voyant-tools.org/docs/tutorial-tools.html>



Come e cosa esportare

- L'esportazione  si può applicare all'intero progetto Voyant o al risultato di un singolo strumento
- È possibile esportare:
 - Una **URL** o un codice incorporabile in un altro sito
 - Un file **.png** statico nel caso delle visualizzazioni (ATTENZIONE: uno screenshot potrebbe avere una migliore qualità dell'immagine)
 - I dati in forma di **tabella** o in HTML

N.B. La URL rimane valida per **un mese** dall'ultima visita

Come cercare

- Sintassi delle ricerche lessicali (campo di ricerca nella parte bassa delle skin)
 - pestilenza: trova il termine esatto
 - pestilen*: trova termini che iniziano con "pestilen" → *pestilenza, pestilenze, pestilente...*
 - ^pest*: trova i termini che iniziano con "pest" e continuano con altre lettere → *peste, pestan, pestava, pestilenza, pestilenze, pestilente...*
 - "opera di misericordia": cerca l'intera espressione
 - "opera misericordia"~5: "opera" e "misericordia" co-occorrono entro 5 termini → *opera di misericordia, opera per fare una gran misericordia*

Come caricare i file

- Aprire Voyant (sito ufficiale, mirror o versione in locale)

- cliccare su opzioni



- sotto **Processing** scegliere **Simple Word Boundaries/Limiti di parola semplici**

The following table summarizes tokenization for the string `What's voyant-tools.org?`:

Tokenization	Count	Tokens	Notes
Automatic	3	<code>what's, voyant, tools.org</code>	the hyphen is split but the tools.org is considered a URL token; tokens are lowercase
Word Boundaries	5	<code>what, s, voyant, tools, org</code>	any non-word character is a delimiter, tokens are lowercase
Whitespace Only	2	<code>What's, voyant-tools.org?</code>	punctuation is kept in tokens and case is unchanged

- cliccare su **Upload**, aprire la cartella **capitoli**, selezionare (ctrl+a) tutti i file e cliccare su **Apri**: sono i capitoli de *I Promessi Sposi*

Aggiungere le *stopword* - 1

- Parole da ignorare: di solito parole grammaticali come congiunzioni, preposizioni, articoli...
- Dove trovare le liste di *stopword*?
 - Lingue moderne: <https://www.ranks.nl/stopwords>
 - Latino e greco antico:
<https://github.com/aurelberra/stopwords/tree/master/ancientstopwords/data>
 - Lista di *stopword* per l'italiano nella cartella Voyant: aprire il file **stopword-manzoni.txt** con un editor di testo, selezionare e copiare la lista

Aggiungere le *stopword* - 2

- Su Voyant, cliccare sulle opzioni della skin Cirrus



1) Selezionare "Italian"

The screenshot shows the 'Options' dialog box with the following settings:

- Stopwords: Auto-detect (dropdown)
- White List: (empty dropdown)
- Categories: auto (dropdown)
- Font family: Palatino (dropdown)
- Palette: default (dropdown)

Buttons and checkboxes:

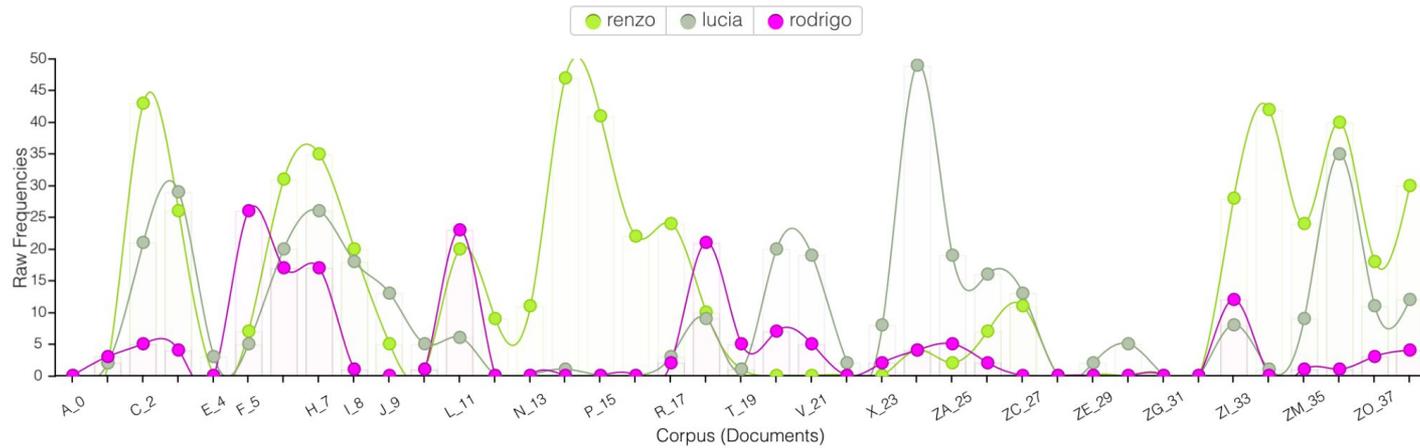
- Stopwords: Edit List
- White List: Edit List
- Categories: Edit
- Palette: Edit List
- apply globally (checked checkbox)
- Reset
- Cancel
- Confirm

