# linfo2263
*2023*

**UCLouvain**

# Computational Linguistics

| 5.00 credits | 30.0 h + 15.0 h | Q1 |
| --- | --- | --- |

| | |
| --- | --- |
| Teacher(s) | Dupont Pierre ; |
| Language : | English<br>> French-friendly |
| Place of the course | Louvain-la-Neuve |
| Main themes | • Various levels of linguistic analysis<br>• Corpus processing<br>• Part-of-speech tagging<br>• Probabilistic language modeling (N-grams and Hidden Markov Models)<br>• Formal grammars and parsing algorithms<br>• Machine translation, deep learning<br>• Linguistics engineering applications such as automatic completion software, POS tagging, parsing or machine translation |
| Learning outcomes | **At the end of this learning unit, the student is able to :** |

Given the learning outcomes of the "Master in Computer Science and Engineering" program, this course contributes to the development, acquisition and evaluation of the following learning outcomes:

INFO1.1-3

INFO2.3-4

INFO5.3-5

INFO6.1, INFO6.4

Given the learning outcomes of the "Master [120] in Computer Science" program, this course contributes to the development, acquisition and evaluation of the following learning outcomes:

SINF1.M4

SINF2.3-4

SINF5.3-5

SINF6.1, SINF6.4

1    Students completing successfully this course should be able to

- describe the fundamental concepts of natural language modeling
- master the methodology of using linguistic resources, in particular large scale corpora, possibly annotated or structured
- apply in a relevant way statistical language modeling techniques
- implement recent machine learning methods applied to language processing
- develop linguistic engineering applications

Students will have developed skills and operational methodology. In particular, they have developed their ability to

- integrate a multidisciplinary approach between computer science and linguistics, using wisely the terminology, tools and existing methods,
- manage the time available to complete projects of medium size,
- manipulate and exploit large amounts of data.

| | |
|---|---|
| Evaluation methods | ## Computation of the overall course grade<br><br>The projects are worth 30 % of the final grade, 70 % for the final exam (closed-book).<br><br>The projects **cannot** be implemented again in second session. The global project grade is fixed at the end of the semester and included as such in the global score for the second session.<br><br>The final exam is, by default, a written exam (on paper or, when appropriate, on a computer).<br><br>## Rules for student collaboration and use of external resources<br><br>Collaborative studying among students is encouraged during project follow-up sessions and via an exchange forum on Moodle.<br><br>Each student is expected to submit a personal solution to each project. The use of public resources (e.g. stackoverflow.com), including generative AIs (e.g. chatGPT) is permitted, as long as each (fragment of) code submitted by the student mentions all the resources used.<br><br>The distribution or exchange between students of (fragments of) code is not authorized by any means (GitHub, Facebook, Discord, etc.), even after the project deadlines.<br><br>Failure to comply with these rules for any project may result in an overall grade of 0 for all projects.<br><br>These rules are explained in detail during the first class (see course Moodle site). |
| Teaching methods | • Lectures<br>• Practical projects implemented in Python on the *Inginious* platform. |
| Content | • Various levels of linguistic analysis<br>• (Automated) corpus processing: formating, tokenization, data tagging<br>• Probabilistic language models: N-grams, HMMs<br>• Part-of-Speech Tagging<br>• Introduction to Deep Learning<br>• Question Answering<br>• Machine Translation<br>• Typical linguistic applications such as automated text generation, POS taggers, machine translation or chatbots |
| Inline resources | moodle.uclouvain.be/course/view.php?id=1182 |
| Bibliography | One recommended textbook - un ouvrage conseillé :<br><br>• Speech and Language Processing, D. Jurafsky and J.H. Martin, Prentice Hall. |
| Faculty or entity in charge | INFO |

| Programmes containing this learning unit (UE) | | | | |
|---|---|---|---|---|
| Program title | Acronym | Credits | Prerequisite | Learning outcomes |
| Master [120] in Data Science : Statistic | DATS2M | 5 | | 🔍 |
| Master [120] in Linguistics | LING2M | 5 | | 🔍 |
| Master [120] in Computer Science and Engineering | INFO2M | 5 | | 🔍 |
| Master [120] in Computer Science | SINF2M | 5 | | 🔍 |
| Master [120] in Data Science Engineering | DATE2M | 5 | | 🔍 |
| Master [120] in Data Science: Information Technology | DATI2M | 5 | | 🔍 |