Universidad de Alcalá

GUÍA DOCENTE

LINGÜÍSTICA COMPUTACIONAL / *COMPUTATIONAL LINGUISTICS*

**Grado en Estudios Ingleses
Grado en Lenguas Modernas y Traducción
Universidad de Alcalá**

Curso Académico 2022-2023
Cursos 3º- 4º– Cuatrimestre 1º

GUÍA DOCENTE

Nombre de la asignatura:	Lingüística computacional / Computational linguistics
Código:	251068
Titulación en la que se imparte:	Grado en Estudios Ingleses Grado en Lenguas Modernas y Traducción
Departamento y Área de conocimiento:	Departamento de Filología Moderna Área de Filología Inglesa
Carácter:	Optativo
Créditos ECTS:	8
Curso y cuatrimestre:	3º / 4º curso – 1er cuatrimestre
Profesorado:	Antonio Pareja Lora
Horario de tutoría:	Se indicará al comienzo del cuatrimestre
Idioma en el que se imparte:	Inglés

COURSE SUMMARY

Lingüística Computacional (Computational linguistics) is an optional course offered for students in the 3rd or 4th year of the Degree in English Studies and the Degree in Modern Languages and Translation. It aims at developing the appropriate knowledge and skills in the area(s) of human language technologies, language engineering and/or natural language processing (NLP). Students will be provided with relevant theoretical and practical NLP resources used in the field. They will be familiar with the key concepts in computational linguistics and will be able to use different NLP tools for different purposes, as well as to create their own very basic resources in some restricted fields of application of NLP.

1. COURSE DESCRIPTION

The course is structured in such a way as to present the most important theoretical concepts and practical issues in computational linguistics. Taking into account a practical approach, this module is designed to make students aware of the wide range of NLP tools and resources they can use. Students are expected to acquire a critical value on the use of the different tools and resources that will be presented. Students are expected to attend lectures and hands-on technology sessions.

A like for linguistic topics together with a sound background in linguistic disciplines are required. In addition, capability and willingness to use computers and to handle quantitative and qualitative data plus a good command of certain basic computer skills (Windows and/or Linux operating systems, advanced word processing and spreadsheet/database management) are also required.

The course is quite practical, as learning is based on practical tasks and activities that, in some cases, are quite difficult to perform outside the classroom. Therefore, students are explicitly cautioned not to enrol in the course if they are subject to time or other constraints that may hinder their attendance.

Since all activities are conducted in English, the module assumes a good working knowledge of English (Common European Framework of Reference for Languages B2.2 level or higher). Following the guidelines issued by the Department of Modern Philology, students are expected to have acquired a level C1 in all four skills at the end of their 4th year. Students in their 3rd year, for their part, will be expected to be well on their way to achieving this advanced level of English in all four communicative skills.

2. COMPETENCES

Generic Competences:

1. To improve student´s abilities to communicate and work in groups.
2. To enhance language competence in the communication of information, ideas, opinions, problems and solutions.
3. To improve language skills (reading, writing, speaking and listening).
4. To develop analytical and argumentative skills and independent critical thinking based on supporting evidence.
5. To develop student´s autonomy as learners.
6. To develop the capacities for self-learning, self-monitoring, self-improvement, self-assessment and team work.
7. To make efficient use of library and research skills in order to find and organize information.
8. To develop the student's abilities to analyze and synthesize information.
9. To develop the student's abilities for independent and co-operative learning.
10. To get acquainted with the use of advanced computer tools and on-line resources.
11. To develop the capacity to observe and link phenomena establishing cause-effect relationships.
12. To develop an understanding of science and a scientific knowledge of English.

Specific Competences:

1. To understand the key concepts related to computational linguistics.
2. To be able to use different natural language processing tools and resources.
3. To critically evaluate natural language processing tools and resources.

4. To break down a real problem into its component parts and split these into a series of feasible steps.
5. To identify and implement the best solution for a specific problem.
6. To arrange and handle data in appropriate ways.
7. To find computational solutions to common research tasks and problems.
8. To locate suitable computational tools and resources and to make use of them.
9. To implement simple computational tools for specific tasks when necessary.
10. To recognize the scope and possible causes of experimental error.
11. To write well-structured and coherent reports of their studies.

3. COURSE CONTENT

Syllabus	Workload
UNIT 0: INTRODUCTION • Computational Linguistics, natural language processing, language engineering and language technologies	• 3 hours
UNIT 1: FUNDAMENTALS OF PROGRAMMING FOR THE LANGUAGE SCIENCES • Basic datatypes • Basic functions • Basic data structures • Basic programming structures	• 12 hours

UNIT 2: LANGUAGE TECHNOLOGIES (I)
• Computational morphology and morphosyntax
• Computational syntax
   o Dependency parsing
   o (H)PSG parsing
• Computational semantics
• Computational processing of discourse
• Computational pragmatics
	• 12 hours

UNIT 3: LANGUAGE TECHNOLOGIES (II) • Speech technologies • Spell & grammar checkers • Information retrieval • Information extraction and text mining	• 9 hours

UNIT 4: DIGITAL LANGUAGE RESOURCES (I)
• Corpora:
   o Classification
   o Development
   o Annotation
   o Related standards
	• 3 hours

UNIT 5: DIGITAL LANGUAGE RESOURCES (II)
• Digital lexical resources:
    ○ Classification
    ○ Development
    ○ Main examples
    ○ Related standards
	• 3 hours

UNIT 6: DIGITAL LANGUAGE RESOURCES (III)
• Digital terminological resources:
    ○ Classification
    ○ Development
    ○ Main examples
    ○ Related standards
	• 3 hours

4. METHODOLOGY AND CLASS ACTIVITIES

4.1. STUDENT WORKLOAD

Face-to-face and/or classroom sessions	50 (45 one-hour whole group sessions, 2 workshops, 1 small-group tutorial, and a two-hour exam).
Students' autonomous work	150 hours
Total hours	200

4.2. LEARNING ACTIVITIES, MATERIALS AND RESOURCES

Whole group sessions	Lectures and seminars: Introduction to computational linguistics and to natural language processing tools and resources. Individual and group work. Group tutorial and workshops.
Workshops	Activities focused on different computational linguistics issues.
Autonomous work	Readings; projects, research assignments and other tasks involving different areas, tools and resources in the fields of computational linguistics, possibly some basic computational programming and/or development; preparation of the exam.