2) Cliccare su **Edit List**,
incollare la lista del file
stopwords-it.txt

3) Cliccare su **Confirm**

Strumenti: TRENDS / ANDAMENTI

- Grafico a linee che mostra la distribuzione delle occorrenze di una parola
- Diverse opzioni disponibili tramite il pulsante "Visualizza"
- Due modalità di frequenza: assoluta o relativa



Strumenti: SUMMARY / SOMMARIO

- Panoramica del corpus:
 - numero di parole, numero di parole uniche, documenti più lunghi e più brevi, con densità di vocabolario più alta e più bassa, numero medio di parole per frase, livello di leggibilità (ATTENZIONE: solo per testi inglesi, indice di Coleman-Liau), parole più frequenti, parole distintive per documento (TF-IDF)

Questo corpus contiene 39 documenti con 223,830 totale parole e con 20,902 forme di parola uniche. Creato circa 32 minuti fa.

Lunghezza del documento: 

- Il più lungo/a: Y_24 (9054); K_10 (7855); ZL_34 (7591); I_8 (7513); ZD_28 (7473)
- Il più corto/a: A_0 (1309); ZF_30 (4275); W_22 (4359); C_2 (4475); ZM_35 (4681)

Densità del vocabolario: 

- Maggiore: A_0 (0.536); W_22 (0.376); M_12 (0.364); ZH_32 (0.357); U_20 (0.350)
- Minore: Y_24 (0.265); ZN_36 (0.275); ZP_38 (0.285); O_14 (0.293); K_10 (0.296)

Strumenti: DOCUMENTS / DOCUMENTI

- Tabella con informazioni numeriche su ciascun documento (nel nostro caso, su ogni capitolo)
- Include una specifica funzionalità per modificare il corpus

	Titolo	Parole	Tipi	Ratio	Words/Sentence
1	A_0	1,309	701	54%	38.5
2	B_1	6,396	2,163	34%	32.8
3	C_2	4,475	1,506	34%	18.0
4	D_3	5,421	1,727	32%	21.4
5	E_4	5,399	1,836	34%	29.7

Strumenti: CONTEXTS / CONTESTI

- Mostra ogni occorrenza della parola o espressione cercata con una porzione di testo circostante (il contesto), sia alla sua destra che alla sua sinistra.
- Clicca sull'icona più (+) per espandere una riga qualsiasi e visualizzare più contesto
- Cosa succede se cerchi "e"?

	Documento	Sinistra	Parola	Destra
	A_0	per disporle con qualche ordine,	misericordia	! venivano a fare un libro
	doveva esser fondato e posteale, con loro gran sorpresa, insieme, le mandavamo insieme a spasso. Non vi sarebbe mai stato autore che provasse così ad evidenza d' aver fatto bene. Ma che? quando siamo stati a quello di raccapezzare tutte le dette obiezioni e risposte, per disporle con qualche ordine, misericordia ! venivano a fare un libro. Il che veduto, ponemmo da canto il pensiero, per due ragioni che il lettore troverà certamente valide la prima, che un libro impiegato a giustificarne un altro, anzi lo stile d' un altro, potrebbe parere cosa ridicola la seconda, che di libri basta uno per			
	B_1	qualche cosa di bene straordinario. «	misericordia	! che ha ella, signor padrone
	B_1	dicendo: «per amor del cielo!» «	misericordia	!» sciamò Perpetua. «Oh ch...
	C_2	che gli usciva dalla tasca. «	misericordia	!» sciamò con voce fioca don
	E_4	seria. – E quell' altro disgraziato! –	misericordia	! che spettacolo! – Salvatelo...

