

Linguistica e Filologia Digitale

Simone Ciccolone

L.M. in Filologie e Letterature Classiche e Moderne

a.a. 2019/2020

Obiettivi formativi

Il corso si pone come obiettivo didattico generale quello di introdurre il campo disciplinare delle *Digital Humanities*, proponendo alcune riflessioni e strumenti di analisi in campo linguistico e filologico. In particolare,

- verranno presentate le riflessioni preliminari dell'analisi linguistica e filologica applicata al testo digitale "nativo" (*born-digital*, come blog, comunicazione mediata dal computer e *interactive fiction*) e alla trasposizione digitale del testo;
- si proporrà una riflessione critica riguardo alla computabilità del testo e alla sua trasposizione in dato analizzabile quantitativamente;
- verranno introdotti i principali strumenti di codifica digitale del testo (orale e scritto), evidenziando la multidimensionalità del dato linguistico e la complessità del processo interpretativo di un testo;
- verrà illustrato il flusso di lavoro necessario alla creazione ed elaborazione di un corpus, dalla raccolta dei testi all'annotazione, passando per la trascrizione di interazioni dialogiche orali;
- verranno presentati gli strumenti di analisi quantitativa del testo, dalle basi della lessicostatistica all'individuazione di *patterns* combinatori.

Al termine del corso, gli studenti avranno raggiunto i seguenti obiettivi formativi:

A. Conoscenze e capacità di comprensione (*knowledge and understanding*)

- Conoscenza dei principali strumenti di analisi quantitativa di dati linguistici;
- Conoscenze di base di linguistica dei corpora (creazione di un corpus, trascrizione, annotazione, ricerca tramite *query*, concordanze);
- Comprensione, a livello teorico, del funzionamento dei principali strumenti di elaborazione automatica del linguaggio;
- Capacità di comprensione dei problemi legati alla testualità digitale e di elaborazione di nuove soluzioni.

B. Applicazione di conoscenze e capacità di comprensione (*applying knowledge and understanding*)

- Capacità di applicare le competenze acquisite per la raccolta ed elaborazione di un piccolo corpus di testi;
- Abilità di base di codifica, trascrizione e annotazione di testi orali e scritti (trascrizione del parlato, lemmatizzazione, *POS tagging*);

- Capacità di ricerca selettiva avanzata all'interno di testi e corpora (*regular expressions*, estrazione di *patterns*);
- Capacità di elaborare progetti originali nel campo delle *Digital Humanities* e di proporre soluzioni innovative.

C. Autonomia di giudizio (*making judgements*)

- Riflessione critica del rapporto tra studi umanistici, digitalizzazione della cultura e strumenti di trasmissione e interpretazione della conoscenza e dell'informazione, ovvero del rapporto tra accesso ai contenuti e trattamento automatico dei testi, interpretazione individuale o collettiva e controllo automatico, privato o politico, delle fonti;
- Autonomia di giudizio riguardo all'applicazione di strumenti della riflessione linguistica e filologica alla realtà attuale o a fatti e prodotti culturali specifici;
- Autonomia nella ricerca di informazioni qualificate e nell'approfondimento delle tematiche riguardanti il campo delle *Digital Humanities*.

D. Abilità comunicative (*communication skills*)

- Capacità di presentazione, in forma orale o scritta, di dati estrapolati da corpora e di analisi dei testi;
- Capacità di discussione, in forma orale o scritta, dei temi trattati durante il corso;
- Capacità di comprensione di testi scientifici riguardanti temi di informatica umanistica.

E. Abilità di apprendimento (*learning skills*)

- Capacità di autoapprendimento all'uso di nuovi strumenti o "linguaggi" (in senso informatico) per l'elaborazione di dati linguistici e di testi;
- Capacità di estendere le competenze acquisite tramite approfondimenti teorici e applicazioni concrete nel campo interdisciplinare delle *Digital Humanities*.

Prerequisiti

Conoscenze di base di linguistica generale (flessione e derivazione; parti del discorso; funzioni sintattiche; struttura informativa); conoscenze di base dei metodi di critica testuale (principi di base del metodo del Lachmann); comprensione di testi scientifici in lingua italiana (possibilmente, anche in lingua inglese); competenze logico-matematiche di base; competenze informatiche di base (videoscrittura, elaborazione di tabelle). Ulteriori competenze informatiche, benché non rappresentino un prerequisito, potranno agevolare notevolmente la partecipazione alle attività del corso.

In caso di dubbi o carenze riguardo ai prerequisiti indicati, si consiglia di contattare il docente e di consultare i manuali di riferimento delle relative discipline. Si segnalano qui alcuni testi utili in merito:

- Berruto, Gaetano / Cerruti, Massimo (2016), *La linguistica. Un corso introduttivo*. UTET, Torino.
- Graffi, Giorgio / Scalise, Sergio (2013), *Le lingue e il linguaggio*. Il Mulino, Bologna.
- Cesarini Martinelli, Lucia (1994), *La filologia*. Editori Riuniti, Milano.

