# Linguistica computazionale

## Prof. Francesco Mambrini

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

L’insegnamento si propone di fornire agli studenti le nozioni di base della Linguistica Computazionale, nonché di offrire un primo orientamento al mondo delle risorse linguistiche, in particolare testuali. Il corso sarà articolato in tre parti. Dopo una sezione introduttiva dedicata alla definizione e storia della Linguistica Computazionale, il corso si concentrerà sulla nozione di *corpus* e sulla rappresentazione digitale dell’informazione testuale. Nella seconda parte, verranno illustrate alcune tecniche fondamentali per il trattamento e l’elaborazione computazionale dei dati testuali. Infine, verranno discussi alcuni metodi basilari di esplorazione del testo (concordanze, collocazioni) e si farà cenno alle principali tecniche di Trattamento Automatico del Linguaggio (o “Natural Language Processing”).

Al termine dell’insegnamento, lo studente:

* conoscerà le principali caratteristiche dell’informazione testuale digitale;
* sarà in grado di eseguire semplici operazioni di elaborazione automatica dei testi attraverso espressioni regolari e programmi a linea di comando (*wc*, *grep*, *cut, tr* e altri);
* 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* a cavallo tra Linguistica dei Corpora e Linguistica Computazionale.

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

4. Imparare a trattare il testo digitale: le espressioni regolari e alcuni programmi eseguibili a linea di comando.

5. L’analisi quantitativa dei testi: frequenze e distribuzioni delle parole.

6. Esplorare un corpus: concordanze e collocazioni.

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

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

Lenci A.-Montemagni S. & Pirrelli V. (2016). *Testo e computer. Introduzione alla linguistica computazionale*. Roma, Carocci (seconda edizione). **Obbligatori:** capitoli 1, 2. **Facoltativi:** capitoli 4, 5, 7, 8. [*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 al secondo piano del palazzo Franciscanum. A distanza, su Teams, inviando preventivamente una mail a francesco.mambrini@unicatt.it. 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)