

Linguistica e filologia digitale

Scheda corso (a.a. 2024/25)

Simone Ciccolone

Obiettivi

Il corso si pone come obiettivo didattico generale quello di introdurre gli ambiti tematici e gli strumenti di base della linguistica dei corpora e della filologia digitale, proponendo alcune riflessioni, attività laboratoriali e di analisi nel campo delle scienze del linguaggio e della filologia contemporanea. In particolare,

- verranno presentate le riflessioni preliminari dell'analisi linguistica e filologica applicata al testo "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)
 - A1. Conoscenze di base di linguistica dei corpora (definizione di corpus, trascrizione, annotazione, ricerca tramite query, concordanze);
 - A2. Comprensione, a livello teorico, del funzionamento degli strumenti di base di elaborazione automatica del linguaggio;
 - A3. Capacità di comprensione dei problemi legati alla testualità digitale e di elaborazione di nuove soluzioni.
- (B. Applicazione di conoscenze e capacità di comprensione)
 - B1. Capacità di applicare le competenze acquisite per la raccolta ed elaborazione di un piccolo corpus di testi;
 - B2. Abilità di base di codifica, trascrizione e annotazione di testi orali e scritti (trascrizione del parlato, rappresentazione di documenti scritti, tokenizzazione, lemmatizzazione, *POS tagging*);
 - B3. Capacità di ricerca selettiva avanzata all'interno di testi e corpora (*regular expressions*, estrazione di *patterns*);
 - B4. Capacità di elaborare progetti originali nel campo della linguistica e filologia digitale e di proporre soluzioni innovative.

- (C. Autonomia di giudizio)
 - C1. Riflessione critica del rapporto tra studi umanistici, digitalizzazione della cultura e strumenti di trasmissione e interpretazione della conoscenza e dell'informazione;
 - C2. Autonomia di giudizio riguardo all'applicazione di strumenti della riflessione linguistica e filologica alla realtà attuale o a fatti e prodotti culturali specifici;
 - C3. Autonomia nella ricerca di informazioni qualificate e nell'approfondimento delle tematiche riguardanti il campo delle *Digital Humanities*.
- (D. Abilità comunicative)
 - D1. Capacità di presentazione, in forma orale o scritta, di dati estrapolati da corpora e di analisi di testi;
 - D2. Capacità di discussione, in forma orale o scritta, dei temi trattati durante il corso;
 - D3. Capacità di comprensione di testi scientifici riguardanti temi di umanistica digitale.
- (E. Abilità di apprendimento)
 - E1. Capacità di autoapprendimento all'uso di nuovi strumenti per l'elaborazione di dati linguistici e di testi;
 - E2. 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 (collazione, *stemma codicum*, 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.
- Beltrami, Pietro (2017), *La filologia romanza*. Il Mulino, Bologna.
- Stussi, Alfredo (1994), *Introduzione agli studi di filologia italiana*. Il Mulino, Bologna.

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

www.labcd.unipi.it/tutorial/diamoci-le-basi/

Contenuti

Argomenti trattati:

- Testo, testualità, interpretazione
- 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
- Codifica TEI
- Fondamenti di filologia digitale
- Le *Scholarly Digital Editions*
- Testualità digitale e intermedialità

Attività laboratoriali:

- Ricerca di concordanze nel Corpus Repubblica
- Prova di trascrizione di 1 min. di parlato
- Tokenizzazione e *POS tagging* di una porzione di testo
- Analisi dello spettro di frequenze di un testo
- Analisi dei collocati di una parola nel corpus *Repubblica*
- Codifica di un testo in formato ipertestuale
- Creazione di voci di glossario o wiki su temi legati al corso

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. Gli studenti frequentanti sono poi invitati a svolgere attività laboratoriali, individuali o in piccoli gruppi, da consegnare via e-mail o presentare di persona durante il periodo delle lezioni.

Eventuali attività integrative potranno essere accessibili tramite il canale Teams del corso. Inoltre, le esercitazioni potranno essere svolte. Gli studenti (in particolare coloro che intendono sostenere l'esame da frequentanti) sono invitati ad iscriversi al **canale Teams**, usando il codice: **68vrn45**.

Verifica dell'apprendimento

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 delle attività laboratoriali assegnate o concordate col docente; per gli studenti non frequentanti, l'esame orale includerà anche la discussione degli approfondimenti indicati nel programma d'esame per non frequentanti.

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

Gli studenti frequentanti dovranno integrare i materiali didattici presentati a lezione con lo studio dei seguenti manuali di riferimento:

- Freddi, Maria (2014), *Linguistica dei corpora*. Carocci, Roma.
- Mancinelli, Tiziana / Pierazzo, Elena (2020), *Che cos'è un'edizione scientifica digitale*. Carocci, Roma.

Gli studenti non frequentanti sono tenuti a sostituire la partecipazione alle lezioni con lo studio dei manuali di riferimento indicati sopra e ad integrare il programma d'esame con il seguente testo:

- Lenci, Alessandro / Montemagni, Simonetta / Pirrelli, Vito (2005), *Testo e computer. Elementi di linguistica computazionale*. Carocci, Roma (esclusi capp. 3 e 6).

e con uno degli approfondimenti indicati di seguito, a scelta dello studente:

- Numerico, Teresa / Fiormonte, Domenico / Tomasi, Francesca (2010), *L'umanista digitale*. Il Mulino, Bologna (un capitolo a scelta tra 2-4);
- Tavosanis, Mirko (2018), *Lingue e intelligenza artificiale*. Carocci, Roma (capitoli 1-2 e un capitolo a scelta tra 3-6);
- Tomasin, Lorenzo (2017), *L'impronta digitale. Cultura umanistica e tecnologia*. Carocci, Roma (capitoli 1-3 e un capitolo a scelta tra 4-7).

Indipendentemente dalla effettiva partecipazione alle lezioni, ogni studente può scegliere se completare il programma per frequentanti oppure il programma per non frequentanti. Tuttavia, per sostenere l'esame col programma per frequentanti è necessario concordare la consegna delle attività laboratoriali direttamente col docente e ricevere una valutazione preventiva degli elaborati consegnati almeno due settimane prima della data d'esame. Gli studenti che intendono presentare il programma per frequentanti sono quindi caldamente invitati a partecipare alle lezioni e agli incontri laboratoriali.

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.

Il programma d'esame con l'elenco completo dei testi e le indicazioni sulle attività laboratoriali, nonché ulteriori dettagli sui materiali del corso, saranno resi disponibili tramite i canali di web learning del corso (**elearning.unica.it** e MS Teams) e sulla **pagina personale del docente** (https://unica.it/unica/it/ateneo_s07_ss01.page?contentId=SHD190002).

Altre Informazioni

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:

- *SublimeText* (<https://sublimetext.com>)
- *ELAN* (<https://tla.mpi.nl/tools/tla-tools/elan/>)
- *AntConc* (<https://www.laurenceanthony.net/software/antconc/>)

Corpora disponibili online:

- *Corpus Repubblica* (<https://corpora.dipintra.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, tutti 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.

(Simone Ciccolone – 23/9/2024)