# Testi del Syllabus

| | | |
|---|---|---|
| Resp. Did. | **JEZEK ELISABETTA** | **Matricola: 007755** |

| | |
|---|---|
| Docente | **JEZEK ELISABETTA, 6 CFU** |

| | |
|---|---|
| Anno offerta: | **2018/2019** |
| Insegnamento: | **501166 - LABORATORIO DI ANALISI DI DATI LINGUISTICI (C. P.)** |
| Corso di studio: | **05409 - LINGUISTICA TEORICA, APPLICATA E DELLE LINGUE MODERNE** |
| Anno regolamento: | **2018** |
| CFU: | **6** |
| Settore: | **L-LIN/01** |
| Tipo Attività: | **B - Caratterizzante** |
| Anno corso: | **1** |
| Periodo: | **Secondo Semestre** |

## Testi in italiano

| | |
|---|---|
| **Lingua insegnamento** | ITALIANO |
| **Prerequisiti** | Nozioni di base di linguistica generale, in particolare morfologia, sintassi, semantica e pragmatica, come vengono fornite negli insegnamenti di laurea triennale di area umanistica. |
| **Obiettivi formativi e risultati di apprendimento** | Il corso ha l'obiettivo di rendere gli studenti capaci di raccogliere e analizzare dati linguistici da una molteplicità di prospettive, e conoscere le principali risorse linguistiche digitali a disposizione (corpora, lessici, concordance tools, banche dati, basi di conoscenza, datasets, ontologie, ecc.). Al termine del corso lo studente possiederà gli strumenti per progettare e condurre autonomamente un'analisi linguistica utilizzando metodologie basate prevalentemente sull'annotazione manuale o semiautomatica dei dati, allo scopo di estrarre o verificare generalizzazioni linguistiche per scopi teorici o applicativi. |
| **Programma e contenuti** | DATI E MODELLI PER RISORSE MULTILINGUE<br><br>Il corso di quest'anno introduce gli studenti alla varietà di dati disponibili per l'analisi linguistica (corpora, giudizi di accettabilità, dati elicitati, sperimentali, ecc.), concentrando l'attenzione sulle risorse multilingui.<br><br>Con l'ausilio di letture selezionate, sono esaminate la creazione, l'annotazione e la struttura di tali risorse e è sperimentato in laboratorio il loro utilizzo per l'analisi linguistica e le applicazioni nel trattamento automatico del linguaggio. |
| **Metodi didattici** | Lezioni frontali interattive<br>Slides<br>Incontri seminariali con presentazioni di gruppo delle letture e discussione<br>Laboratorio |
| **Testi di riferimento** | Letture<br><br>Baisa Vít, Može Sara and Irene Renau. 2016. "Multilingual CPA: Linking Verb Patterns across Languages." In: Margalitadze, Tinatin and George |

Meladze (eds) Proceedings of the XVII Euralex International Congress: Lexicography and Linguistic Diversity, pp. 410-417.

Boas, Hans C. 2005. "Semantic frames as interlingual representations for multilingual lexical databases." International Journal of Lexicography 18, no. 4, pp. 445-478.

Fellbaum Christiane, and Piek Vossen. 2012. "Challenges for a multilingual wordnet." Language Resources and Evaluation 46, no. 2, pp. 313-326.

Havasi Catherine, Speer Robert, and Jason Alonso. 2007. "ConceptNet 3: a flexible, multilingual semantic network for common sense knowledge." In Recent advances in natural language processing, pp. 27-29. Philadelphia, PA: John Benjamins.

Lenci Alessandro, Bel Nuria, Busa Federica, Calzolari Nicoletta, Gola Elisabetta, Monachini Monica, Ogonowski Antoine et al. 2000. SIMPLE: A general framework for the development of multilingual lexicons, International Journal of Lexicography 13, pp. 249–263.

Navigli, Roberto, and Simone Paolo Ponzetto. 2012. "BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network." Artificial Intelligence 193, pp. 217-250.

Pianta Emanuele, Bentivogli Luisa and Christian Girardi. 2002. MultiWordNet: Developing an aligned multilingual database, in: Proceedings of the 1st International Global WordNet Conference, pp. 21–25.

Ponti, Edoardo Maria, O'Horan Helen, Berzak Yevgeni, Vulić Ivan, Reichart Roi, Poibeau Thierry, Shutova Ekaterina and Anna Korhonen. 2018. "Modeling Language Variation and Universals: A Survey on Typological Linguistics for Natural Language Processing." arXiv preprint arXiv:1807.00914.

Ulteriori letture saranno indicate durante il corso e caricate sulla piattaforma Kiro.

