



Universidad
de Alcalá

GUÍA DOCENTE

Herramientas informáticas en la investigación lingüística / Computer tools in language research

**Grado en Estudios Ingleses
Grado en Lenguas Modernas y
Traducción**

Universidad de Alcalá

Curso Académico 2018/2019

Curso 3º- 4º– Cuatrimestre 1º

GUÍA DOCENTE

| | |
|---|---|
| Nombre de la asignatura: | Herramientas informáticas en la investigación lingüística / Computer tools in language research |
| Código: | 251043 |
| 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: | 3º / 4º curso - 1º cuatrimestre |
| Profesorado: | José Simón Granda |
| Horario de Tutoría: | A determinar en función del horario de la asignatura |
| Idioma en el que se imparte: | Inglés |

1. MODULE DESCRIPTION

Much linguistic research relies on the study of data derived from actual observation or experimentation. Nowadays, this kind of research (empirical research) entails extensive use of computer tools and familiarity with quantitative and statistical methods of analysis.

This course is designed to meet the needs of students seeking to get acquainted with the scientific foundations and standard techniques and protocols of empirical research as applied in different linguistic domains: Morphology-Syntax, Semantics, Pragmatics, Discourse Analysis, Phonology-Phonetics, Lexicography-Terminology, Stylometry, Forensic Linguistics or Applied Linguistics. The course will introduce students in the core areas of research planning and organisation, corpus linguistics, quantitative/qualitative analysis, data mining and data-driven language engineering. Common tasks in the different stages of linguistic enquiry will be discussed and practised making use of suitable computational tools.

Prerequisites

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.

Since all activities are conducted in English, the module assumes a good working knowledge of English (Common European Framework level B2.2 or above).

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. AIMS

Upon successful completion of the course, students will have acquired the following skills and competences:

Generic competences:

1. To develop analytical and argumentative skills and independent critical thinking based on supporting evidence.
2. To improve language skills (reading, writing, speaking and listening).
3. To enhance language competence in the communication of information, ideas, opinions, problems and solutions.
4. To develop the student's abilities to analyse and synthesize information.
5. To develop the student's abilities for independent and co-operative learning.
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 organise information.
8. To get acquainted with the use of advanced computer tools and on-line resources.
9. To develop the capacity to observe and link phenomena establishing cause-effect relationships.
10. To develop an understanding of science and a scientific knowledge of English.

Subject specific competences:

1. To plan and design different types of linguistic survey setting clear objectives and appropriate working hypotheses.
2. To break down a real problem into its component parts and split these into a series of feasible steps.
3. To identify and implement the best solution for a specific problem.
4. To identify and make use of relevant material in general-purpose and ad-hoc corpora.
5. To arrange and handle data in appropriate ways.

6. To understand the significance of careful, accurate and precise recording of data.
7. To analyse data qualitative and quantitatively in an objective and systematic way.
8. To make the appropriate choice of statistical tests for each particular problem making convenient use of them.
9. To find computational solutions to common research tasks and problems.
10. To locate suitable computational tools and to make use of them.
11. To implement simple computational tools for specific tasks when necessary.
12. To recognise the scope and possible causes of experimental error.
13. To draw conclusions out of data analysis and to extrapolate results.
14. To write well-structured and coherent reports of their studies.

3. MODULE CONTENTS

| Units | Topics | Time |
|-----------------------------|---|---------|
| • Foundations | General scientific methods Research methodologies and techniques Computing in the linguistic sciences | 1 week |
| • Survey design | First steps: identifying research questions, aims and working hypotheses Types of empirical research Study/survey/experiment design Project proposal Literature review | 2 weeks |
| • Working with textual data | Data and metadata Coding and mark-up conventions Corpus compilation, annotation and exploitation | 2 weeks |
| • Computational techniques | Data handling: coding, tagging, lemmatising and parsing data Tagsets and tagset mapping Generating lists, indices and concordances Computing frequencies, calculations and graphs Data analysis and data mining | 3 weeks |

| | | |
|--|---|---------|
| <ul style="list-style-type: none"> • Statistical methods | Introduction to the R and SPSS packages Summary statistics Inferential statistics Describing relationships: correlation; regression; sample comparison Testing hypotheses | 3 weeks |
| <ul style="list-style-type: none"> • Suggested prototypical projects (might change depending on students' skills and interests) | Stylometry: forensic applications Assessing text complexity and readability Lexicography and terminology: neologism watch Prototypical corpus annotation Automated assessment of categories and text annotation | 4 weeks |

