LING 483/682: Computational Linguistics 2, Winter 2020

Tue/Thur 11:35 – 12:55, PE1085 002

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. This course follows COMP/LING 445/645 in sequence, using the same materials and picking up where that course left off.

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

Having previously taken COMP/LING 445/645 is a strict requirement for this course.

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 criticial. 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 and will start where COMP/LING 445/645 left off.

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

MyCourses website: `https://mycourses2.mcgill.ca/d2l/home/441875`

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 if 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 three substantial problem sets and a final problem set of the same format as the other problem sets but due after the end of classes.

Problems will generally consist of some mix of: (i) programming problems in Clojure (ii) mathematical problems and (iii) linguistics thought exercises. Problems will be submitted as ClojureScript files (i.e., with an `.clj` extension) or LaTeX-based pdfs. Problem sets must be submitted on MyCourses before 11:30 (class time) on each due date.

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 MyCourses website.

8 Evaluation

Note that details below are subject to change.

- Class Attendance and Participation (20%)
- Problem Sets (60%) 3 problem sets equally weighted.
- Final Problem Set (20%) This problem set will be the same format as the others, but will be completed after the end of class.

For the three term-time problem sets, we will allow students to resubmit corrections within one week after they are returned for up to half the points that were missed. This will not be possible for the final problem set due to time constraints.

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.

- Tuesday, January 7, 2020: Regular Languages
- Thursday, January 9, 2020:
- Tuesday, January 14, 2020:
- Thursday, January 16, 2020

- Tuesday, January 21, 2020
- Thursday, January 23, 2020
- Tuesday, January 28, 2020
  - ✎: Problem Set 1 Released
- Thursday, January 30, 2020
- Tuesday, February 4, 2020
- Thursday, February 6, 2020
- Tuesday, February 11, 2020
- Thursday, February 13, 2020
- Tuesday, February 18, 2020
  - ✐: Problem Set 1 Due
  - ✎: Problem Set 2 Released
- Thursday, February 20, 2020
- Tuesday, February 25, 2020
  - ✔: Problem Set 1 Graded
- Thursday, February 27, 2020
- Tuesday, March 3, 2020: Study Break
  - ✐: Problem Set 1 Corrections Due
- Thursday, March 5, 2020: Study Break
- Tuesday, March 10, 2020
  - ✐: Problem Set 2 Due
  - ✎: Problem Set 3 Released
- Thursday, March 12, 2020
- Tuesday, March 17, 2020
  - ✔: Problem Set 2 Graded
- Thursday, March 19, 2020
- Tuesday, March 24, 2020
  - ✐: Problem Set 2 Corrections Due
- Thursday, March 26, 2020
- Tuesday, March 31, 2020
  - ✐: Problem Set 3 Due
  - ✎: Problem Set 4 Released
- Thursday, April 2, 2020
- Tuesday, April 7, 2020
  - ✔: Problem Set 3 Graded
- Thursday, April 9, 2020: Last Class

- Tuesday, April 14, 2020: After Classes
  - ✐: Problem Set 3 Corrections Due
- Tuesday, April 21, 2020: After Classes
  - ✐: Problem Set 4 Due