COURSE OUTLINE

(1) GENERAL

SCHOOL	PHILOSOPHY
ACADEMIC UNIT	PHILOLOGY
LEVEL OF STUDIES	Undergraduate
COURSE CODE	**GLOF 322** **SEMESTER** 3-8
COURSE TITLE	Computational and Statistical Approaches in Dialectology

INDEPENDENT TEACHING ACTIVITIES *if credits are awarded for separate components of the course, e.g. lectures, laboratory exercises, etc. If the credits are awarded for the whole of the course, give the weekly teaching hours and the total credits*	**WEEKLY TEACHING HOURS**	**CREDITS**
Seminar/Workshop	3	10
Add rows if necessary. The organisation of teaching and the teaching methods used are described in detail at (d).		

COURSE TYPE *general background, special background, specialised general knowledge, skills development*	special background, specialised general knowledge, skills development
PREREQUISITE COURSES:	There are no prerequisite courses.
LANGUAGE OF INSTRUCTION and EXAMINATIONS:	Greek
IS THE COURSE OFFERED TO ERASMUS STUDENTS	Yes (in Greek)
COURSE WEBSITE (URL)	https://elearn.uoc.gr/

(2) LEARNING OUTCOMES

Learning outcomes

The course learning outcomes, specific knowledge, skills and competences of an appropriate level, which the students will acquire with the successful completion of the course are described.

Consult Appendix A

- *Description of the level of learning outcomes for each qualifications cycle, according to the Qualifications Framework of the European Higher Education Area*

- *Descriptors for Levels 6, 7 & 8 of the European Qualifications Framework for Lifelong Learning and Appendix B*

- *Guidelines for writing Learning Outcomes*

Upon successful completion of this seminar/workshop, students will acquire specialized knowledge, skills, and competencies in computational and statistical methodologies for dialectology, with a focus on dialect distance measurement and dialectometric tools. Specifically, students will be able to:

1. **Understand and apply** methodologies for collecting and analyzing dialectal data, including grammatical descriptions, dialect dictionaries, spoken and written corpora, word lists, and georeferenced linguistic data from linguistic atlases and databases.
2. **Process linguistic data** using advanced tools, including *optical character recognition* (OCR), speech-to-text conversion, normalization, and semi-automatic annotation through machine learning techniques.

3. **Utilize computational and statistical methods** to quantify dialect distances and conduct comparative analyses of dialectal variation.
4. **Engage with and apply** dialectometric tools and fundamental machine learning techniques to dialectological research.
5. **Implement comparative research methodologies,** integrating computational and statistical techniques to investigate dialect variation and change.
6. **Collect, analyze, and systematically present** data from Modern Greek dialects or independently compiled datasets, bridging theoretical insights with empirical research.
7. **Develop and communicate research findings effectively** through structured academic reports and oral presentations, demonstrating critical engagement with computational and statistical approaches in dialectology.

General Competences

Taking into consideration the general competences that the degree-holder must acquire (as these appear in the Diploma Supplement and appear below), at which of the following does the course aim?

Search for, analysis and synthesis of data and information, with the use of the necessary technology	*Project planning and management*
Adapting to new situations	*Respect for difference and multiculturalism*
Decision-making	*Respect for the natural environment*
Working independently	*Showing social, professional and ethical responsibility and sensitivity to gender issues*
Team work	*Criticism and self-criticism*
Working in an international environment	*Production of free, creative and inductive thinking*
Working in an interdisciplinary environment	*……*
Production of new research ideas	*Others…*
	…….

Search for, analysis and synthesis of data and information, with the use of the necessary technology

Working in an international/interdisciplinary environment

Decision-making

Production of free, creative and inductive thinking

Project planning and management

Respect for difference and multiculturalism

Working independently

Team work

Criticism and self-criticism

(3) SYLLABUS

The seminar/workshop provides an in-depth exploration of computational and statistical methodologies in dialectology, guiding students through the full analytical pipeline. It begins with data collection methodologies, covering spoken and written sources such as grammars, text corpora, word lists, and georeferenced linguistic data. Students will then engage with modern processing techniques, including *Optical Character Recognition* (OCR), speech-to-text conversion, normalization, and semi-automatic annotation using machine learning methods.

