GHENT UNIVERSITY

Course Specifications

Valid as from the academic year 2026-2027

NLP and Linguistic Analysis (A005870)

Course size *(nominal values; actual values may depend on programme)*

Credits 5.0 **Study time 150 h**

Course offerings in academic year 2026-2027

A (semester 1)	English	Gent

Lecturers in academic year 2026-2027

Macken, Lieve	LW22	lecturer-in-charge
Tezcan, Arda	LW22	co-lecturer
Van Hee, Cynthia	LW22	co-lecturer

Offered in the following programmes in 2026-2027	crdts	offering
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, English, Language Technology)	5	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, French, Language Technology)	5	A
Bachelor of Arts in Applied Language Studies: a combination of at least two languages(main subject Dutch, German, Language Technology)	5	A
Postgraduate Certificate Computer-Assisted Language Mediation	5	A
Preparatory Course Master of Arts in Multilingual Communication: a combination of at least two languages	5	A
Preparatory Course Master of Arts in Translation: a combination of at least two languages	5	A

Teaching languages

English

Keywords

Natural Language Processing, Linguistic analysis

Position of the course

In this course we cover the basics of natural language processing. We will learn how algorithms can be used to teach computers to understand artificial language. Given the ambiguity of language, we start our discussion at the morphological and word level, build up through the syntactic level, and end with the complexity of semantic and discourse modelling.
The aim of this course is twofold. On the one hand, it aims to provide students with a theoretical insight into the basic principles of natural language processing. On the other hand, it builds on previously acquired programming skills in Python and teaches students how to integrate available libraries, such as spaCy, into Python code. By doing so, they actively apply the acquired knowledge in developing NLP applications.
This course builds on the course Introduction to Processing Language with Python and is taught over 5 weeks (weeks 4-8).

Contents

The course deals with the following topics:

- Basic principles of text (pre)processing: tokenisation, part-of-speech tagging, named entity recognition, lemmatising, dependency parsing
- Corpus analysis with a focus on word frequency, lexical diversity, density and sophistication, and topic analysis
- Application of regular expressions, edit distance and dynamic programming
- Syntactic analysis with a focus on PoS tagging and parsing
- Introduction to language models and representing meaning using word vectors

- Ethical issues related to language models and data processing, such as bias
- Integration of existing NLP libraries, such as spaCy, into Python code
- Visualisation of linguistic data to support text analysis and interpretation

Initial competences

A good basic knowledge of Python programming (see final competences 'Introduction to Processing Language with Python')

Final competences

1 Students will have a basic knowledge and understanding of Natural Language Processing (NLP) as a fundamental subfield within Artificial Intelligence.
2 Students will have the practical knowledge and skills necessary to integrate existing NLP libraries into Python code, to perform linguistic analysis of texts and to visualise this information.

Conditions for credit contract

This course unit cannot be taken via a credit contract

Conditions for exam contract

This course unit cannot be taken via an exam contract

Teaching methods

Seminar, Lecture, Independent work

Extra information on the teaching methods

Lectures and hands-on seminars
Guided self-study at home.
Recordings of the lectures will be made available

Study material

Type: Slides

Name: NLP and linguistic analysis
Indicative price: Free or paid by faculty
Optional: no
Language : English
Available on Ufora : Yes

Type: Software

Name: Google Colab Notebooks
Indicative price: Free or paid by faculty
Optional: no

References

- Daniel Jurafsky & James Martin, "Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition" (ed3).
- Python Software Foundation. Official Python documentation. http://www.python.org/doc
- SpaCy. https://spacy.io/

Course content-related study coaching

Individual coaching after class or after appointment with the tutor.

Assessment moments

end-of-term and continuous assessment

Examination methods in case of periodic assessment during the first examination period

Written assessment

Examination methods in case of periodic assessment during the second examination period

Written assessment

Examination methods in case of permanent assessment

Participation, Assignment

Possibilities of retake in case of permanent assessment

examination during the second examination period is possible

Extra information on the examination methods

First session:

- Assignments and class attendance (10%)
- Written exam (90%)

Second session:

- Written exam (100%)

Calculation of the examination mark

A combination of continuous assessment (10%) and end-of-term assessment (90%)

Facilities for Working Students

Class attendance is strongly recommended.
Limited possibility of feedback via e-mail, restricted to answering specific questions.