**Digital Humanities (Computational Linguistics)**

Prof. Guido Milanese; Prof. Alessandra Lombardi

**Module 1: Digital Humanities (Prof. Guido Milanese)**

***COURSE AIMS***

The course aims to provide students with the basic skills (1) for analysing texts of various types with the aid of computers and (2) for structured scientific writing.

***INTENDED LEARNING OUTCOMES***

At the end of the course, students will be able to:

1. Understand the relationship between IT, linguistics and philology;
2. Be able to identify the right tool for various typologies of linguistic and literary analysis.

***COURSE CONTENT***

1. Introduction to the history of philology and computational linguistics.
2. Hardware and software: definitions of “humanistic IT”.
3. Free software and proprietary software
4. Digitalisation and OCR.
5. “Markup” and digital texts (XML, TEI, Markdown).
6. Textual and bibliographical data banks.
7. Analysing texts by computer (levels of analysis).
8. Structured writing with LaTeX (LyX) and Markdown, BibTeX.

IT tools to be demonstrated:

* OCR programmes, indexing and concordances

lemmatisation programmes and morphological analysis programmes

* LaTeX and related programmes (LyX, BibTeX and Jabref in particular).
* Markdown
* Tools for philological analysis

***READING LIST***

In addition to the course notes: G. Milanese, *Filologia, letteratura, computer*, Milan, Vita e Pensiero 2020

***TEACHING METHOD***

Lectures will be computer based. Students with a portable computer (notebook, netbook, tablet etc.) are welcome to use it during lectures.

***ASSESSMENT METHOD AND CRITERIA***

Students will be required to prove their ability to use *at least one* of the tools demonstrated in lectures, and answer questions on the theoretical and historical elements of the course.

***NOTES AND PREREQUISITES***

The course is introductory in nature and as such does not require previous skills. Basic computer skills are sufficient.

Further information can be found on the lecturer’s webpage at http://docenti.unicatt.it/web/searchByName.do?language=ENG or on the Faculty notice board.

**Module 2: Computational lexicography**

**Prof. Alessandra Lombardi**

***COURSE AIMS***

The aim of the course is to introduce the key concepts of lexicography, tracing their development (from paper format to digital) and describing their main areas of application. Through the logical presentation of the various types of online lexical resources currently available, as well as related challenges and opportunities for development, the course aims to provide students with the practical and intellectual tools they need to use these resources knowledgeably and critically. The course will focus particularly on systems which enable the automatic acquisition, the processing and systematic representation of reoccurring lexical combinations (collocations) extracted from corpora.

***INTENDED LEARNING OUTCOMES***

At the end of the course, students will be able to:

1. Describe, in a relevant and terminologically accurate manner, the macro and micro-structural aspects of traditional lexicographical resources, as well as the main functions of the most commonly used digital resources for language learning and translating (con riferimento all’italiano e alle lingue straniere studiate);
2. Explain, at least provide a general outline of, the development of lexicographical theory and practice over time with reference to different linguistic-cultural traditions;
3. Produce critical considerations on research possibilities, consult and use specific tools offered by different online lexical resources;
4. Adopt targeted, effective strategies for searching for information online and building up a suitable vocabulary for performing various interpreting tasks.

***COURSE CONTENT***

1. Lexicography: theoretical and applicative foundations.

2. Evolution of dictionaries: from the origins to the digital age.

3. Tools for computational lexicography: comparison of typologies

a. Lexical-ontological resources (basic knowledge): online thesauruses, glossaries and ontologies;

b. machine dictionaries (e.g. applications: frequency vocabulary; assisted and automatic translation systems);

c. electronic lexicography products:

computerised dictionaries

electronic dictionaries

collaborative dictionaries

aggregators of lexicographical sources;

4. The impact of corpus on lexicography practices: *corpus-based dictionaries; corpus query tools* (acquiring lexical information via the analysis of corpus (contexts, concordances, *word sketches*).

***READING LIST***

a. Reference handbook: Granger S. &Paquot M.(2013), *Electronic Lexicography,* Oxford University Press

b. For the part on the history of lexicography

Hanks, P. 2012, *Lexicography from Earliest Times to the Present* (lightly edited version of a chapter first published in Keith Allan (ed., 2013), The Oxford Handbook of the History of Linguistics. Oxford: Oxford University Press, p. 503-536.). Also available online.

<http://www.patrickhanks.com/uploads/5/1/4/9/5149363/2012d-lexicography_from_earliest_times.pdf>

c. 3 more recently published single-subject essays, for the in-depth study of topics discussed during the course, the titles of which will be indicated at the end of the course and posted on Blackboard.

d. Additional material (recommended for all, but mandatory for *non-attending students*): *video-relazione* di Lars Trap-Jensen al XVIII EURALEX International Congress *Lexicography in global contexts* (Ljubljana 2018): “Lexicography between NLP and linguistics: aspects of theory and practice”:

http://videolectures.net/euralex2018\_trap\_jensen\_lexicography/

***TEACHING METHOD***

Frontal lectures and practical exercises in the computer laboratory. All the teaching material used during lectures will be available on Blackboard.

***ASSESSMENT METHOD AND CRITERIA***

The final examination will take into account three different areas of assessment:

1. mastery of basic notions (students should demonstrate their knowledge of the key concepts of lexicography, historical development, theoretical foundations and main computational applications, as well as the problems and perspectives related to the production and fruition of online lexicographical resources);

2. ability to recognise and describe, in a critical and argumentative manner, the aims, structure and function of a digital lexicographical resource presented by the lecturer (and chosen from a selection illustrated during the course);

3. presentation during the exam, with related reasoned comments, of an individual mini-project chosen from among those suggested during the last lecture of the module (to be developed according to the methodological criteria indicated by the lecturer and available on Blackboard after the end of lectures, and to be sent by email 15 days before the exam interview has been scheduled).

Students who do not undertake the individual end-of-course project are required to prepare the supplementary essay (see Reading List, point d.) in addition to the three mandatory single-subject essays for all students.

Because the course consists of two modules, the mark for the part on Computational Lexicography (Prof. Lombardi) will count for 50% of the final mark, as the part of the final mark awarded by Prof. Milanese.

***NOTES AND PREREQUISITES***

There are no prerequisites for attending the course. However, students should be interested in broadening their lexical knowledge and skills within digital literacy.

Further information can be found on the lecturer's webpage at http://docenti.unicatt.it/web/searchByName.do?language=ENG or on the Faculty notice board.