5. ASSESSMENT

Assessment procedure

All students will be graded following a process of continuous assessment. Different aspects of the course will be evaluated separately throughout the semester in order to obtain an overall continuous and formative final mark for the course. This means that all the students will have to participate actively and effectively in classes as well as in all the group activities. A minimum of 80% attendance is compulsory.

Exceptionally, those students who have been recognized as eligible for final evaluation in accordance with the terms in article 10 of the "NORMATIVA REGULADORA DE LOS PROCESOS DE EVALUACIÓN DE LOS APRENDIZAJES" (passed on 24-03-2011) may sit for a final exam, which will include questions and activities on all the topics covered in the class and translation activities. Those students who have not passed, either through continuous or final evaluation, will have the right to be assessed again in the corresponding month through an exam that is similar to the first one.

Assessment criteria

Students should demonstrate they have acquired the main concepts and competences related to this course. Thus, by the end of the course, students are expected to be able to:

- Demonstrate knowledge and understanding of the key issues and concepts of each unit.
- Use different natural language processing tools and resources.
- Use NLP techniques, tools and resources translation technology to solve practical cases.

Grading criteria

Taking all this into consideration, students will be graded as follows:

MATRÍCULA DE HONOR (HONOURS):

- Student fully understands computational linguistics concepts, and is able to put all of them into practice by means of the appropriate tools and resources, showing advanced knowledge of the field and is ready to apply it to completely new scenarios.
- Student has no errors of decoding / encoding and he / she has a very good use of English.
- Student shows the ability to critically evaluate NLP technologies in a clear way, demonstrating independent ideas and opinions.

SOBRESALIENTE (FIRST):

- Student has an excellent understanding of computational linguistics concepts and is able to put all of them into practice by means of the appropriate tools and resources.
- Student does not have significant errors of decoding / encoding and he / she has a very good use of English.
- Student shows the ability to critically evaluate NLP technologies in a clear way, demonstrating independent ideas and opinions.

NOTABLE (SECOND)

- Student well understands computational linguistics concepts and is able to put most of them into practice by means of the appropriate tools.
- Student has up to one important decoding / encoding error and has a good use of English.
- Student shows the ability to critically evaluate NLP technologies, demonstrating some independent ideas and opinions.

APROBADO (PASS)
- Student has some important errors in understanding and putting into practice computational linguistics main concepts but minimum is achieved.
- Student has encoding problems (choice of vocabulary, idioms and register) and / or significant grammatical errors.
- Student evaluates NLP technologies, but in a very general way and has few critical opinions.

SUSPENSO (FAIL)
- Student has severe problems and shows inaccuracies in understanding and putting into practice computational linguistics main concepts.
- Student shows persistent serious grammatical inaccuracy and inappropriate choices of vocabulary, idiom and register.
- Student is not able to critically evaluate NLP technologies and does not show any independent thinking.

Assessment breakdown

Evaluation will be based on practice, presentations and a final exam weighted as follows:

- Projects and practices 30%
- Written assignments and oral presentations 40%
- Final exam 30%
  TOTAL 100%

6. REFERENCES

Bolshakov, Igor A. and Gelbukh, Alexander: *COMPUTATIONAL LINGUISTICS – Models, Resources, Applications*. Mexico DF, Universidad Nacional Autónoma de México. 2004.

Indurkhya, Nitin and Fred J. Damerau (editors): *Handbook of Natural Language Processing* (second edition). Boca Raton, FL: CRC Press. 2010.

Jurafsky, D. and Martin, J: *Speech and language processing*. Prentice Hall. Upper Saddle River. 2009.

Manning, Christopher D. and Hinrich Schütze: *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, MA. 1999.

Mitkov, Ruslan (ed.):*The Oxford Handbook of Computational Linguistics*. Oxford University Press, Oxford. 2003.

Pareja-Lora, Antonio: *Providing Linked Linguistic and Semantic Web Annotations: The OntoTag Hybrid Annotation Model*. Saarbrücken, LAP (Lambert Academis Publishing). 2012.

Pareja-Lora, Antonio; Blume, María; Lust, Barbara C. and Chiarcos, Christian (editors): *Development of Linguistic Linked Open Data Resources for Collaborative Data-Intensive Research in the Language Sciences*. Cambridge, MA, The MIT Press. 2020.

Sayers, D., R. Sousa-Silva, S. Höhn et al.: *The Dawn of the Human-Machine Era: A forecast of new and emerging language technologies*. Report for EU COST Action CA19102 'Language In The Human-Machine Era'. https://lithme.eu/. 2021.

7. ADDITIONAL PROVISION

The University of Alcalá guarantees that if, due to health requirements, the public authorities prevent teaching activity from taking place on the University's premises, the teaching plans' objectives will be met through an online teaching and evaluation methodology. The UAH commits to return to face-to-face teaching as soon as said impediments cease.