



UNIVERSITÀ DI PAVIA

Anno Accademico 2014/2015

LABORATORIO DI ANALISI DI DATI LINGUISTICI (C. P.)

Anno immatricolazione	2014/2015
Anno offerta	2014/2015
Normativa	DM270
SSD	L-LIN/01 (GLOTTOLOGIA E LINGUISTICA)
Dipartimento	DIPARTIMENTO DI STUDI UMANISTICI
Corso di studio	LINGUISTICA TEORICA, APPLICATA E DELLE LINGUE MODERNE
Curriculum	PERCORSO COMUNE
Anno di corso	1°
Periodo didattico	Primo Semestre (22/09/2014 - 22/12/2014)
Crediti	6
Ore	36 ore di attività frontale
Lingua insegnamento	ITALIANO
Tipo esame	ORALE
Docente	MAURI CATERINA - 6 CFU
Prerequisiti	E' richiesta la conoscenza della lingua inglese e delle nozioni di base della linguistica generale.
Obiettivi formativi	<p>Il corso ha come obiettivo quello di rendere gli studenti capaci di riconoscere, descrivere e analizzare il dato linguistico da una molteplicità di prospettive, attraverso l'uso dei principali strumenti di analisi a disposizione (corpora, lessici, concordance tools, database tipologici, ecc.).</p> <p>Al termine del corso, gli studenti saranno in grado di comporre e analizzare un campione di dati in vista di specifiche ricerche linguistiche, avendo acquisito la capacità di selezionare gli strumenti di analisi e le risorse linguistiche in funzione del tipo di ricerca da svolgere.</p>
Programma e contenuti	<p>Il corso ha come obiettivo quello di rendere gli studenti capaci di riconoscere, descrivere e analizzare il dato linguistico da una molteplicità di prospettive, attraverso l'uso dei principali strumenti di</p>

analisi a disposizione (corpora, lessici, concordance tools, database tipologici, ecc.).

E' prevista un'articolazione in tre parti:

1) "Dato linguistico": individuazione, descrizione, analisi. La creazione e l'uso di corpora.

Nella prima parte del corso si discuterà la nozione di 'dato linguistico', trattando le principali problematiche metodologiche e teoriche legate alla descrizione, all'analisi e alla spiegazione dei dati linguistici. Dopo un'introduzione generale sugli approcci qualitativi e quantitativi, verranno presentate le nozioni di base per la preparazione e l'analisi di corpora, descrivendo le principali tecniche di elicitazione dei dati, i fondamenti della Corpus Linguistics, i principali tipi di corpora e il loro ruolo centrale in diversi ambiti di ricerca linguistica (ricerche di semantica e pragmatica, analisi di fenomeni morfosintattici, analisi sociolinguistiche).

2) Corpora e database: tipologia e descrizione delle principali risorse linguistiche.

Nella seconda parte del corso verranno presentate le principali risorse disponibili in relazione a diversi filoni di ricerca: risorse linguistiche per la tipologia, la linguistica storica e contrastiva, e l'acquisizione di lingue seconde, illustrando le potenziali applicazioni di concordancing tools, lessici, database e corpora nell'analisi di dati linguistici. Si discuterà anche in che modo si possa utilizzare il Web come corpus, sottolineando i limiti ma anche le potenzialità di un insieme di dati eterogeneo e in costante divenire. Infine, verranno discusse le principali problematiche legate alla creazione e all'uso di risorse linguistiche diversificate e verranno forniti i fondamenti di XML.

3) Laboratorio: analisi applicata di dati linguistici.

La parte finale del corso prevede un'applicazione pratica degli strumenti descritti nelle prime due parti: in una dimensione di laboratorio, gli studenti (individualmente o riuniti in piccoli gruppi) concordano con la docente un piccolo progetto su cui stenderanno poi una tesina, che sarà parte integrante dell'esame. Il progetto può consistere in:

- analisi di un fenomeno all'interno di un corpus
- analisi di un fenomeno per mezzo di uno o più strumenti appresi durante il corso,
- progettazione di un piccolo corpus,
- comparazione e/o recensione di diverse risorse linguistiche

Metodi didattici

Lezioni frontali ed esercitazioni pratiche.

