Natural Language Processing, COMP 550, Fall 2024

Prerequisites:	MATH 323 or ECSE 305, COMP 251 or COMP 252 COMP 424 and LING 201 provide optional useful background.
Course schedule:	Mondays and Wednesdays, 11:35am – 12:55pm Macdonald-Harrington Building, Room G-10
Instructors:	David Adelani, david.adelani@mcgill.ca Office hours: Wednesdays 2pm - 3:30pm in McConnell 204N Jackie Chi Kit Cheung, jackie.cheung@mcgill.ca Office hours: Monday 2pm - 3:30pm in McConnell 108N
TA(s):	Shira Abramovich, shira.abramovich@mail.mcgill.ca Ziling Cheng, ziling.cheng@mail.mcgill.ca Gaurav Iyer, gaurav.iyer2@mail.mcgill.ca Xijuan Sun, xijuan.sun@mail.mcgill.ca Zihan Wang, zihan.wang5@mail.mcgill.ca

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.

Recommended Course Materials

Speech and Language Processing. Jurafsky and Martin. 2nd 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.

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 Content

This list is tentative and subject to modifications.

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

Means of Assessment

Title	Weight	Description	Due Date	Considerations and Late Penalties
Midterm Exam	25%	In person, multiple choice and short answer.	November 6, 2024	Make-up exam to be held roughly one week later for students with a valid reason for missing the assessment.
Programming Assignments	2 x 10%	Includes programming and written portions.	Periodically throughout the term; roughly late September and November	Late assignments accepted up to 24 hours after posted deadline without penalty.
Reading Assignments	4 x 5%	Critical summaries of research papers	Periodically throughout the term; roughly once a month	Late assignments accepted up to 24 hours after posted deadline without penalty.
Final Project	35%	Research project to be completed in a group of three. Includes a proposal	Last day of class	Late submissions accepted up to 2 weeks after last day of class.

		report deadline in the middle of the term.		

Assessments in this course are governed by the Policy on Assessment of Student Learning (PASL), which provides a set of common principles to guide the assessment of students' learning. Also see Faculty of Science-specific rules on the implementation of PASL.

Legally mandated academic accommodations are handled by Student Accessibility and Achievement. For more information see https://www.mcgill.ca/access-achieve/

In accord with McGill University's Charter of Students' Rights, students in this course have the right to submit in English or in French 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." (Approved by Senate on 21 January 2009)

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. (Énoncé approuvé par le Sénat le 21 janvier 2009)

McGill University values academic integrity. Therefore, all students must understand the meaning and consequences of cheating, plagiarism and other academic offences under the Code of Student Conduct and Disciplinary Procedures" (Approved by Senate on 29 January 2003) (See McGill's guide to academic honesty for more information).

In the event of extraordinary circumstances beyond the University's control, the content and/or assessment tasks in this course are subject to change and students will be advised of the change.

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.

Use of generative AI technologies. Unless otherwise specified by the instructor, you may use generative AI technologies (e.g., Chat-GPT, Bard) only in an assistive manner (e.g., to help understand course content, to search for information, to help brainstorm ideas, or to check for grammar and style errors). You may not use such technologies as the primary means with which to complete your assignments or your final project. All substantive uses of these tools must be acknowledged in your submissions.

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.