

## Master in Applied Linguistics (LM-39)

<b>Course title</b>	Computational Linguistics
<b>Course code</b>	54113
<b>Scientific sector</b>	L-LIN/01
<b>Degree</b>	Master in Applied Linguistics
<b>Semester</b>	1st
<b>Course year</b>	2nd
<b>Credits</b>	6 + 2
<b>Modular</b>	Yes

<b>Total lecturing hours</b>	30
<b>Total lab hours</b>	30
<b>Attendance</b>	according to the regulation
<b>Prerequisites</b>	

<b>Specific educational objectives</b>	<p>The objective of the course is to provide the students with a practical understanding of a) some aspects of speech and text data computational processing b) the computational tools and techniques available to achieve so.</p> <p>The computational linguistics module will focus on how to build a pipeline, how to process significant quantities of data, how to train language models and how to test them.</p> <p>The speech technology module will focus in the first part on the implementation of typical speech processing techniques, in particular for feature extraction and speech segmentation. In the second part, the course will address tasks as speaker identification and keyword spotting as a whole.</p>
--	---

<b>Module 1</b>	Computational linguistics
<b>Lecturer</b>	Luca Ducceschi
<b>Scientific sector</b>	L-LIN/01
<b>Teaching language</b>	English
<b>Office hours</b>	from Monday to Friday on request
<b>List of topics covered</b>	
<b>Teaching format</b>	<i>Laboratorial activities and personal projects</i>
<b>Total lecturing hours</b>	30
<b>Credits</b>	6

<b>Module 2</b>	Speech Technologies - LAB
<b>Lecturer</b>	Alessio Brutti
<b>Scientific sector</b>	L-LIN/01
<b>Teaching language</b>	English
<b>Office hours</b>	from Monday to Friday on request
<b>List of topics covered</b>	
<b>Teaching format</b>	<i>Frontal lectures, laboratorial activities and personal projects</i>
<b>Total lecturing hours</b>	30
<b>Credits</b>	2

<b>Learning outcomes</b>	Computational linguistics
--------------------------	---------------------------

	<p>The course's goal is to teach the students how to understand and address non-trivial tasks relative to natural language processing, such as a) accessing and managing gigabytes of textual data (e.g. the whole English Wikipedia), b) processing strings of characters using the basic Python modules and regular expressions, c) train some language models to solve problems like part-of-speech tagging, or text classification, evaluate such models using statistical tools and gold standard datasets. In order to stimulate their independent thinking and problem-solving skills students will be required to experiment with different solutions to a given problem and to evaluate the different options (e.g. rule based approaches vs trained models). Given the nature of the exam (an oral presentation of a computational project), some of the lab sessions will be dedicated to presenting and discussing the project ideas.</p> <p>The goal of the speech technology lab is to introduce the students to digital processing of non-stationary signals and to some common speech processing tasks as: speaker identification, keyword spotting and language recognition. Students will learn how: 1) to process speech signals in python, extracting and evaluating the most relevant features (e.g. pitch, formants, spectrograms, etc.) 2) how to tackle from scratch some typical speech processing tasks (speaker identification and keyword detection) using neural models in Keras. The course will also provide an overview of state-of-the-art end-to-end speech recognition systems.</p>
<p><b>Assessment</b></p>	<p><i>Oral Exam: students are required to discuss the details of a project that encompasses and develops the subjects covered during the course. The project will be defined and agreed upon before the exam. The oral exam will include 1 or 2 questions on the topics addressed during the course and not covered by the project.</i></p> <p><i>Each module will have a separate project/exam.</i></p>
<p><b>Assessment language</b> <b>Evaluation criteria and criteria for awarding marks</b></p>	<p><i>English</i></p> <p><i>The outcome of the exam will be determined by: a) the level of understanding of the topics covered during the course, b) the computational skills of the candidate, c) the clarity of the exposition and of the project.</i></p> <p><i>The final mark will be the average of the marks obtained in each module.</i></p>
<p><b>Required readings</b></p>	<p>Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit, Steven Bird, Ewan Klein, and Edward Loper, available here: <a href="https://www.nltk.org/book/">https://www.nltk.org/book/</a></p>
<p><b>Supplementary readings</b></p>	<p>Speech Communications: human and machine 2<sup>nd</sup> edition, O'Shaughnessy Introduction to Digital Speech Processing by Rabiner and Schafer, 2007</p>