4. TEACHING AND LEARNING METHODS

The course consists of one weekly whole-class session and two weekly seminars, as well as one ECTS tutorial in small groups (5-8 students) plus four one-hour workshops (in groups of 4-6 students). Tutorial and workshops dates will be evenly spaced over the term.

The full contents of the course will be introduced in a practical way since the first session; therefore, attendance and active participation in lectures and seminars are a must. Lectures and seminars will be run in English. Likewise, the students' activities and essays must be written in English.

Prior to each class students are expected to read the set key readings for the week. Weekly seminars will include a critical discussion of the issues raised followed by incremental hands-on practice, first conducted under tutor guidance and then carried out by the students on their own (always under supervision). Most of these activities will involve group work, self-assessment and peer revision.

These activities must be complemented with online tutorials for some of the units and for the computer tools to be used plus further practice the students should carry out either at home or at any of the university computer rooms. All the materials (software, text samples, tutorials and instructions) required for their practice will be uploaded in advance onto our virtual learning platform so that students may access them at any time. Students will deliver two oral presentations and two essays in addition to day-to-day laboratory exercises.

All the activities (both done at the laboratory and at home), proposals, drafts and written assignments should form part of a portfolio. Self-assessment, tutor's corrections/suggestions and peer revisions notes and comments should also be included in the portfolio. At the end of the course the portfolio will be turned in for assessment.

ETCS Tutorial: Presentation of the assignments for the four workshops: brief introduction to empirical research methods; discussion of the methodology for written work and oral presentations.

- Workshop 1: Discussion and peer revision of their first paper (research assignment).
 Workshop 2: Oral presentation of the first research assignment and their portfolio.
 Workshop 3: Discussion and peer revision of their second paper (research assignment).
 Workshop 4: Oral presentation of the second research assignment and their portfolio.

Hopefully, classes, seminars, tutorials and workshops will be held at the new computers room (Caracciolos) or Henry Sweet laboratory, which limits the maximum number of attending students to 24. Accordingly, prior to getting enrolled in the course pre-applicant students will be given an entry test in order to assess their initial level at the pre-required skills; each of them will be informed of his or her results to make them aware of the extra effort they will have to devote in case their levels do not meet the standard.

4.1. Student workload: 200

| | |
|---------------------------|--|
| Class contact hours: 52 | 45 class sessions 1 ECTS tutorial 4 hours workshops 2 final exam Total: 52 |
| Students study hours: 148 | Readings Extra activities Research assignments Total: 148 |
| Total hours | 200 |

4.2. Learning activities, materials and resources

| | |
|-----------------------|--|
| Face to face sessions | Lectures and seminars Group tutorials and workshops |
| Autonomous work | Readings Online tutorials Further practice activities Preparation of assignments Preparation of oral presentations Preparation of exams |
| Tutorials | Individual tutorials on demand |

4.3. Materials and resources

Key readings and basic (mostly public domain) computer tools will be supplied by the tutor (for a complete list of readings and other materials see “bibliography” below).
 Several on-line tutorials.

Blackboard learning platform.

Course intranet.

Language laboratory / computer room (depending on availability).

