



# UNIVERSITÀ DI PISA

---

## LINGUISTICA COMPUTAZIONALE II

**SIMONETTA MONTEMAGNI**

Anno accademico 2019/20  
CdS INFORMATICA UMANISTICA  
Codice 513LL  
CFU 6

Moduli LINGUISTICA COMPUTAZIONALE II	Settore/i L-LIN/01	Tipo LEZIONI	Ore 36	Docente/i SIMONETTA MONTEMAGNI GIULIA VENTURI
--	-----------------------	-----------------	-----------	--

### Obiettivi di apprendimento

#### Conoscenze

Il corso si propone di introdurre lo studente a settori chiave della Linguistica Computazionale caratterizzati da un forte impatto applicativo. In particolare, si articola attorno a due macro-temi, l'annotazione linguistica multi-livello del testo e l'estrazione di conoscenza linguistica da basi documentali annotate, che sono affrontati da una duplice prospettiva, teorica e applicativa.

#### Capacità

Al termine del corso lo studente saprà utilizzare in modo critico e consapevole strumenti di annotazione linguistica automatica e di estrazione di conoscenza linguistica, identificare le problematiche legate al trattamento di varietà non-standard della lingua e ipotizzare possibili soluzioni.

#### Prerequisiti (conoscenze iniziali)

Nozioni di base di linguistica computazionale, di linguistica generale e di linguistica italiana.

#### Indicazioni metodologiche

Durante il corso si alterneranno lezioni frontali, con l'ausilio di slides powerpoint che vengono messe a disposizione degli studenti, ed esercitazioni di laboratorio, sia individuali sia di gruppo (svolte con PC delle aule informatiche e/o PC personali), in cui gli studenti sono invitati a confrontarsi con l'applicazione di strumenti software di annotazione linguistica del testo e di estrazione di conoscenza disponibili come demo online e ad analizzarne criticamente i risultati ottenuti in relazione a diverse varietà d'uso della lingua.

#### Programma (contenuti dell'insegnamento)

I contenuti del corso sono suddivisi in due macro-temi, per ciascuno dei quali segue una lista dei principali argomenti trattati:

- **Annotazione linguistica**
  - annotazione linguistica come processo incrementale; strumenti software per l'annotazione linguistica del testo; schemi di annotazione per l'annotazione morfo-sintattica e sintattica, con particolare attenzione allo schema delle "Universal Dependencies"; costruzione di corpora annotati e valutazione dell'annotazione; adattamento al dominio o altre varietà d'uso della lingua (es. storiche) di strumenti di annotazione;
- **Estrazione di conoscenza linguistica**
  - ricostruzione del profilo linguistico di collezioni di testi; monitoraggio linguistico di diverse tipologie testuali e/o varietà d'uso della lingua; uso dei risultati del monitoraggio linguistico all'interno di diversi scenari applicativi, ad esempio per la classificazione di generi testuali o per l'identificazione della lingua materna di produzioni L2; analisi della leggibilità del testo.

#### Bibliografia e materiale didattico

**Bibliografia essenziale del corso, che potrà subire variazioni e/o integrazioni sulla base del progetto che verrà definito per l'esame finale**

#### Annotazione linguistica del testo: letture generali

- Nivre, J. (2005) *Two Notions of Parsing*. In Arppe, A., Carlson, L., Lindén, K., Piitulainen, J., Suominen, M., Vainio, M., Westerlund,



## UNIVERSITÀ DI PISA

H. and Yli-Jyrä, A. (eds.) *Inquiries into Words, Constraints and Contexts*. Festschrift in the Honour of Kimmo Koskenniemi on his 60th Birthday. CSLI Publications, 106-115.

- Nivre, J. (2006) *Two Strategies for Text Parsing*. In Suominen, M., Arppe, A., Airola, A., Heinämäki, O., Miestamo, M., Määttä, U., Niemi, J., Pitkänen, K. K. and Sinnemäki, K. (eds.) *A Man of Measure: Festschrift in Honour of Fred Karlsson on his 60th Birthday*. Turku: The Linguistic Association of Finland.
- Alessandro Lenci, Simonetta Montemagni, Vito Pirrelli, 2009, Annotazione sintattica di corpora: aspetti metodologici, in Cecilia Andorno, Stefano Rastelli (a cura di), *Corpora di italiano L2: tecnologie, metodi, spunti teorici*, Perugia, Guerra Edizioni, pp. 25-46.
- Alessandro Lenci, Simonetta Montemagni, Vito Pirrelli 2005, *Testo e computer*, Carocci, Roma – Capitolo 8 Annotazione linguistica del testo

