



Università
per Stranieri
di Perugia

Anno Accademico 2022-2023

PROGRAMMA D'ESAME

Laurea: **Digital Humanities per l'italiano (DHI)**

Insegnamento: **Linguistica dei corpora - Laboratorio di codifica dei testi**

Anno di corso: **III**

Semestre: **II**

Docenti: **Luciana Forti, Irene Fioravanti, Fabio Zanda**

SSD: **L-LIN/01**

CFU: **9**

Carico di lavoro globale: **225 ore**

Ripartizione del carico di lavoro: **60 ore di lezione e 165 ore di studio individuale**

Lingua di insegnamento: **Italiano**

PREREQUISITI

Sono necessarie una buona conoscenza della lingua e della grammatica italiana, e conoscenze di base nel campo della Linguistica generale. In particolare:

- livelli di analisi delle lingue: fonologia, morfologia, sintassi, lessico;
- principali unità di analisi linguistica: fonema, morfema, lessema, sintagma, frase, testo.

È utile, in tal senso, aver già sostenuto l'esame di Linguistica generale. Eventuali lacune potranno essere recuperate facendo riferimento ai seguenti testi:

- Berruto, B., Cerruti, M. (2017). La linguistica. Un corso introduttivo, Seconda Edizione, UTET Università.

Gli studenti possono comunque utilizzare lo strumento di LOL per interagire con il docente e chiarire eventuali dubbi sulle proprie conoscenze disciplinari di base.

È inoltre necessaria una comprensione della lingua italiana orale e scritta almeno al livello B2 del QCER. Gli studenti stranieri devono, se necessario e in particolare se segnalato nei colloqui in ingresso, avvalersi del servizio di tutorato linguistico offerto dall'Ateneo.

OBIETTIVI FORMATIVI

Conoscenza e comprensione:

- conoscere le principali tappe di sviluppo, i metodi e gli strumenti di analisi della Linguistica dei corpora;
- comprendere le opportunità legate all'uso di corpora linguistici nel campo della ricerca linguistica;
- conoscere il modo in cui i dati linguistici estratti da corpora, le risorse e le metodologie usate dalla linguistica dei corpora sono usati per analizzare l'uso linguistico, anche come supporto alla didattica delle lingue;
- conoscere gli standard, le metodologie e gli strumenti necessari all'elaborazione e alla codifica di dati linguistici, prodotti da parlanti nativi e da apprendenti.

Applicazione della conoscenza e della comprensione:

- essere in grado di analizzare e valutare dati estratti da corpora, attraverso le metodologie e le risorse della Linguistica dei corpora;

- saper utilizzare le metodologie e le risorse necessarie ad annotare dati linguistici secondo gli standard internazionali di codifica.

CONTENUTO DEL CORSO

Il corso è articolato in tre sezioni:

Nella prima sezione viene delineata la posizione della Linguistica dei corpora all'interno delle Scienze del linguaggio, e ne vengono descritte le caratteristiche salienti. Si introduce e si problematizza la nozione di corpus in linguistica, e si affrontano due aspetti centrali della disciplina: il trattamento di dati linguistici (codifica testuale e annotazione linguistica), e la loro successiva analisi. A partire dall'interrogazione di corpora, vengono descritte alcune metodologie di analisi di dati linguistici strutturati, basate su misure di tipo quantitativo (frequenza, dispersione, misure di associazione e di ricchezza e densità lessicale, analisi per keywords) e qualitativo (analisi di concordanze). Alcuni casi di studio proposti riguardano un fenomeno lessico-grammaticale rilevante in molti ambiti applicativi, tra cui quello della didattica delle lingue: le combinazioni lessicali. Un approfondimento finale verterà sui corpora e l'apprendimento linguistico, con rimando alla Learner Corpus Research. (20 ore)

La seconda sezione è finalizzata a coinvolgere gli studenti in attività pratiche di codifica testuale di dati. Particolare enfasi viene posta sulle fasi di progettazione di un *learner corpus* e sulla codifica di testi di apprendenti. Inoltre, si approfondisce il tema dell'annotazione di dati linguistici e della strutturazione di documenti XML. Infine, si introducono le pratiche di etichettatura e analisi degli errori, con particolare riferimento alle combinazioni lessicali. (20 ore)