5. ASSESSMENT

The course will be graded by continuous assessment except for those students who are granted evaluation by final exam, according to the university regulations. Students who follow the continuous assessment process will be graded in accordance with the following criteria:

Assessment criteria

Assessment will rely on the degree of achievement of the target competences. It will take into account to what extent the student:

- Can plan, design and carry out different types of linguistic studies.
- Manages to find suitable computational tools and knows how to use them.
- Is able to identify and make use of relevant evidence in data sources.
- Makes the appropriate choice and use of statistical tests.
- Knows how to analyse data qualitative and quantitatively.
- Can justify findings and draw conclusions.
- Writes well-structured and coherent reports of their studies.
- Shows capacity to present and debate a project orally.

Marking criteria

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

| | |
|---|-----|
| – Laboratory and classroom practice | 30% |
| – Self-practice (registered in the student's portfolio) | 10% |
| – Written assignments and oral presentations | 40% |
| – Final exam | 20% |

| | |
|--------------|-------------|
| Total | 100% |
|--------------|-------------|

Marking procedures

- Laboratory practice will be continuously assessed, randomly checking, correcting and marking every quiz and activity the students do. They will be regularly informed of their progression and possible problems detected by the tutor, who will supply them with complementary activities when necessary.
- Self-practice will be assessed by means of their portfolio.
- Since students are expected to have at least a C1.1 level (upon entry), activities will be graded progressively and requirements will range from this initial level to the final C1.2 level.
- In all the activities, both oral and written, overall language performance at academic level will also be assessed.
- In oral presentations, fluency and accuracy in pronunciation will be equally taken into account together with the ability to discuss ideas convincingly in English.

- The final exam is compulsory for everyone and will consist of a questionnaire plus a set of practical activities to be done at the laboratory.

Students will then be assessed through a combination of:

- Level of achievement in class and home activities and assignments.
- Participation in class discussions and practice.
- Essays and presentations to be delivered along the term.
- An overall final exam.

According to the instructions contained in the 'normativa reguladora de los procesos de evaluación de los aprendizajes (aprobada en Consejo de Gobierno de 24 de marzo de 2011)', there will also be one final summative exam (100% of the grade) of the materials covered in the course for those students who have requested exemption from the continuous evaluation process and have been thus authorized.

Likewise, in the extraordinary examination session (June) students will be graded by means of a summative exam (100% of the grade) of the material covered in the course.

6. BIBLIOGRAPHY

Basic bibliography:

Booth, Wayne C., Colomb, Gregory G. & Joseph M. Williams (2008) *The Craft of Research*. 3rd ed. Chicago/London: the University of Chicago Press.

Evans, David (2011) *Introduction to Computing: Explorations in Language, Logic, and Machines*. Available from <http://www.computingbook.org/FrontMatter.pdf>.

Lindquist, Hans (2009) *Corpus Linguistics and the Description of English*. Edinburg: EUP.

Oakes, Michael (1998) *Statistics for Corpus Linguistics*. Edinburg: EUP.

Swales, John M. and Christine B. Feak (2012) *Academic Writing for Graduate Students Essential Skills and Tasks*. 3rd ed. Ann Arbor: University of Michigan Press.

Additional bibliography:

Huck, Schuyler W. (2012) *Reading Statistics and Research*. 6th ed. Boston: Pearson.

Mansfield, Richard (2010) *Mastering VBA for Office 2010*. Indianapolis: Wiley.

Rockoff, Larry (2011) *The Language of SQL: How to Access Data in Relational Databases*. Boston: Course Technology.

Valiela, Ivan (2001) *Doing Science: Design, Analysis, and Communication of Scientific Research*. Oxford: Oxford University Press.

Weisser, Martin (2009) *Essential Programming for Linguistics*. Edinburg: EUP.

Online resources:

Association for Computers in the Humanities: <http://www.ach.org/>

OTA (Oxford Text Archive): <http://ota.ox.ac.uk/>

Purdue Online Writing Lab: <http://owl.english.purdue.edu/owl/>

SageMath: <http://www.sagemath.org/>

Rweb (R statistics online): <http://pbil.univ-lyon1.fr/Rweb/>