

# Course Outline – Natural Language Processing

## General Information

Course #	COMP 550
Term / Year	Fall 2020
Prerequisites	MATH 323 or ECSE 305, COMP 251 or COMP 252 COMP 424 and LING 201 provide optional useful background.
Course schedule	Lectures released Mondays and Wednesdays
Number of credits	3

## Instructor Information

Name and title	Jackie Chi Kit Cheung, Prof.
E-mail	<a href="mailto:jcheung@cs.mcgill.ca">jcheung@cs.mcgill.ca</a>
Virtual office hours	To be posted on myCourses

## TA Information

Names and e-mails	Ali Emami	<a href="mailto:ali.emami@mail.mcgill.ca">ali.emami@mail.mcgill.ca</a>
	Zhi Wen	<a href="mailto:zhi.wen@mail.mcgill.ca">zhi.wen@mail.mcgill.ca</a>
	Dora Jambor	<a href="mailto:dora.jambor@mail.mcgill.ca">dora.jambor@mail.mcgill.ca</a>
	David Venuto	<a href="mailto:david.venuto@mail.mcgill.ca">david.venuto@mail.mcgill.ca</a>
	Ashita Diwan	<a href="mailto:ashita.diwan@mail.mcgill.ca">ashita.diwan@mail.mcgill.ca</a>

## Course Overview

An introduction to the computational modelling of natural language, including algorithms, formalisms, and applications. Computational morphology, language modelling, syntactic parsing, lexical and compositional semantics, and discourse analysis. Selected applications such as automatic summarization, machine translation, and speech processing. Machine learning techniques for natural language processing.

## Message Regarding Remote Delivery

Lectures will be pre-recorded and posted on myCourses. This term will be unusual, and may present extra challenges for many students. We will do our best to provide a supportive learning environment, and to actively engage you in this course through contact hours with the teaching staff, and discussion opportunities with other students.

## 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 Materials

*Speech and Language Processing*. Jurafsky and Martin. 2<sup>nd</sup> edition.

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 at <https://web.stanford.edu/~jurafsky/slp3/>.

### Other references:

- *Natural Language Toolkit*. Bird and Loper, and other developers. Available for free at <http://www.nltk.org/>
- *Natural Language Processing*. Eisenstein. Draft available for free at <https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf>
- *Foundations of Statistical Natural Language Processing*. Manning and Schütze.

## Course Content

This list is tentative and subject to modifications.

- Introduction to NLP
- Language modelling
- Syntax and parsing
- Part-of-speech tagging
- Lexical semantics
- Compositional semantics
- Computational discourse and pragmatics
- NLP applications
- Other topics in computational linguistics (e.g., historical linguistics, language acquisition)
- Machine learning for NLP

## Evaluation

Group final project	40%
Programming assignments	20%
Reading assignments	20%
Study groups	10%
Online quizzes	10%

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

## Final Project

You will be expected to complete a final project in a group of three on a topic of your choice in NLP. This project will give you a chance to explore one topic in more depth. You will be expected to turn in a

written report and any accompanying code or dataset at the end of the term. More details will be released about the expectations and criteria.

### **Assignments**

There will be two types of assignments: programming and reading. To receive full grades, assignments (as well as all other course work) must represent your own personal efforts unless otherwise stated on the assignment handout.

**Late policy.** Assignments that are less than 15 minutes late will be accepted without penalty. Assignments that are between 15 minutes and 24 hours late will be accepted with a 10% absolute penalty. Assignments that are more than 24 hours late will not be accepted without a valid extenuating circumstance.

Assignment submission will take place on myCourses. 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 reserves 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.**

### **Study Groups**

You will be expected to participate in online group discussions with other students, and then to write up a short summary of your discussions to be submitted on myCourses. We will make efforts to accommodate any issues with technology, time zone differences, or other personal circumstances. We will release more details and guidelines for these discussions.

### **Online Quizzes**

Short online quizzes will be posted on myCourses. The purpose of these quizzes is to ensure that you follow along with the lectures, and that you can check if you have understood the content. You will be allowed multiple trials to correct any mistakes. These quizzes must be completed individually.

### **Plagiarism Policy**

**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 unless otherwise stated. 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. The only exceptions are for source code supplied by the instructor explicitly for an assignment, and for the final project, where usual research citation practices apply. 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 use 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.

## **Language of Submission**

"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. This does not apply to courses in which acquiring proficiency in a language is one of the objectives."

« Conformément à la Charte des droits de l'étudiant de l'Université McGill, chaque étudiant a le droit de soumettre en français ou en anglais tout travail écrit devant être noté (sauf dans le cas des cours dont l'un des objets est la maîtrise d'une langue). »

## **Academic Integrity**

"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 McGill's guide to academic honesty (<https://www.mcgill.ca/students/srr/honest>) } for more information)."