

LING/COMP 445, LING 645: Computational Linguistics, Fall 2019

Mon/Wed 16:05 (04:05 PM) – 17:25 (05:05 PM), ARTS W 20

Instructor	Tim O'Donnell timothy.odonnell@mcgill.ca 1085 Dr. Penfield, 304 Office hours: By Appointment
Grader	Vanna Willerton savanna.willerton@mail.mcgill.ca 1085 Dr. Penfield, 301

1 Overview

The fields of computer science and modern generative linguistics have grown together since their very beginnings. In this course, we will learn about the tools that have been developed over the years for the analysis of natural language in a computational setting. The emphasis will be on learning the methods and logic underlying the computational generative approach, rather than on learning recent applications from natural language processing.

2 Goals

The goal of this course is to learn how to think about linguistic structure formally, define computational models of linguistic structure, and evaluate their ability to account for empirical linguistic data rigorously.

Students will:

- Learn how to formalize notions like languages and grammars
- Learn how to define probability distributions over formal languages and grammars
- Learn how to work with such probability distributions algorithmically
- Learn how to evaluate models empirically
- Develop practical programming skills and concepts that support the above goals

3 Prerequisites

Programming background at the level of COMP 250 or equivalent. Mathematics background at the level of MATH 240 or equivalent. Basic calculus will be helpful but not critical. Introductory level linguistics at the level of LING 201, especially introductory syntax will be helpful but not critical. We will emphasize building up all basic tools from scratch, but the class will move fast.

4 McGill Policy Statements

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 (see www.mcgill.ca/students/srr/honest/ for more information).

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.

Instructor generated course materials (e.g., handouts, notes, summaries, exam questions, etc.) are protected by law and may not be copied or distributed in any form or in any medium without explicit permission of the instructor. Note that infringements of copyright can be subject to follow up by the University under the Code of Student Conduct and Disciplinary Procedures.

5 Logistics

Piazza website: <https://piazza.com/mcgill.ca/fall2019/compling445645/home>

Course Webbook: <https://foundations-computational-linguistics.github.io/>

Other Materials: Any other materials that are needed over the course of the term will be made available electronically.

6 Readings

The class will make use of an interactive textbook using the Clojure programming language. Other materials will be made available online when appropriate.

7 Course structure

The course will be taught using an interactive textbook built using the Clojure programming language. The class is structured around this textbook and the problem sets we will cover in this class.

There will be four substantial problem sets, a midterm examination, and a final problems set of the same format as the problem sets but due after the end of classes.

Problems will generally consist of a mix of: (i) programming problems in Clojure (ii) mathematical problems and (iii) linguistics thought exercises. Problems will be submitted as ClojureScript files or L^AT_EX-based PDFs via the email to the course grader and instructor (i.e., timothy.odonnell@mcgill.ca and savanna.willerton@mail.mcgill.ca) as ClojureScript files (i.e., with an `.clj` extension). Problem sets are due before the beginning of class at 16:05 in the afternoon on the due dates specified below.

Participation and interaction is encouraged in this course. We will sometimes use a short amount of class time (15–20 minutes) to cover questions on preceding problem set or other difficult material.

Class communication and discussion will take place through the Piazza website.

8 Evaluation

Note that details below are subject to change.

- Term Time Examination (25%)

Problem Sets (60%) 4 problem sets equally weighted.

Final Problem Set (15%) This problem set will be the same format as the others, but will be completed after the end of class.

9 Course schedule

Note that the topics covered below and exact dates are subject to change, depending on how quickly we make progress through the topics the course.

- Wednesday, September 4, 2019: Introduction/Models of Computation
- Monday, September 9, 2019: Chapter 1/2
- Wednesday, September 11, 2019: Chapter 3/4
- Monday, September 16, 2019: Chapter 5/6
 - 📧: Problem Set 1 Released
- Wednesday, September 18, 2019: Chapter 7
- Monday, September 23, 2019: Chapter 8
- Wednesday, September 25, 2019: Chapter 9

- Monday, September 30, 2019: Chapter 10
- Wednesday, October 2, 2019: Chapter 11
- Monday, October 7, 2019: Chapter 12
 - 📎: Problem Set 1 Due
 - 📎: Problem Set 2 Released
- Wednesday, October 9, 2019: Chapter 13
- Monday, October 14, 2019: No Class
 - ✓: Problem Set 1 Graded
- Wednesday, October 16, 2019: Chapter 14/15
- Monday, October 21, 2019:
- Wednesday, October 23, 2019:
- Monday, October 28, 2019:
 - 📎: Problem Set 2 Due
 - 📎: Problem Set 3 Released
- Thursday, October 25, 2018:
- Wednesday, October 30, 2019:
- Monday, November 4, 2019:
 - ✓: Problem Set 2 Graded
- Wednesday, November 6, 2019:
- Monday, November 11, 2019:
 - 📎: Problem Set 3 Due
 - 📎: Problem Set 4 Released
- Wednesday, November 13, 2019: Review
- Monday, November 18, 2019: Exam
 - ✓: Problem Set 3 Graded
- Wednesday, November 20, 2019:
- Monday, November 25, 2019:
 - 📎: Problem Set 4 Due
 - 📎: Problem Set 5 Released
- Wednesday, November 27, 2019:
- Monday, December 2, 2019:
 - ✓: Problem Set 4 Graded
- Tuesday, December 3, 2019: Last Class
- Wednesday, December 18, 2019: After Classes
 - 📎: Problem Set 5 Due