### Annotazione morfosintattica

- Venturi G. (2009). *Rassegna comparativa degli schemi di annotazione morfosintattica per la lingua italiana*, Technical report TRIPLE - RTT/1, February 2009.
- Slav Petrov, Dipanjan Das, and Ryan McDonald. 2012. *A universal part-of-speech tagset*. In *Proceedings of LREC 2012*.
- Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. 2011. *Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments*. In *Proceedings of ACL 2011*. Markus Dickinson and Marwa Ragheb (2009). *Dependency Annotation for Learner Corpora*. *Proceedings of the Eighth Workshop on Treebanks and Linguistic Theories (TLT-8)*. Milan, Italy.
- Christopher D. Manning. 2011. *Part-of-Speech Tagging from 97% to 100%: Is It Time for Some Linguistics?* In Alexander Gelbukh (ed.), *Computational Linguistics and Intelligent Text Processing, 12th International Conference, CICLing 2011, Proceedings, Part I*. *Lecture Notes in Computer Science 6608*, pp. 171--189. Springer.

### Annotazione sintattica

- Joakim Nivre. 2015. *Towards a Universal Grammar for Natural Language Processing*. In A. Gelbukh (Ed.), *Proceedings of CICLing 2015, Part I, LNCS 9041*, pp. 3–16, Springer International Publishing Switzerland.
- Sito Universal Dependencies (UD) project
  - <http://http://universalddependencies.org/>
- Bosco C., Montemagni S., Simi M. (2013). *Converting Italian Treebanks: Towards an Italian Stanford Dependency Treebank*. In *Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse (LAW VII & ID at ACL-2013)*, Sofia, Bulgaria, August 8-9, pp. 61-69.
- Dell'Orletta F., Venturi G. (2016) "ULISSE: una strategia di adattamento al dominio per l'annotazione sintattica automatica". In E. M. Ponti e M. Baudassi (a cura di) "Computer parler soigner: tra linguistica e intelligenza artificiale", *Atti del convegno 15-17 dicembre 2014*, Pavia University Press, pp. 55-79.

### Monitoraggio linguistico e sue applicazioni

- Montemagni S. (2013) "Tecnologie linguistico-computazionali e monitoraggio della lingua italiana". In *Studi Italiani di Linguistica Teorica e Applicata (SILTA) Anno XLII, Numero 1*, pp. 145-172,

Un articolo a scelta tra i seguenti:

- Dell'Orletta F., Montemagni S., Venturi G. (2013), "Linguistic Profiling of Texts Across Textual Genre and Readability Level. An Exploratory Study on Italian Fictional Prose". In *Proceedings of the Recent Advances in Natural Language Processing Conference (RANLP-2013)*, 7-11 September, Hissar, Bulgaria, pp. 189-197
- Dell'Orletta F., Montemagni S. e Venturi G. (2016) "Esplorazioni computazionali nello spazio dell'interlingua: verso una nuova metodologia di indagine". In R. Bombi e V. Orioles (a cura di), *Atti del XLVIII Congresso Internazionale di Studi della Società di Linguistica Italiana (SLI 2014)*, 25-27 settembre Udine.
- Barbagli A., Lucisano P., Dell'Orletta F., Montemagni S., Venturi G. (2015) "Il ruolo delle tecnologie del linguaggio nel monitoraggio dell'evoluzione delle abilità di scrittura: primi risultati". In *Italian Journal of Computational Linguistics (IJCoL)*, vol. 1, n. 1, pp. 99-117.

### Analisi della leggibilità del testo

- *Recent Advances in Automatic Readability Assessment and Text Simplification*. Special issue of the *International Journal of Applied Linguistics*, 2014, 165:2, John Benjamins Publishing Company
  - "Computational Assessment of Text Readability: A Survey of Current and Future Research", Kevyn Collins-Thompson
  - "Assessing document and sentence readability in less resourced languages and across textual genres", Dell'Orletta F., Montemagni S., Venturi G. (2014)
  - Siddharthan A., (2014) *A survey of research on text simplification*, *ITL - International Journal of Applied Linguistics*, Volume 165, Issue 2, 2014, pp. 259-298

### Indicazioni per non frequentanti

Contattare le docenti per concordare il programma d'esame.

### Modalità d'esame



## UNIVERSITÀ DI PISA

---

L'esame consisterà nell'illustrazione, sotto forma di relazione scritta, dei risultati di un progetto di annotazione linguistica di testi rappresentativi di una specifica varietà d'uso della lingua italiana e in un colloquio volto a discutere i risultati del progetto e a verificare l'acquisizione dei concetti introdotti nel corso e nei materiali didattici indicati.

### Note

Il corso di Linguistica Computazionale II inizierà il **26 febbraio 2020**.

*Ultimo aggiornamento 14/02/2020 17:08*