Building on these foundations, the course introduces computational and statistical techniques for measuring and analyzing dialect distances. Students will employ dialectometric tools and applications, ranging from fundamental statistical approaches to advanced machine learning techniques. Emphasis is placed on comparative methodologies, with case studies from Modern Greek dialects and datasets compiled by students.

The course integrates theoretical instruction with hands-on practice, ensuring students gain experience across all stages of data processing and analysis. By the end of the seminar, students will be able to effectively apply modern computational tools to compare and study dialects, bridging theoretical insights with empirical research.

Course Outline:

Week	Course Content	Activity
1	Introduction to dialectology and course framework	**Theory:** Overview of the field, methodologies, and tools.
		Discussion: Assignment and guidance for individual/group projects.
2	Collection of dialectal data: spoken and written sources	**Theory:** Data collection methodologies and tools.
		Discussion: Guidance on project planning.
3	Data processing tools: OCR, Speech-to-Text	**Workshop:** Hands-on application of OCR and Speech-to-Text tools for data digitization.
4	Data normalization and annotation	**Workshop:** Application of normalization and semi-automatic annotation in text corpora.
5	Computational methods for measuring dialect distances	**Theory:** Introduction to dialect distance calculation methods.
		Discussion: Review of project progress and troubleshooting.
6	Dialectometric tools and software	**Workshop:** Practical application of dialectometric tools on spoken, written, and georeferenced data.
7	Statistical methods for analyzing dialect distances	**Workshop:** Application of statistical techniques to dialectal data.
8	Machine learning applications in dialectometry	**Theory:** Introduction to machine learning techniques for dialect analysis.
		Discussion: Planning the application of methods in student projects.
9	Preparation for oral presentations and final written reports	**Workshop:** Individualized support for structuring presentations and academic writing.
10-13	Final oral presentations and report submission	**Presentation:** Student presentations, discussion, and evaluation of research findings.
		Discussion: Final refinements before written report submission.

This structured approach ensures a gradual and comprehensive development of both theoretical understanding and practical expertise. The course promotes a balance between independent research and collaborative work while fostering academic writing and presentation skills.

(4) TEACHING and LEARNING METHODS - EVALUATION

DELIVERY *Face-to-face, Distance learning, etc.*	**1. Face-to-Face – Synchronous Learning** • **Lectures:** Presentation of theoretical concepts, methodologies, and tools, followed by interactive discussions. • **Workshops:** Hands-on application of modern computational and statistical tools for dialect analysis. • **Collaborative Learning:** Group-based problem-solving activities to reinforce practical skills. • **Case Studies:** Analysis of linguistic data from Modern Greek dialects or datasets compiled by students. •

	2. Individual and Group Work
	• **Data Collection and Analysis:** Students apply course methodologies to independent or group research projects.
	• **Guidance and Supervision:** Support in structuring research, preparing oral presentations, and writing reports.
	3. Asynchronous Learning
	• **Material Uploads:** Course materials, including lecture slides, academic articles, sample datasets, and additional resources, available on the eLearn platform.
	• **Online Communication Tools:** Announcements, discussion forums, and messaging through eLearn for ongoing support.
	• **Personalized Feedback:** Individualized guidance via email or scheduled personal meetings.
USE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY *Use of ICT in teaching, laboratory education, communication with students*	• Class lectures, presentations, slides • Class notes, announcements & communication via eLearn • Communication via email
TEACHING METHODS *The manner and methods of teaching are described in detail.* *Lectures, seminars, laboratory practice, fieldwork, study and analysis of bibliography, tutorials, placements, clinical practice, art workshop, interactive teaching, educational visits, project, essay writing, artistic creativity, etc.* *The student's study hours for each learning activity are given as well as the hours of non-directed study according to the principles of the ECTS*	(see table below)