La finale sezione affronta il tema dell'uso di corpora nella didattica dell'italiano. In particolare, si fornisce un'introduzione al Data-driven learning (DDL), facendo riferimento ai principi pedagogici a cui si ispira e alle teorie linguistiche che ne hanno determinato la nascita. Si traccia, inoltre, una storia degli studi e delle sperimentazioni relative al DDL, con particolare riferimento all'Italiano L2. Infine, vengono dimostrate e realizzate varie tipologie di attività DDL per l'apprendimento delle combinazioni lessicali in Italiano L2. (20 ore)

METODI DIDATTICI

Per studenti frequentanti

Lezioni frontali e attività applicative, da integrare con lo studio progressivo dei testi di riferimento e dei materiali proposti sulla piattaforma LOL (<https://lol.unistrapg.it/>).

Nel caso in cui le condizioni generali relative alla pandemia lo richiedano, saranno adottate modalità di didattica blended o totalmente a distanza, anche grazie alle piattaforme digitali a disposizione.

Per studenti non frequentanti

Studio individuale dei testi di riferimento, col supporto delle dispense e dei materiali scaricabili dalla piattaforma LOL (<https://lol.unistrapg.it/>).

La docente è disponibile a seguire gli studenti non frequentanti anche attraverso gli strumenti telematici presenti nella stessa piattaforma (es. forum di discussione).

METODI DI ACCERTAMENTO

L'esame si svolgerà in modalità orale. Gli studenti dovranno dimostrare di conoscere e di aver compreso i contenuti del corso, anche attraverso la loro applicazione concreta. Un'attenzione particolare sarà rivolta all'uso adeguato della terminologia specifica che caratterizza la disciplina, nonché alla capacità di sintesi e di

rielaborazione critica dei contenuti del corso.

Per gli studenti con DSA, la cui certificazione sia depositata presso la Segreteria Studenti, sono previste misure compensative e/o dispensative. Le richieste saranno valutate caso per caso allo scopo di adattare il programma e le modalità d'esame alle singole esigenze. A tal fine è necessario contattare il docente con congruo anticipo, anche mediante la Commissione per le Differenze e l'Inclusione (differenze-inclusione@unistrapg.it).

TESTI DI RIFERIMENTO

Per studenti frequentanti:

- Freddi M. (2014). *Linguistica dei corpora*. Carocci.
- Sinclair J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Johns T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *English Language Research Journal*, 4, 1–16.
- Hadley G. (2002). An introduction to data-driven learning. *RELC Journal* 33(2), pp. 99-124.
- Materiali integrativi (slide presentate a lezione) messi a disposizione sulla piattaforma LOL (<https://lol.unistrapg.it/>).

Per studenti non frequentanti:

- Freddi M. (2014). *Linguistica dei corpora*. Carocci.
- Sinclair J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Johns T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *English Language Research Journal*, 4, 1–16.
- Hadley G. (2002). An introduction to data-driven learning. *RELC Journal* 33(2), pp. 99-124.
- Materiali integrativi (slide presentate a lezione) messi a disposizione sulla piattaforma LOL (<https://lol.unistrapg.it/>).

TESTI DI CONSULTAZIONE E APPROFONDIMENTO

- Tony McEnery & Andrew Hardie (2012), *Corpus Linguistics. Method, Theory and Practice*. Cambridge, Cambridge University Press.
- Stefanowitsch, A. 2020. *Corpus linguistics: A guide to the methodology*. Berlin: LanguageScience Press.
- Pierazzo E. (2005). *La codifica dei testi: un'introduzione*. Carocci.

ALTRE INFORMAZIONI

È richiesta l'iscrizione on-line al corso sulla piattaforma LOL.

È raccomandata una regolare frequenza.

Per ricevimenti, si potranno prendere accordi con i docenti tramite posta elettronica.

E-mail dei docenti: luciana.forti@unistrapg.it / irene.fioravanti@unistrapg.it / fabio.zanda@unistrapg.it