

COMP-599: Introduction to Natural Language Processing

McGill University, Fall 2016

Course Details

Instructor:	Jackie Chi Kit Cheung
Office:	McConnell Engineering Building (MC) 108N
Office hour:	W 14:30–15:30
Contact info:	jcheung@cs.mcgill.ca
Lecture room:	McConnell Engineering Building (MC) 103
Class times:	MW 13:05–14:25

Teaching Assistant:	Jad Kabbara
Contact info:	jad.kabbara@mail.mcgill.ca

Important Links

- myCourses : <http://www.mcgill.ca/lms/>

Course Description

This course presents an introduction to the computational modelling of natural language. Topics covered include: computational morphology, language modelling, syntactic parsing, lexical and compositional semantics, and discourse analysis. We will consider selected applications such as automatic summarization, machine translation, and speech processing. We will also study machine learning algorithms that are used in natural language processing.

Learning Outcomes

By the end of the course, students should have a broad understanding of the field of natural language processing. They should have a sense of the capabilities and limitations of current natural language technologies, and some of the algorithms and techniques that underlie these technologies. They should also understand the theoretical underpinnings of natural language processing in linguistics and formal language theory.

Course Prerequisites and Textbooks

Prerequisites:

- Probabilities and statistics (MATH-323 or ECSE-305)
- Algorithms (COMP 251/252)

- Programming experience

Optional, useful background:

- Artificial intelligence (COMP 424)
- Introductory linguistics (LING 201).

Textbooks:

- *Speech and Language Processing*, any edition. Jurafsky and Martin.

You may purchase a copy of this textbook from the McGill bookstore or through an online retailer. Drafts of selected chapters from the third edition are available online.

Other references:

- *Foundations of Statistical Natural Language Processing*. Manning and Schuetze.
- *Natural Language Toolkit*. Bird and Loper, and other developers. Available for free at:
– <http://www.nltk.org/>

Topics

This list is tentative and subject to modifications.

- Introduction to NLP
- Computational morphology
- Finite-state machines
- Language modelling
- Syntax and parsing
- POS tagging
- Lexical semantics
- Compositional semantics
- Computational discourse
- NLP applications
- Computational linguistics proper (e.g., historical linguistics, language acquisition)
- Machine learning for NLP

Grading Scheme and Deadline Policy

Your final grade in the course is calculated as follows:

- **Assignments:** $10\% \times 4 = 40\%$
- **In-class Midterm Examination:** 20%
- **Research Project:** 30%
- **Oral final exam:** 10%

Students who receive unsatisfactory final grades will **NOT** have the option to submit additional work in order to improve their grades.

Official language policy for graded work: In accord with McGill University's Charter of Students' Rights, students in this course have the right to submit in English or in French any written work that is to be graded.

Assignments

There will be **four** assignments, worth 10% each, which will involve programming, solving problem sets, and reading and writing academic papers. To receive full grades, assignments (as well as all other course work) **MUST** represent your own personal efforts (see the section on Plagiarism Policy and Assignments below).

Late Policy: Late assignments will not be accepted except with a valid medical excuse.

Assignment submission will take place on paper in class, and on myCourses, as described in the assignment handout. Every student is responsible for verifying that their submissions are successful. If you believe the content of your myCourses submission box is different from what you have submitted, you must e-mail me immediately to provide evidence of your correct submission.

The instructor reserve the right to modify the lateness policy for a particular assignment; any such modifications will be clearly indicated at the beginning of the relevant assignment specifications. **Plan appropriately and do not submit to myCourses only minutes before the assignment deadline.**

Midterm Examination

The midterm examination is scheduled for the following date in class:

- Wednesday, November 9th from 13:00 to 14:30

There will be no alternate time for the midterm. If you will miss the midterm due to extenuating circumstances, please e-mail me immediately.

Final Project and Oral Exam

You will conduct a research project and submit a written report on it, worth 30% of your final grade. In addition, you will present your research project in an oral final exam, to be scheduled during the exam period. This presentation will be worth 10% of your final grade. More details of the project and presentation will follow.

Personal Computers and Required Software

You will use Python 2.7 to complete the programming parts of the assignments. You can use any editor or environment of your choice to write your programs. You will be expected to install several Python packages for natural language processing for the assignments.

Plagiarism Policy

Official policy: McGill University values academic integrity. Therefore all students must understand the meaning and consequences of cheating, plagiarism, and other academic offenses under the Code of Student

Conduct and Disciplinary Procedures (see www.mcgill.ca/integrity/ for more information).

Plagiarism Policy and Assignments

You must include your name and McGill ID number at the top of each program or module that you implement and submit. By doing so, you are certifying that the program or module is entirely your own, and represents only the result of your own efforts.

Work submitted for this course must represent your own efforts. Assignments **must** be done **individually**; you **must not** work in groups. You **must not** copy any other person's work in any manner (electronically or otherwise), even if this work is in the public domain or you have permission from its author to use it and/or modify it in your own work (obviously, this prohibition does not apply to source code supplied by the instructor explicitly for this purpose). Furthermore, you **must not** give a copy of your work to any other person.

The plagiarism policy is not meant to discourage interaction or discussion among students. You are encouraged to discuss assignment questions with the instructor, TA, and your fellow students. However, there is a difference between discussing ideas and working in groups or copying someone else's solution. A good rule of thumb is that when you discuss assignments with your fellow students, you should not leave the discussion with written notes. Also, when you write your solution to an assignment, you should do it on your own.

Students who require assistance with their assignments should see the TA or instructor during their office hours. If you have only partially finished an assignment, **document the parts that do not work**, and submit what you managed to complete for partial credit.

We may be using automated software similarity detection tools to compare your assignment submissions to that of all other students registered in the course, and these tools are very effective at what they have been designed for. However, note that the main use of these tools is to determine which submissions should be manually checked for similarity by an instructor or TA; we will not accuse anyone of copying or working in groups based solely on the output of these tools.

You may also be asked to present and explain your assignment submissions to an instructor at any time.

Students who put their name on programs or modules that are not entirely their own work will be referred to the appropriate university official who will assess the need for disciplinary action.