La parte finale del corso prevede un'applicazione pratica degli strumenti descritti nel corso.

Testi di riferimento

Studenti frequentanti

La frequenza è fortemente consigliata, sia per la natura laboratoriale del corso sia perché non esiste un manuale di riferimento che tratti in modo sistematico le nozioni e gli strumenti che verranno discussi nel corso. La docente renderà settimanalmente disponibili online i materiali proiettati o distribuiti a lezione.

Qui di seguito vengono indicati preliminarmente alcuni testi e lavori a cui si farà riferimento durante il corso, si tenga presente che la lista che segue non è esaustiva. Al termine di ogni lezione, la docente indicherà le parti che sono state trattate e che andranno preparate in vista

dell'esame.

1) Evans, N. and A. Dench. "Introduction: Catching language". In Evans, N., A. Dench & F. Ameka (eds.), *Catching Language: The Art and Craft of Grammar Writing*, 1-39. Berlin: Mouton de Gruyter. 2006.

2) Dryer, Matthew S. "Descriptive theories, explanatory theories and Basic Linguistic Theory". In Evans, N., A. Dench & F. Ameka (eds.), *Catching Language: The Art and Craft of Grammar Writing*, 207-234. Berlin: Mouton de Gruyter. 2006.

3) Ide, Nancy. "Preparation and Analysis of Linguistic Corpora". In Susan Schreibman, Ray Siemens, John Unsworth (eds.), *A Companion to Digital Humanities*. Oxford: Blackwell. 2004.

<http://www.digitalhumanities.org/companion/>

4) Bradley, John "Text tools". In Susan Schreibman, Ray Siemens, John Unsworth (eds.), *A Companion to Digital Humanities*. Oxford: Blackwell. 2004.

<http://www.digitalhumanities.org/companion/>

5) Steven Bird, Gary Simons, "Seven dimensions of portability for language documentation and description", *Language* 79 (3), 2003, pp. 557-582.

6) McEnery?, Anthony, Richard Xiao and Yukio Tono. *Corpus-Based Language Studies: An advanced resource book*. London: Routledge, 2006.

7) Everaert, M., S. Musgrave, and A. Dimitriadis (eds). *The Use of Databases in Cross-Linguistic Studies*. Berlin: Mouton de Gruyter. 2009.

8) A. Kilgarriff and G. Grefenstette. Introduction to the Special Issue on the Web as Corpus. *Computational Linguistics* 29 (2003).

9) B. Turchetta. 2000. *La ricerca di campo in linguistica: metodi e tecniche d'indagine*. Roma: Carocci, 2000. (consultabile in biblioteca). Soprattutto Cap. 2-3.

10) Lüdeling, Anke; Evert, Stefan; Baroni, Marco. 2006. Using web data for linguistic purposes. In 'Language and Computers, Corpus Linguistics and the Web'. Edited by Marianne Hundt, Nadja Nesselhauf and Carolin Biewer , pp. 7-24(18). Rodopi Publisher.

Studenti non frequentanti

Gli studenti non frequentanti sono pregati di contattare con sufficiente anticipo la docente via mail per concordare il programma da portare all'esame.

Modalità verifica apprendimento

Studenti frequentanti

L'esame finale per i frequentanti consisterà in una prova orale in cui si discuteranno la tesina e l'attività di laboratorio/applicazione. Tale discussione costituirà un punto di partenza anche per domande volte a verificare la comprensione del percorso svolto durante il corso.

Studenti non frequentanti

Gli studenti non frequentanti sono pregati di contattare con sufficiente anticipo la docente via mail per concordare il programma da portare all'esame.

Altre informazioni

Studenti frequentanti

L'esame finale per i frequentanti consisterà in una prova orale in cui si discuteranno la tesina e l'attività di laboratorio/applicazione. Tale discussione costituirà un punto di partenza anche per domande volte a verificare la comprensione del percorso svolto durante il corso.

Studenti non frequentanti

Gli studenti non frequentanti sono pregati di contattare con sufficiente anticipo la docente via mail per concordare il programma da portare all'esame.

**Obiettivi Agenda 2030 per lo
sviluppo sostenibile**

[\\$bl legenda sviluppo sostenibile](#)