






In view of the health context linked to the spread of the coronavirus, the methods of organisation and evaluation of the learning units could be adapted in different situations; these possible new methods have been - or will be - communicated by the teachers to the students.

5 credits	30.0 h + 15.0 h	Q1
-----------	-----------------	----

Teacher(s)	Dupont Pierre ;Fairon Cédric ;
Language :	English
Place of the course	Louvain-la-Neuve
Main themes	<ul style="list-style-type: none"> • Basics in phonology, morphology, syntax and semantics • Linguistic resources • Part-of-speech tagging • Statistical language modeling (N-grams and Hidden Markov Models) • Robust parsing techniques, probabilistic context-free grammars • Linguistics engineering applications such as spell or syntax checking software, POS tagging, document indexing and retrieval, text categorization
Aims	<p>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:</p> <ul style="list-style-type: none"> • INFO1.1-3 • INFO2.3-4 • INFO5.3-5 • INFO6.1, INFO6.4 <p>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:</p> <ul style="list-style-type: none"> • SIN1.M4 • SIN2.3-4 • SIN5.3-5 • SIN6.1, SIN6.4 <p>1 Students completing successfully this course should be able to</p> <ul style="list-style-type: none"> • describe the fundamental concepts of natural language modeling • master the methodology of using linguistic resources (corpora, dictionaries, semantic networks, etc) and make an argued choice between various linguistic resources • apply in a relevant way statistical language modeling techniques • develop linguistic engineering applications <p>Students will have developed skills and operational methodology . In particular , they have developed their ability to</p> <ul style="list-style-type: none"> • integrate a multidisciplinary approach to the edge between computer science and linguistics , using wisely the terminology and tools of one or the other discipline , • manage the time available to complete mini -projects , • manipulate and exploit large amounts of data . <p>-----</p> <p><i>The contribution of this Teaching Unit to the development and command of the skills and learning outcomes of the programme(s) can be accessed at the end of this sheet, in the section entitled "Programmes/courses offering this Teaching Unit".</i></p>
Evaluation methods	<p>Due to the COVID-19 crisis, the information in this section is particularly likely to change.</p> <p>The mini-projects are worth 25 % of the final grade, 75 % for the final exam (closed-book). The mini-projects can NOT be implemented again in second session. The 25 % for the mini-projects are fixed at the end of the semester and included as such in the global score for the second session. The final exam is, by default, a written exam (on paper or, when appropriate, on a UCLouvain computer).</p>

Teaching methods	<p>Due to the COVID-19 crisis, the information in this section is particularly likely to change.</p> <ul style="list-style-type: none"> • Lectures • Mini-projects (2 to 3 weeks) implemented, by default, in Python and by groups of 2 students • Feedback sessions about the projects
Content	<ul style="list-style-type: none"> • Linguistic essentials: morphology, part-of-speech, phrase structure, semantics and pragmatics • Corpus analysis: formating, tokenization, morphology, data tagging • Probabilistic language models: N-grams, HMMs • Part-of-Speech Tagging • Probabilistic Context-Free Grammars: parameter estimation and parsing algorithms, tree banks • Introduction to Machine Translation • Lexical semantics • Information extraction • Typical Applications: POS taggers, information extraction tools, probabilistic parsers
Inline resources	<p>http://moodleucl.uclouvain.be/course/view.php?id=7865</p>
Bibliography	<p>One recommended textbook - un ouvrage conseillé :</p> <ul style="list-style-type: none"> • Speech and Language Processing (2nd Edition), D. Jurafsky and J.H. Martin, Prentice Hall, 2009.
Faculty or entity in charge	<p>INFO</p>

Programmes containing this learning unit (UE)				
Program title	Acronym	Credits	Prerequisite	Aims
Master [120] in Data Science Engineering	DATE2M	5		
Master [120] in Computer Science and Engineering	INFO2M	5		
Master [120] in Linguistics	LING2M	5		
Master [120] in Computer Science	SINF2M	5		
Master [120] in Data Science : Statistic	DATS2M	5		
Master [120] in Data Science: Information Technology	DAT12M	5		