# Linguistica computazionale

## Prof. Francesco Mambrini

***OBIETTIVO DEL CORSO E RISULTATI DI APPRENDIMENTO ATTESI***

L’insegnamento si propone di fornire agli studenti un’introduzione ai principali concetti e metodi della Linguistica Computazionale (LC), nonché di offrire un primo orientamento al mondo delle risorse linguistiche digitali, in particolare testuali. Il corso sarà articolato in una parte teorica e in una serie di lezioni di carattere più illustrative e pratiche. Nella parte teorica, dopo un’introduzione dedicata alla definizione e storia della LC, il corso si concentrerà sulla nozione di *corpus* e sulla rappresentazione digitale dell’informazione testuale. Verranno illustrate alcune tecniche fondamentali per l’esplorazione e l’elaborazione computazionale dei dati testuali (concordanze, espressioni regolari, collocazioni). Infine, si farà cenno ai principali compiti di Trattamento Automatico del Linguaggio (o “Natural Language Processing”), e si introdurranno alcuni modelli linguistici sviluppati per risolvere task quali l’etichettatura delle Parti del Discorso o la “Sentiment Analysis”. Le lezioni verranno accompagnate da dimostrazioni pratiche volte a illustrare come tali operazioni di esplorazione o annotazione del testo possono essere eseguite utilizzando il linguaggio di programmazione Python.

Al termine dell’insegnamento, lo studente:

* conoscerà le principali caratteristiche dell’informazione testuale digitale;
* saprà come semplici operazioni di elaborazione automatica dei testi possano essere svolte utilizzando Python;
* saprà individuare le metodologie computazionali di trattamento del testo più adatte ai diversi tipi di analisi linguistica.

***PROGRAMMA DEL CORSO***

1. Introduzione alla linguistica computazionale: definizione, obiettivi, cenni storici.

2. La nozione di *corpus*, tra Linguistica dei Corpora e Linguistica Computazionale.

3. Il testo digitale: la codifica dell’informazione testuale.

4. Imparare a esplorare il testo digitale: espressioni regolari, concordanze e collocazioni.

5. Annotare un testo: introduzione al Trattamento Automatico del Linguaggio.

6. Apprendimento automatico e modelli linguistici.

7. Introduzione ai task del Trattamento Automatico del Linguaggio.

***BIBLIOGRAFIA[[1]](#footnote-1)***

Jezek E. e Sprugnoli R.. (2023). *Linguistica Computazionale. Introduzione all’analisi automatica dei testi*. Bologna, il Mulino. **Obbligatori:** Capitoli 1, 6, 7. **Facoltativi**: Capitoli 4, 8. [*Acquista da VP*](https://librerie.unicatt.it/scheda-libro/elisabetta-jezek-rachele-sprugnoli/linguistica-computazionale-introduzione-allanalisi-automatica-dei-testi-9788815290359-724386.html)

Nissim M. e Pannitto L. (2022). *Che cos’è la linguistica computazionale*. Roma, Carocci. **Facoltativo**. [*Acquista da VP*](https://librerie.unicatt.it/scheda-libro/malvina-nissim-ludovica-pannitto/che-cose-la-linguistica-computazionale-9788829013418-709199.html)

Lenci A., Montemagni S., Pirrelli V. (2016). *Testo e Computer. Elementi di linguistica computazionale*. Roma, Carocci (2° edizione). **Facoltativi**: Capitoli 1, 7. [*Acquista da VP*](https://librerie.unicatt.it/scheda-libro/alessandro-lenci-simonetta-montemagni-vito-pirrelli/testo-e-computer-elementi-di-linguistica-computazionale-9788843083060-241751.html)

***DIDATTICA DEL CORSO***

Le lezioni avranno carattere sia teorico sia pratico. Chi dispone di un computer portatile è invitato a servirsene durante le lezioni.

***METODO E CRITERI DI VALUTAZIONE***

La prova finale consisterà in un esame orale. Il colloquio avrà lo scopo di verificare l’acquisizione dei contenuti presentati a lezione e la capacità di utilizzare le nozioni apprese per risolvere alcuni esercizi pratici relativi all’analisi computazionale di dati testuali.

In particolare, le domande dell’esame verteranno su: a) aspetti teorici relativi a testi digitali, corpora testuali, strumenti e metodi per la loro analisi; b) scenari pratici, quali ad esempio: 1. la scelta dello strumento corretto per estrarre un certo tipo di informazioni da un testo, 2. la progettazione di un ipotetico scenario proposto dal docente (raccolta di dati, scelta degli strumenti, elaborazione dei dati) per svolgere un’attività di ricerca linguistica sui corpora.

Entrambe le tipologie di domande avranno pari peso ai fini della valutazione finale.

## ***AVVERTENZE E PREREQUISITI***

Il corso ha un carattere introduttivo; non è, pertanto, richiesta alcuna conoscenza preliminare di tipo informatico. Una generica familiarità con il computer è sufficiente.

*Orario e luogo di ricevimento*

Su appuntamento, inviando preventivamente una mail a francesco.mambrini@unicatt.it, presso l’ufficio 219, 2° piano dell’edificio Franciscanum. Alla fine delle lezioni, se in presenza.

1. I testi indicati nella bibliografia sono acquistabili presso le librerie di Ateneo; è possibile acquistarli anche presso altri rivenditori. [↑](#footnote-ref-1)