Strumenti: TERMS / TERMINI

- Tabella delle frequenze dei termini nell'intero corpus
 - Conteggio = frequenza assoluta
 - Tendenze = grafico con la distribuzione delle frequenze relative tra i documenti del corpus
- Per la stessa analisi, ma a livello di documento: Termini del documento (Document Terms)

			Parola	Conteggio	Tendenza
<input checked="" type="checkbox"/>	<input type="checkbox"/>	1	renzo	561	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	2	disse	560	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	3	don	442	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	4	lucia	391	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	5	tempo	375	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	6	dire	340	

Strumenti: COLLOCATES / COLLOCAZIONI

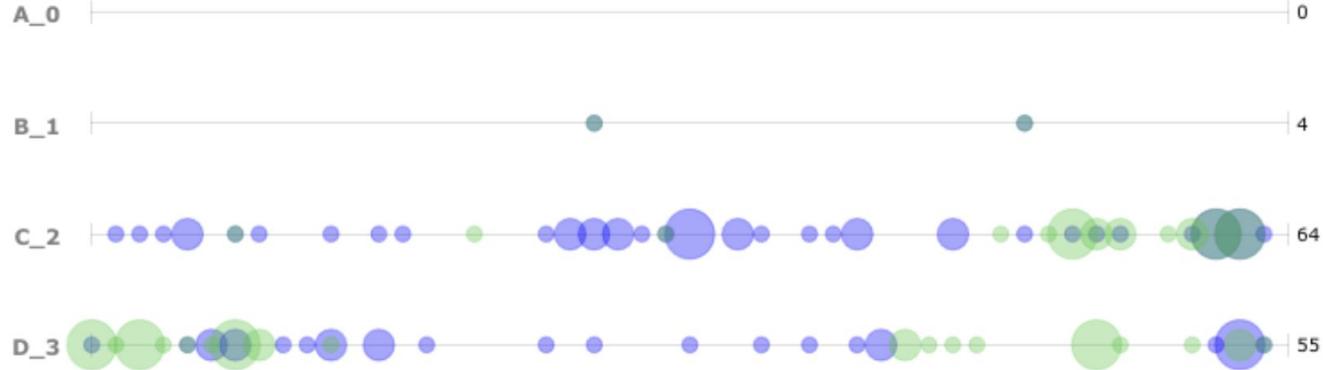
- Mostra quali termini compaiono più frequentemente in prossimità della parola cercata
 - Per impostazione predefinita, i collocati più frequenti in un contesto di 5 parole vengono mostrati per le 10 parole chiave più frequenti nel corpus
 - Il contesto (numero di parole da considerare su ciascun lato della parola chiave) può essere modificato utilizzando un cursore

	Parola	Collocazione	Conteggio (contesto)
<input type="checkbox"/>	don	abbondio	227
<input type="checkbox"/>	don	rodrigo	163
<input type="checkbox"/>	padre	cristoforo	97

Strumenti: BUBBLELINES

- Visualizza la frequenza e la distribuzione dei termini in un corpus
- Mostra i punti del testo in cui i termini sono menzionati

renzo lucia



Strumenti: TOPICS

- Esegue la modellazione degli argomenti secondo l'algoritmo LDA (Latent Dirichlet Allocation)
 - 10 termini per argomento
 - Dettagli per documento

Topics

don lucia disse abbondio agnese donna signor detto casa curato

pane grida milano ferrer casa peste vicario numero effetto generale

dire porta gran nome giovane donne tempo momento piedi dice

padre voce uomo poteva cristoforo signore tratto gran detto conte

mezzo modo buon dire vedere povero dico pensiero caso mondo

renzo rodrigo mano disse casa frate notte voglia capo griso

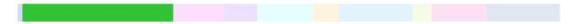
ZF_30



ZG_31



ZH_32



ZI_33



ZL_34



Esempi di utilizzo sui *Promessi Sposi*

1. Come individuare gli hapax legomena? → TERMS
2. Quali sono le parole più frequenti del cap. 32? → CIRRUS + SCALE
3. Quali sono le parole più caratteristiche del cap. 10? --> DISTINCTIVE WORDS (in SOMMARIO)
4. Renzo e Lucia appaiono sempre insieme? --> ANDAMENTI, BUBBLELINES, MICROSEARCH
5. Qual è il capitolo con più densità lessicale? → DOCUMENTI
6. Come individuare gli usi metaforici di una parola? CONTESTI
7. Come individuare forme desuete? → CONTESTI (es. cercare *is** oppure **j**)

Esempi di utilizzo sulle tre redazioni

- Caricate i 3 file della cartella **redazioni**, caricate la lista di stopwords e provate a rispondere alle seguenti domande:
 1. Ci sono differenze evidenti tra i 3 testi dal punto di vista quantitativo?
 2. La presenza di Lucia è simile nelle 3 versioni?
 3. Il personaggio di Geltrude appare in "Fermo e Lucia" similmente a come appare Gertrude ne "I Promessi Sposi"?
 4. Quante parole contengono la lettera "j" nelle varie versioni?
 5. Cosa notate nella frequenza di "egli"?

