

CSC 485/2501: Introduction to Computational Linguistics
Course Information

Instructor: Gerald Penn
CSC485H1F LEC0101 MWF 11–12 ES B149
Lectures: CSC485H1F LEC0201 MWF 12–1 ES B149
CSC485H1F LEC2001 MWF 11–12 ES B149
CSC2501HF LEC0101 MWF 12–1 ES B149
(Note: some lecture days will be used for tutorials)
Office: BA 283A
Tel: (416)978-7390
Office Hours: MW 1–2, or by appointment
Email: gpenn@teach.cs.utoronto.ca

Remote Attendance: All in-person lectures and office hours are suspended until 23rd September. During this time, lectures will take place on Zoom, and office hours will take place on the class Discord server. The necessary Zoom and Discord links have been made available to registered students on Quercus.

After 23rd September, in-person lectures will take place at the times listed above, and will be simulcast with audio and synchronized slides over Zoom. Office hours will be conducted both in person and on Discord, but priority will be given to in-person attendees.

	Name	Assignment
Teaching Assistants:	Aditya Bhargava	1
	Jinman Zhao	2
	Frank Niu	3

Textbooks:

- Required Jurafsky, Daniel, and Martin, James H. *Speech and Language Processing*, 2nd edition, Pearson Prentice-Hall, 2009. Available in paper and e-book rental versions (for the latter, go to CourseSmart.com and search for *Jurafsky*). We'll also be referring to the draft 3rd edition: <https://web.stanford.edu/~jurafsky/slp3/>. See also the errata list for the 2nd edition: www.cs.colorado.edu/~martin/SLP/Errata/SLP2-PH-Errata.html.
- Required Bird, Steven; Klein, Ewan; and Loper, Edward. *Natural Language Processing with Python*, O'Reilly, 2009. Free (in HTML) with online extras at www.nltk.org/book.
- Recommended Mertz, David. *Text Processing in Python*. Addison-Wesley, 2003. Free ASCII version at Gnosis.cx/TPiP.
- Optional Allen, James. *Natural Language Understanding*, 2nd edition. Benjamin/Cummings, 1995.
- Recommended Martelli, Ravenscroft and Holden. *Python in a Nutshell*, 3rd ed., O'Reilly, 2017.

Course Web Page: <http://www.cs.toronto.edu/~gpenn/csc485/>

Evaluation: For undergraduates registered in CSC 485, there will be three homework assignments worth one third of your final mark each. Those registered for CSC 2501 must complete the three homework assignments (25% each), as well as five essays on assigned research papers ($5 \times 5 = 25\%$). There is no final examination for either course code.

- *No late homeworks will be accepted, except in case of documented medical or other emergencies.*

Policy on collaboration: No collaboration on homeworks or essays is permitted. The work you submit must be your own. No student is permitted to discuss the homeworks with any other student unless the instructor or TAs make the solutions publicly available.

Failure to observe this policy is an academic offense, carrying a penalty ranging from a zero on the homework to suspension from the university.

Course Goals: This course is an introduction to a statistical and computational characterization of natural language. You will also have the chance to practice programming in Python.

Prerequisites: For undergraduates, STA247H1/STA255H1/STA257H1 and CSC209H1, but CSC324H1/CSC330H1/CSC384H1 is strongly recommended. Engineering students may substitute APS105H1/ APS106H1/ ESC180H1/ CSC180H1 for the CSC 209 requirement, although experience with the Unix operating system is strongly recommended, and/or ECE302H1/ STA286H1/ CHE223H1/ CME263H1/ MIE231H1/ MIE236H1/ MSE238H1/ ECE286H1 for the statistics requirement. Note that the University's automatic registration system does not check for prerequisites: even if you have registered for the course, you will not receive credit for it unless you had satisfied the prerequisite before you registered. For advice, contact the Undergraduate Office on the fourth floor of the Bahen Centre or the instructor.

Newsgroup: The course newsgroup is on the web at <https://piazza.com/utoronto.ca/fall2021/csc4852501>. Your teaching assistants will be monitoring it.

Tentative Syllabus:

Date	Topic	Advance reading*
10–17 Sept	Intro to CL	RP ; J&M: 1; BK&L: 1, <i>2.3, 4 as necessary</i>
17–24 Sept	Grammars and parsing	RP ; J&M: 5.0–1, 12.0–12.3.3, 12.3.7, <i>13.1–2</i> ; BK&L: 8.0–8.4
27 Sept– 1 Oct	Lexical semantics	J&M: 19.1–4, 20.8
4–8 Oct	Word sense disambiguation	RP ; J&M: 20.1–5
11–15 Oct	Chart parsing	J&M: 13.3–4; <i>A: 3.4, 3.6</i> ; BK&L: 8.4 and online extras section 8.2 on chart parsing
18–22 Oct	Typed Feature Structures	RP ; J&M: 12.3.4–6, 15.0–3; <i>A: 4.1–5</i> ; BK&L: 9
25–29 Oct	Ambiguity Resolution	
1–5 Nov	Attachment Disambiguation	
15–19 Nov	Stochastic Grammars	RP ; J&M: 5.2–5.5.2, 5.6, 12.4, 14.0–1, 14.3–7
22–24 Nov	Categorial Grammars	
24–26 Nov	Supertagging	
29 Nov– 1 Dec	Dependency Graphs	
3–6 Dec	Question Answering	
8 Dec	Anaphora resolution	J&M: 21.0, 21.2–8
9 Dec	Parsing for FWO Languages	

*J&M = Jurafsky and Martin; BK&L = Bird, Klein, and Loper; A = Allen; **RP** = research paper distributed on-line; *italics indicates optional additional reading*.

Course Calendar:

Fri, 10 September	First lecture
Mon, 20 September	Last day to add course (CSC 2501)
Mon, 20 September	Essay 1 due (CSC 2501)
Wed, 22 September	Last day to add course (CSC 485)
Fri, 1 October	Essay 2 due (CSC 2501)
Fri, 8 October	Assignment 1 due
Mon, 11 October	Thanksgiving holiday
Fri, 15 October	Essay 3 due (CSC 2501)
Mon, 25 October	Last day to drop course (CSC 2501)
Fri, 29 October	Essay 4 due (CSC 2501)
Fri, 5 November	Assignment 2 due
Mon, 8 November	Last day to drop course (CSC 485)
8–12 November	Reading Week — no lectures or tutorial
Fri, 26 November	Essay 5 due (CSC 2501)
Thu, 9 December	Last lecture
Thu, 9 December	Assignment 3 due