Per le competenze informatiche di base, si consultino i corsi online messi a disposizione dal Laboratorio di Cultura Digitale:

- <http://www.labcd.unipi.it/tutorial/diamoci-le-basi/>
-

Contenuti

Argomenti trattati

- Testo, testualità, interpretazione
- Modalità, dialogicità e medium
- *Computer Mediated Communication* (CMC)
- Fondamenti di lessicostatistica
- Fondamenti di linguistica dei corpora
- Corpora scritti e orali
- Frequenza lessicale e analisi distribuzionali
- Effetti di frequenza e sistema linguistico
- Trascrizione, annotazione, codifica di testi
- Codifica XML e rappresentazione multidimensionale del testo
- *Le Scholarly Digital Editions*
- Documentazione, accessibilità e riproducibilità nell'era digitale

Attività laboratoriali

- Ricerca di concordanze nel Corpus *Repubblica*
- Prova di trascrizione di 1 min. di parlato
- *POS tagging* in ELAN
- Analisi dello spettro di frequenze di un testo
- Codifica di un testo in formato ipertestuale

Alcune di queste attività laboratoriali saranno richieste in sede d'esame agli studenti frequentanti e potranno influire sulla valutazione finale.

Metodi didattici

Il corso è organizzato in 30 ore di lezione frontale, in cui sono incluse attività laboratoriali ed esercitazioni pratiche. Eventuali attività integrative potranno essere accessibili tramite la piattaforma di *e-learning*.

Verifica dell'apprendimento

Modalità d'esame

L'esame si svolge in forma orale, tramite una discussione sui contenuti del corso e sugli argomenti trattati nei testi di riferimento del programma d'esame. Per gli **studenti frequentanti**, l'esame orale includerà anche la revisione di una delle attività laboratoriali assegnate o concordate col docente; per gli **studenti non frequentanti**, l'esame orale includerà anche la discussione di un capitolo a scelta dello studente tra quelli indicati nel programma d'esame per non frequentanti.

Criteri di valutazione

La valutazione finale, espressa in trentesimi, terrà conto:

- della capacità dello studente di discutere dei temi trattati;
- delle potenziali capacità applicative emerse durante la discussione orale;
- per i frequentanti, delle capacità applicative dimostrate tramite le attività laboratoriali;

- per i non frequentanti, della capacità di riflessione critica e dell'autonomia di giudizio sviluppata sui temi trattati.

Testi di riferimento

Gli **studenti frequentanti** sono invitati a integrare i materiali didattici presentati a lezione con lo studio del manuale di riferimento:

- Lenci, Alessandro / Montemagni, Simonetta / Pirrelli, Vito (2005), Testo e computer. Elementi di linguistica computazionale. Carocci, Roma (escluso il cap. 6).

Gli **studenti non frequentanti** sono tenuti a sostituire la partecipazione alle lezioni con lo studio del manuale di riferimento indicato sopra (escluso il cap. 6) e ad integrare il programma d'esame con un approfondimento a scelta tra le due opzioni seguenti:

- [LINGUISTICA] Tavosanis, Mirko (2018), Lingue e intelligenza artificiale. Carocci, Roma (capp. 1-2 più un capitolo a scelta tra 3-6).
- [FILOLOGIA] Fiormonte, Domenico (2018), Per una critica del testo digitale. Letteratura, filologia e rete. Bulzoni, Roma (capp. 1-2 più un capitolo a scelta tra 5-8).

Lo studio dei testi di riferimento indicati è indispensabile per la verifica finale dell'apprendimento. Eventuali letture alternative possono essere discusse col docente e devono essere approvate in tempo utile prima della data d'esame.

Altre informazioni

Materiali didattici verranno distribuiti a lezione e tramite la piattaforma di *e-learning*. Nel corso delle lezioni e delle attività laboratoriali si rimanderà a una serie di strumenti e risorse di libero accesso disponibili online. Gli studenti possono prendere visione di tali risorse tramite i seguenti link:

Strumenti software da scaricare e installare:

- ELAN (<https://tla.mpi.nl/tools/tla-tools/elan/>)
- SublimeText (<https://sublimetext.com>)
- TextStat (<http://neon.niederlandistik.fu-berlin.de/en/textstat/>)
- Simple Concordance Program (<http://www.textworld.com/scp/>)

Strumenti *web-based* (utilizzabili online):

- Textalyser (<http://textalyser.net/>)
- Indice Gulpease (<http://www.corrige.it/prova-gratis/>)

Corpora disponibili online (tramite la piattaforma *NoSketchEngine*):

- Corpus Repubblica (<https://corpora.djpintra.it/>)
- Corpus KIParla (<https://kiparla.it>)

Gli studenti non frequentanti devono mettersi in contatto col docente ALMENO due mesi prima del loro esame per concordare il programma.

Data la natura particolarmente tecnica e specialistica del corso, gli studenti sono caldamente invitati a contattare il docente tramite e-mail (simone.ciccolone@unica.it) e a partecipare agli incontri di ricevimento per qualsiasi chiarimento sui contenuti e sulle modalità d'esame.