Activity	Semester workload
Lectures	39
Independent study and exam preparation	83
Final written exam	3
Course total	*125*

| **STUDENT PERFORMANCE EVALUATION** *Description of the evaluation procedure* *Language of evaluation, methods of evaluation, summative or conclusive, multiple choice questionnaires, short-answer questions, open-ended questions, problem solving, written work, essay/report, oral examination, public presentation, laboratory work, clinical examination of patient, art* | **1. Assessment Method** The evaluation consists of two components: **i. Oral Presentation of the Project in Class (40%)** **Criteria for Assessment:** • Scientific presentation of data collection, analysis, and results. • Organization and clarity of presentation. • Communication skills. • Use of visual aids (e.g., PowerPoint). |

| *interpretation, other*

Specifically-defined evaluation criteria are given, and if and where they are accessible to students. | **ii. Final Written Assignment (60%)**

Criteria for Assessment:
• **Data Collection:** Accuracy and completeness in the collection of dialectal material.
• **Data Processing:** Application of appropriate computational and statistical analysis methods.
• **Structure and Organization:** Clear structure including introduction, methodology, results, discussion, and conclusions; logical flow and coherence.
• **Scientific Documentation:** Use of academic references and sources to support findings.
• **Presentation of Results:** Accurate interpretation and visualization of findings using tables, graphs, and examples.
• **Critical Analysis:** Sufficient discussion of results and formulation of well-supported conclusions.
• **Language and Academic Style:** Clarity of writing, use of appropriate terminology, and consistency in academic tone.

2. Transparency and Student Information
The assessment method and criteria are presented:
• During the first lecture.
• Throughout the course to reinforce understanding of the evaluation process.
• As a downloadable file on the eLearn platform for continuous access. |

(5) ATTACHED BIBLIOGRAPHY

1. General Bibliography:

Auer, P., & Schmidt, J. E. (Eds.). (2009). *Language and Space: An International Handbook of Linguistic Variation. Volume 1. Theories and Methods.* Walter de Gruyter. https://doi.org/10.1515/9783110220278 (with a specific reference to Part VI).

Boberg, C., Nerbonne, J. A., & Watt, D. J. L. (Eds.). (2017). *The Handbook of Dialectology* (First edition). John Wiley & Sons, Inc. (with a specific reference to Section 2).

Lameli, A., Kehrein, R., & Rabanus, S. (Eds.). (2010). L*anguage and Space: An International Handbook of Linguistic Variation. Volume 2. Language Mapping. Part I. Part II: Maps*. Walter de Gruyter. https://doi.org/10.1515/9783110219166 (with a specific reference to Part III).

2. Tools/Methods:

Guibon, G., Courtin, M., Gerdes, K., & Guillaume, B. (2020). When Collaborative Treebank Curation Meets Graph Grammars. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 5291–5300). European Language Resources Association. https://aclanthology.org/2020.lrec-1.651.

Heeringa, W., Van Heuven, V., & Van de Velde, H. (2023). *LED-A: Levenshtein Edit Distance App* [computer program] [Software]. https://www.led-a.org/.

Leinonen, T., Çöltekin, Ç., & Nerbonne, J. (2016). Using Gabmap. Lingua, 178, 71–83. https://doi.org/10.1016/j.lingua.2015.02.004.

Pichel, J. R., Gamallo, P., Alegria, I., & Neves, M. (2021). *A Methodology to Measure the Diachronic Language Distance between Three Languages Based on Perplexity*. Journal of Quantitative Linguistics, 28(4), 306–336. https://doi.org/10.1080/09296174.2020.1732177.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). *Robust Speech Recognition via Large-Scale Weak Supervision* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2212.04356.

Smith, R. (2007). An Overview of the Tesseract OCR Engine. Ninth *International Conference on Document Analysis and Recognition* (ICDAR 2007), Vol. 2, 629–633. https://doi.org/10.1109/ICDAR.2007.4376991.