# Theory and methods of linguistic annotation

## Prof. Eleonora Litta Modignani Picozzi

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

*Course Aims*

This course aims at providing the student with in-depth knowledge of linguistic annotation as the practice of adding descriptive notations to raw language data, and of all the necessary steps needed to model and create an annotation language, build a linguistic resource and evaluate its correctness.

Linguistic annotation is a crucial activity in the field of (computational) linguistics because of its role in both the development of language models for natural language processing applications and the empirical evaluation of theoretical statements.

*Learning Outcomes*

At the end of this course, the student will have a thorough overview on linguistic annotation types and standards, evaluation practices and annotation tools. During the course the students will be given elements of statistics and data visualisation, and will know how to add annotations to a corpus, build their own annotation model, draw linguistic theoretical information from it and evaluate the results. Students will be able to demonstate operative knowledge of what they have learned through thematic activities held during the course.

***COURSE CONTENT***

Part one of the course will describe all stages of the linguistic annotation process:

Annotation methods and models

Annotation standards

Building an annotation model

Annotation creation and tools

Evaluation

Iterative improvement of annotation accuracy

Data representation and exploitation

The second part of the course will include operative examples from significant annotation projects across a variety of linguistic annotation typologies, among which are morpho-syntactic tagging, semantic analyses (semantic roles, named entities, sentiment and opinion), discourse level analyses including discourse structure, co-reference, etc. The course will involve students with thematic activities on the theories and techniques they have learned in class.

***READING LIST[[1]](#footnote-1)***

Suggested Reading List

Nancy Ide – James Pustejovsky (editors), *Handbook of Linguistic Annotation*, Springer

(pp. 21-192, 297-314, 379-440)

Emily M. Bender, *Linguistic fundamentals for Natural Language Processing*, Springer

Bernd Heine (ed.), Heiko Narrog (ed.), *Oxford handbook of linguistic analysis*, OUP

(Chapter 8 and chapter 10)

***TEACHING METHOD***

Lectures (in English) with exercises.

***ASSESSMENT METHOD AND CRITERIA***

Oral exam with exercises to verify the degree of acquisition of the course contents. The exam includes also one or more practical exercises about what has been learned during the course.

In particular, questions are about two main kinds of contents: (a) theoretical aspects of linguistic annotation methods, like for instance annotation standards and methods for annotation agreement; (b) practical issues, as they have been experienced during the course.

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***

As this is a second year course, a good knowledge of the Linguistic Computing 1 and 2 syllabus is required. Basic knowledge of lemmatisation and post tagging tools is recommended as it is a revision of how to access local folders and launch NLP tools using a command line interface.

*Place and time of consultation hours*

On appointment, after each lesson, by sending an e-mail to eleonoramaria.litta@unicatt.it.

Place: CIRCSE Research Center, Franciscanum building, second floor, room n. 215.

1. I testi indicati nella bibliografia sono acquistabili presso le librerie di Ateneo; è possibile acquistarli anche presso altri rivenditori. [↑](#footnote-ref-1)