University of Toronto
St. George Campus

Computer Science 485/2501
Autumn 2025

CSC 485/2501: Introduction to Computational Linguistics
Course Information

Instructor: Gerald Penn

Lectures:	CSC485H1F LEC0101	M 10–11	BA 1170
		W 10–11	ES B142
		F 10–11	GB 244
	CSC485H1F LEC0201	M 12–1	BA 1190
		W 12–1	GB 244
		F 12–1	BA 1170
	CSC485H1F LEC2001	M 10–11	BA 1170
		W 10–11	ES B142
		F 10–11	GB 244
	CSC2501HF LEC0101	M 12–1	BA 1190
		W 12–1	GB 244
		F 12–1	BA 1170

(Note: some lecture days will be used for tutorials)

Office: Bahen west vestibule (outside top of BA 1190)
Tel: (416)978-7390
Office Hours: MW 11–12, or by appointment
Email: gpenn@teach.cs.utoronto.ca

Teaching Assistants:

Name	Assignment
Samarendra Dash	1
Teddy Wang	2
Andrew Liu	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 `VitalSource.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 on the course webpage.
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, 1994.
Recommended	Martelli, Ravenscroft and Holden. *Python in a Nutshell*, 4th ed., O'Reilly, 2023.

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

Evaluation: For undergraduates registered in CSC 485, there will be three homework assignments worth 30% 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 ($5 \times 3 = 15\%$). Small one-question quizzes (1/3 of a mark each) make up the remaining 10% of your final mark. Class attendance on the same day is required for quiz credit. 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: Collaboration on and discussion of quiz content is encouraged. No collaboration on homeworks or essays is permitted. The work you submit must be your own. You must also not submit code that is partly or entirely AI-generated. No student is permitted to discuss or share homeworks with any other student from either this or previous years unless the instructor or TAs make the solutions publicly available.

For CSC 2501, essays may be generated in part or wholly by AI, if you wish, on the conditions that 1) you must not tell the TAs or the instructor that you used AI tools on your essays until the final day of class, 2) on the final day of class, you must disclose, for each essay, to which extent you did use AI tools, which tools you used and the prompts that you used to generate them (the instructor will document the procedure for disclosure on that day), and 3) you will not be allowed to resubmit any essay on the grounds that you did not write all or part of your original submission.

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, STA237H1/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 checks for prerequisites: even if you have registered for the class, you will be dropped from it unless you had satisfied the prerequisite before you registered or you had received a prerequisite waiver. 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/fall2025/csc4852501`. Your teaching assistants will be monitoring it.

Course Calendar:

Wed, 3 September	First lecture
Fri, 12 September	Essay 1 due (CSC 2501)
Mon, 15 September	Last day to add course (CSC 485)
Wed, 17 September	Last day to add course (CSC 2501)
Fri, 26 September	Essay 2 due (CSC 2501)
Mon, 6 October	Assignment 1 due
Fri, 10 October	Essay 3 due (CSC 2501)
Mon, 13 October	Thanksgiving holiday
Fri, 24 October	Essay 4 due (CSC 2501)
Mon, 27 October	Last day to drop course (CSC 2501)
27–31 October	Reading Week — no lectures or tutorial
Mon, 3 November	Assignment 2 due
Mon, 11 November	Last day to drop course (CSC 485)
Fri, 21 November	Essay 5 due (CSC 2501)
Tue, 2 December	Last lecture
Tue, 2 December	Assignment 3 due

Tentative Syllabus[1]:

Date	**Topic**	**Advance reading***
2–8 Sept	Intro to CL	**RP**; J&M: 1; BK&L: 1, *2.3, 4 as necessary*
10–17 Sept	Grammars and parsing	**RP**; J&M: 5.0–1, 12.0–12.3.3, 12.3.7, *13.1–2*; BK&L: 8.0–8.4
19–24 Sept	Lexical semantics	J&M: 19.1–4, 20.8
26 Sep–1 Oct	Word sense disambiguation	**RP**; J&M: 20.1–5
1–8 Oct	Language Modelling	
10–15 Oct	Chart parsing	J&M: 13.3–4; *A: 3.4, 3.6;* BK&L: 8.4 and online extras section 8.2 on chart parsing
17 Oct	Ambiguity Resolution	
17–22 Oct	Typed Feature Structures	**RP**; J&M: 12.3.4–6, 15.0–3; *A: 4.1–5*; BK&L: 9
24 Oct–5 Nov	Attachment Disambiguation	
7–12 Nov	Stochastic Grammars	**RP**; J&M: 5.2–5.5.2, 5.6, 12.4, 14.0–1, 14.3–7
14–17 Nov	Categorial Grammars	
17–19 Nov	Supertagging	
21–24 Nov	Dependency Graphs	
26–28 Nov	Question Answering	
1 Dec	Anaphora resolution	J&M: 21.0, 21.2–8
2 Dec	Parsing for FWO Languages	

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