**Modalità di verifica dell'apprendimento**

Prova orale di verifica dell’apprendimento dei contenuti del corso.
Discussione dell'indagine empirica di un fenomeno linguistico a scelta dello studente, concordato con la docente, utilizzando una delle risorse analizzare o creando ex novo una risorsa della stessa tipologia.
Elaborato scritto di 5 cartelle riportante ipotesi, scopo, metodologia e risultati dell'analisi empirica, da inviare a jezek@unipv.it al più tardi 7 gg prima della data dell'appello d'esame.

**Altre informazioni**

Tutto il materiale didattico - elenco aggiornato delle letture, slides delle lezioni, link a risorse linguistiche, istruzioni per l'elaborato finale - è disponibile sul portale della didattica KIRO (accesso con credenziali di Ateneo).

## Testi in inglese

Italian

Familiarity with basic notion in general linguistics, particularly morphology, syntax, semantics and pragmatics, as they are offered in the three-year Bachelor's degrees in Humanities.

The aim of the course is to provide the students with the knowledge and skills needed to collect and examine linguistic data from a variety of perspectives, and be acquainted with digital resources such as corpora, lexicons, concordance tools, databases, knowledge bases, datasets, and

| | |
|---|---|
| | ontologies. At the end of the course the students will be able to autonomously design and perform a linguistic analysis using methodologies primarily based on manual or semiautomatic annotation of data, with the goal of extracting or verifying linguistic generalizations for theoretical or applied purposes. |
| | DATA AND MODELS FOR MULTILINGUAL RESOURCES<br><br>This year's course introduces the students to the variety of data available for linguistic analysis (digital corpora, acceptability judgments, elicited data, experimental data, etc.), focusing the attention on multilingual resources.<br><br>With the help of selected readings, we examine the creation, the annotation and the structure of these resources and we use them in the lab for linguistic analysis and applications in natural language processing tasks. |
| | Face-to-fact interactive lectures<br>Slides<br>Seminars with group presentations of the readings and discussion<br>Lab |
| | Readings<br><br>Baisa Vít, Može Sara and Irene Renau. 2016. "Multilingual CPA: Linking Verb Patterns across Languages." In: Margalitadze, Tinatin and George Meladze (eds) Proceedings of the XVII Euralex International Congress: Lexicography and Linguistic Diversity, pp. 410-417.<br><br>Boas, Hans C. 2005. "Semantic frames as interlingual representations for multilingual lexical databases." International Journal of Lexicography 18, no. 4, pp. 445-478.<br><br>Fellbaum Christiane, and Piek Vossen. 2012. "Challenges for a multilingual wordnet." Language Resources and Evaluation 46, no. 2, pp. 313-326.<br><br>Havasi Catherine, Speer Robert, and Jason Alonso. 2007. "ConceptNet 3: a flexible, multilingual semantic network for common sense knowledge." In Recent advances in natural language processing, pp. 27-29. Philadelphia, PA: John Benjamins.<br><br>Lenci Alessandro, Bel Nuria, Busa Federica, Calzolari Nicoletta, Gola Elisabetta, Monachini Monica, Ogonowski Antoine et al. 2000. SIMPLE: A general framework for the development of multilingual lexicons, International Journal of Lexicography 13, pp. 249–263.<br><br>Navigli, Roberto, and Simone Paolo Ponzetto. 2012. "BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network." Artificial Intelligence 193, pp. 217-250.<br><br>Pianta Emanuele, Bentivogli Luisa and Christian Girardi. 2002. MultiWordNet: Developing an aligned multilingual database, in: Proceedings of the 1st International Global WordNet Conference, pp. 21–25.<br><br>Ponti, Edoardo Maria, O'Horan Helen, Berzak Yevgeni, Vulić Ivan, Reichart Roi, Poibeau Thierry, Shutova Ekaterina and Anna Korhonen. 2018. "Modeling Language Variation and Universals: A Survey on Typological Linguistics for Natural Language Processing." arXiv preprint arXiv:1807.00914.<br><br>Additional readings will be indicated in class and uploaded on the Kiro platform. |
| | Final oral exam covering the material from the entire course.<br>Final assignment (5 pages) reporting goal, research question, methodology and results of an in-depth corpus-based analysis of a linguistic phenomenon previously agreed during office hours. The text in |

| | |
|---|---|
| | pdf format must be sent to jezek@unipv.it 7 days before the exam. |
| | Material for the course - including the updated list of readings, the slides of the lectures, links to linguistic resources, instructions for the final assignment - are available on the KIRO platform (access with personal username and password). |