Altre risorse del portale Manzoni Online

- Sezione *Strumenti* → sotto-sezione *Visualizzare* sul portale <https://www.alessandromanconi.org/>
 - Video tutorial su Voyant
 - Canale YouTube: <https://www.youtube.com/@ManzoniOnline>
- Sezione *Strumenti* → sotto-sezione *Altre attività didattiche*
 - Spunti per lezioni: tecniche attive di insegnamento
 - Escape room



Altre risorse: spunti per lezioni

1. Epistolarità da Manzoni ai **social**: esercizio di scrittura e riscrittura

- Lettura e comprensione di una lettera di Manzoni
- Risposta in stile ottocentesco
- Riscrittura in italiano contemporaneo
- Riscrittura come su WhatsApp

Manzoni al figlio Pietro, 1840

Avrò caro che mi si mandino le tavolette di mano in mano che saran fatte, per poter presto proseguire il lavoro mio.



*Pietro, mandami i file appena pronti! Senza di quelli sono bloccato 😱!
DAAAIIIIIIII!*



Altre risorse: spunti per lezioni

2. Alfabetizzazione informativa: dalla biblioteca di Manzoni alla biblioteca scolastica

- Cosa leggeva Manzoni?
- Quali dizionari possedeva?
- Come possono essere classificati e catalogati i dizionari? → creazione collaborativa di schede a partire da quelle del portale

MANZ. 16. 0205 [Postillato] 
Milano, Biblioteca Nazionale Braidense

Profilo [Postille \(150\)](#)

{ Accademia della Crusca }

{ *Vocabolario degli Accademici della Crusca oltre le giunte fatteci finora, cresciuto d'assai migliaia di voci e modi de' classici, le piu trovate da veronesi ... Tomo primo [settima] 1: A-B*
Verona : dalla stamperia di Dionigi Ramanzini, 1806
XXXII, 409, [3] p

Lingua: italiano

Osservazioni sull'esemplare Legatura in mezza pelle e carta marmorizzata. - Postille mss. di A. Manzoni alle pp. XI, XIII, XIX e in gran parte del testo.

Fa parte di

- [MANZ. 16. 0205-0211: Vocabolario degli Accademici della Crusca oltre le giunte fatteci finora, cresciuto d'assai migliaia di voci e modi de' classici, le piu trovate da veronesi ... Tomo primo \[settima\]](#) [Postillato]

Esemplari dello stesso insieme

- [MANZ. 16. 0205: 1: A-B](#) [Postillato]
- [MANZ. 16. 0206: 2: C-D](#) [Postillato]
- [MANZ. 16. 0207: 3: E-F](#) [Postillato]
- [MANZ. 16. 0208: 4: K-O](#) [Postillato]
- [MANZ. 16. 0209: 5: P-R](#) [Postillato]
- [MANZ. 16. 0210: 6: S-T](#) [Postillato]
- [MANZ. 16. 0211: 7: V-Z](#) [Postillato]



Vedi immagini (456)
[Riproduzione completa]

La Sentiment Analysis in classe

- Riconoscere e classificare automaticamente opinioni, emozioni e valutazioni personali espresse nei testi

Amo questo romanzo

POSITIVO

Questo romanzo non mi
piace!

NEGATIVO

Manzoni è morto a Milano
nel 1873

NEUTRO

- Principalmente usato per monitorare le opinioni degli utenti sui social network e sui siti di recensioni, come può essere applicato allo studio della letteratura?
- Integrazione alla didattica tradizionale:
 - Rendere la lettura più coinvolgente e personale
 - Educare alle emozioni
 - Usare in modo critico strumenti digitali

La Sentiment Analysis in classe

- Annotare manualmente il testo a livello di frase: come sono le emozioni evocate dall'autore nella frase? Prevalentemente positive, negative, miste o neutre (assenti)?

Entrati che furono, il padre Cristoforo riaccostò la porta adagio adagio.	NEUTRE
Allora il sagrestano non poté più reggere, e, chiamato il padre da una parte, gli andava susurrando all'orecchio: «ma padre, padre! di notte... in chiesa... con donne... chiudere... la regola... ma padre!»	NEGATIVE

- Chiedere a ChatGPT (Gemini, Claude...) di classificare le stesse frasi
 - Quale prompt usare? Prompt diversi danno risultati diversi?
- Confrontare annotazione manuale e predizione automatica: quanto è stato bravo il sistema automatico?
 - Analisi critica dei risultati e delle potenzialità dei sistemi