Course guide
2709952 - IQL - Introduction to Quantitative Linguistics

Last modified: 30/01/2026

Unit in charge: Barcelona School of Informatics
Teaching unit: 723 - CS - Department of Computer Science.

Degree: MASTER'S DEGREE IN INFORMATICS ENGINEERING (Syllabus 2012). (Optional subject).
MASTER'S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Optional subject).
MASTER'S DEGREE IN DATA SCIENCE (Syllabus 2021). (Optional subject).

Academic year: 2025 **ECTS Credits:** 6.0 **Languages:** English

LECTURER

Coordinating lecturer:

Others:

PRIOR SKILLS

Programming
Basic probability and statistics

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Transversal:
CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.
CTR9. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

TEACHING METHODOLOGY

The theory sessions will be done primarily by the professor using either the blackboard or projected slides.

The lab work will be done in front of the computer. Students are expected to be working on their assignment, and the professor will explain all that is necessary to follow the class in the beginning of the session. Each lab session will be accompanied by a thorough guide describing the work that needs to be done.

The research project will be carried out under the supervision of the professor.

All the material relevant for the course will be available from Racó or the course's website .

LEARNING OBJECTIVES OF THE SUBJECT

1.Know the foundations of science and the scientific method. Understand the difference between hypothesis and theory, between modeling and understanding, between describing and explaining, between manifestation and principle. Understand the value of prediction and the types of prediction.
2.Learn about the statistical laws of language and their origins.
3.Know and understand the principles of organization of languages and other communication systems
4.Know the mathematical foundations of quantitative linguistics. Know basic probability theory and information theory.
5.Know the statistical analysis methods of quantitative linguistics.
6.Learn how to write a scientific article. Know how to distinguish between a laboratory report and a research paper.

STUDY LOAD

Type	Hours	Percentage
Hours large group	24,0	16.00
Hours small group	24,0	16.00
Self study	102,0	68.00

Total learning time: 150 h

CONTENTS

Introduction to Quantitative Linguistics

Description:
What is quantitative linguistics? Overview of linguistic laws, key concepts and research problems in quantitative linguistics.

Law of abbreviation and problem of compression

Description:
The law of abbreviation in humans and other species. Methods of analysis of the law of abbreviation. Introduction to information theory. Predictions of optimal coding.

Information theory

Description:
Classic information theory and extensions for natural communication systems.

Theory of power laws

Description:
Relationships between power laws. Inference of power laws. Power-law analysis methods.

Models of Zipf's law for word frequencies

Description:
Debowski's bounds. Classic models. Zipfian optimization models of communication.

The statistical structure of symbolic sequences

Description:
Word returns. Correlations in symbolic sequences. Persistence and antipersistence. n-gram models. Generative models.

Dependency syntax

Description:
Introduction to dependency syntax. Formal constraints on syntactic dependency structures.

Word order theory

Description:
Word order principles. Predictions. Ordre de subjecte (S), complement directe (O) and Verb (V).

Theory construction

Description:
The scientific method. A general theory. Closing.

ACTIVITIES

Introduction

Specific objectives:
1, 2

Related competencies :
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

Full-or-part-time: 8h 18m
Theory classes: 3h
Self study: 5h 18m

Law of abbreviation and the problem of compression

Specific objectives:
4, 5

Related competencies :
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

Full-or-part-time: 15h
Theory classes: 3h 48m
Practical classes: 1h
Laboratory classes: 3h 30m
Self study: 6h 42m

Information theory

Specific objectives:
3, 4

Related competencies :
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

Full-or-part-time: 12h 06m
Theory classes: 4h
Practical classes: 1h
Self study: 7h 06m

Theory of power laws

Specific objectives:
2, 4

Related competencies :
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

Full-or-part-time: 15h 36m
Theory classes: 4h
Practical classes: 1h
Laboratory classes: 3h 30m
Self study: 7h 06m

Models of Zipf's law for word frequencies

Specific objectives:
1, 2, 3, 4

Related competencies :
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

Full-or-part-time: 11h 54m
Theory classes: 4h
Practical classes: 0h 48m
Self study: 7h 06m

The statistical structure of symbolic sequences

Specific objectives:
2, 3, 4, 5

Related competencies :
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

Full-or-part-time: 15h 06m
Theory classes: 4h
Practical classes: 1h
Laboratory classes: 3h
Self study: 7h 06m

Dependency syntax

Specific objectives:
2, 5

Related competencies :
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

Full-or-part-time: 12h 06m
Theory classes: 4h
Practical classes: 1h
Self study: 7h 06m

Word order theory

Specific objectives:
3, 4, 5

Related competencies :
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

Full-or-part-time: 15h 36m
Theory classes: 4h
Practical classes: 1h
Laboratory classes: 3h 30m
Self study: 7h 06m

Theory construction

Specific objectives:
1, 3

Related competencies :
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

Full-or-part-time: 8h 18m
Theory classes: 3h
Self study: 5h 18m

Research project

Specific objectives:
1, 2, 3, 4, 5, 6

Related competencies :
CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

Full-or-part-time: 36h
Self study: 36h

GRADING SYSTEM

Grading is done by means of exams, reports on various tasks (labs and a research project) throughout the course.

There will two partial exams which count toward 30% of the score. Students are expected to hand in 4 lab work reports about two weeks after its corresponding lab session, which count toward 30% of the final grade. Finally, students will have to deliver a research project by the end of the course that accounts for 40% of the final grade. The research project is the most important activity and must be understood as a course project (not as one more lab). Labs must be understood as a training for the research project.

The formula to compute the final grade is therefore

0.4 * (P1 + P2) + 0.2 * (L1 + L2 + L3 + L4) + 0.4 * RP

where P1 is the score of the first partial exam, P2 is the score of the 2nd partial exam, Li stands for the grade for i-th lab and RP is the grade of the research project.

BIBLIOGRAPHY

Basic:
- "Linguistic laws in biology". Linguistic laws in biology.
- Zipf, George K.. Human behavior and the principle of least effort.. Addison-Wesley Press, 1949.
- Best, Karl-Heinz; Rottmann, Otto. Quantitative linguistics, an invitation. Ram-Verlag, 2017. ISBN 9783942303514.
- Bentz, Christian. Adaptive languages: an information-theoretic account of linguistic diversity. 1st ed. 2018. ISBN 9783110557770.
- Tanaka-Ishii, Kumiko. Statistical universals of language: mathematical chance vs human choice. Springer, 2021. ISBN 9783030593773.
- Levshina, Natalia. Communicative efficiency: Language structure and use. Cambridge University Press, 2023.
- Baayen, R. Harald. Analyzing linguistic data: a practical introduction to statistics using R. Cambridge University, 2008. ISBN 9780521709187.
- Bunge, Mario. La Ciencia: su método y su filosofía. Ed. corr. y aum. Sudamericana, 1995. ISBN 9500710439.

Complementary:
- Herdan, Gustav. The calculus of linguistic observations. Reprint 2021. 1962. ISBN 9783112415443.
- Debowski, Lukasz. Information theory meets power laws: Stochastic processes and language models. Wiley, 2021.

RESOURCES

Hyperlink:
- https://cqllab.upc.edu/biblio/laws/