# Computational linguistics 2

## Prof. Marco Carlo Passarotti

***COURSE AIMS AND INTENDED LEARNING OUTCOMES***

The course aims to introduce advanced concepts and methodologies of the discipline, specifically focusing on semantic-based linguistic resources, and tools and methods for the automatic processing of semantics, for tasks like semantic role labeling, named entity recognition, word sense disambiguation and topic analysis.

At the end of the course, students will be provided with (a) expertise about the representation of semantic metadata in linguistic resources and (b) skills to process automatically at the semantic level sets of linguistic data with Natural Language Processing tools.

***COURSE CONTENT***

The course is divided into two sections:

1. presentation of layers of semantic annotation in linguistic resources, particularly focusing on:

a. semantic annotation standards (tagsets, formats, guidelines);

b. finding, using and querying resources in linguistic infrastructures (e.g., CLARIN).

2. Methods and tools for automatic semantic processing. A number of linguistic resources, like word/sense embeddings for various languages, will be introduced.

***READING LIST***

[All the readings reported below are not mandatory for the exam]

E.M. Bender, A. Lascarides, Linguistic fundamentals for natural language processing ii: 100 essentials from semantics and pragmatics Synthesis Lectures on Human Language Technologies, 2019, 12(3), pp. 1-268.

P. Cimiano, C. Chiarcos, J.P. McCrae, J. Gracia, Linguistic Linked Data. Representation, Generation and Applications, Springer, 2020.

J. Devlin, et al., Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018). <https://arxiv.org/pdf/1810.04805.pdf&usg=ALkJrhhzxlCL6yTht2BRmH9atgvKFxHsxQ>

M.T. Pilehvar, J. Camacho-Collados, Embeddings in natural language processing: theory and advances in vector representations of meaning, Synthesis Lectures on Human Language Technologies 13.4, 2020, pp. 1-175.

***TEACHING METHOD***

Lectures (in English) with exercises.

***ASSESSMENT METHOD AND CRITERIA***

Oral exam to verify the degree of acquisition of course contents. The exam includes also one or more practical exercises about the use of linguistic resources and/or Natural Language Processing tools.

In particular, questions are about two main kinds of contents: (a) theoretical issues, like for instance the different theoretical approaches to bulding and using reosurces like word and sense embeddings for Natural Language Processing purposes; (b) practical issues, like for instance running a pipeline of Natural Language Processing tools.

In terms of assesment criteria, there is no difference between the questions about the theoretical questions of the course and those on the practical aspects.

***NOTES AND PREREQUISITES***

Given the advanced nature of the course, students are expected to have attended the course of Computational Linguistics 1.

*Place and time of consultation hours*

On appointment, by sending an e-mail to marco.passarotti@unicatt.it.

Place: CIRCSE Research Center, Franciscanum building, second floor, room n. 210.