LIN4770C/5770C Introduction to Computational Linguistics

T 1:55-2:45; R 1:55-3:50
Matherly 102
Spring 2024

Instructor:	Prof. Zoey Liu	**Office:**	Turlington 4016
Email:	liu.ying@ufl.edu	**Office Hour:**	W 1:55-2:45; R 12:50-1:40
TA:	Minghao Zheng	**Office:**	Dauer 36
Email:	mighao.zheng@ufl.edu	**Office Hour:**	M 13:55-14:45; F 10:40-11:30
TA:	Nitin Venkateswaran	**Office:**	TBD
Email:	venkateswaran.n@ufl.edu	**Office Hour:**	M 10.40-11.40; F 9.30-10.30 (M in person; F on zoom)

Course objectives

This course surveys selected topics and tasks that are fundamental to computational linguistics. Students will learn to gain (1) understanding of different linguistic problems that could be solved with mathematical means; (2) understanding of different computational methods to derive automatic analysis of language structures at different linguistic levels; (3) technical programming skills to model linguistic phenomena in a computational framework.

Prerequisite

Given the title of the course, where the word *computational* functions as the modifier of *linguistics*, **an interest in Linguistics is required**. In addition, students are expected to have taken LIN4930/6932 Programming for Linguists or the equivalent to ensure they have sufficient background in the programming language, `Python`.

Course website

We will be using Canvas as the course website. All lecture notes, code, ans supplementary materials will be posted on Canvas. Grades will be posted to the Canvas grade book.

Textbook

There is no required textbook for the course. At the same time, students are encouraged to use relevant chapters from the following (free!) book as references:

- Speech and Language Processing (3rd ed. draft). Dan Jurafsky & James H. Martin.

Attendance

Attendance is required. You are responsible for all the material that is covered during class, even when your absence from class is excused. While in attendance, you are expected to actively participate. When you are excused, download the lecture notes from Canvas and ask a classmate what else you might have missed.

Laptop

Class meetings will include lectures, quizzes, and programming-related activities. Bring your laptop to class, but leave it closed unless you are writing code or taking a quiz.

Course outline

Note: the following course outline is subject to change.

Week	Topic	Lab	Assignment
Week 1	Introduction; Text Normalization		
Week 2	Regular Expressions with Python; Edit Distance	Lab 1	
Week 3	*N*-gram Language Modeling		Assignment 1
Week 4	Noisy Channel; Statistical Machine Learning	Lab 2	
Week 5	Statistical Machine Learning, continued		Assignment 2
Week 6	Part-of-Speech (POS) Tagging	Lab 3	
Week 7	Formal Language		Assignment 3
Week 8	Context-free Parsing	Lab 4	
Week 9	**Spring Break**		
Week 10	Dependency Parsing		Midterm
Week 11	Dependency Parsing, continued	Lab 5	
Week 12	Word Senses; Vector Semantics		Assignment 4
Week 13	Vector Semantics, Continued; Document Similarity	Lab 6	
Week 14	Topic Modeling; Large Language Models		Assignment 5
Week 15	Final Project Presentation		
Week 16	No class		

Grade breakdown

Participation	5%
Quizzes	10%
Labs	10%
Midterm	15%
Assignments	40%
Final project	
Three ideas	5%
Project poster	10%
Final presentation	5%

Grade scale

We will be using the default canvas grading scheme.

A	94-100	A-	90-94	B+	87-90	B	84-87
B-	80-84	C+	77-80	C	74-77	C-	70-74
D+	67-70	D	64-67	D-	61-64	E	< 61

Participation

Students are expected to actively participate in class.

Quizzes

There will be approximately 6-7 pop quizzes throughout the semester; students are expected to take the quiz in class with their laptop. A student's lowest quiz score will be dropped.

Labs

There will be approximately 8 lab sessions throughout the semester, led by the TAs; students are expected to participate in lab sessions with their laptop.

Assignment & collaboration

There will be a total of five assignments. Students are encouraged to collaborate with each other on questions that involve programming in the assignments; if that were the case, the name(s) of the collaborating students are required to be mentioned as a comment on Canvas upon assignment submission.

For questions that involve qualitative data analysis, students are expected to complete them independently.

Midterm

There will be an in-class midterm; students will take the midterm with pen and paper.

Final project

Students will form teams to complete their final project; the ideal size of the team is 2 people. The final project consists of three components: (1) a write-up of three potential ideas for the final project (one short paragraph for each idea); (2) a poster in PDF format of the project; (3) presentation of final project in class.

Late work

Late quizzes/labs/assignments/final projects are not accepted, unless they are accompanied by a letter of explanation from your class dean or a medical professional.

Academic Integrity

UF students are bound by The Honor Pledge which states, "We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honor and integrity by abiding by the Honor Code." On all work submitted for credit by students at the University of Florida, the following pledge is either required or implied: "On my honor, I have neither given nor received unauthorized aid in doing this assignment." The Honor Code specifies a number of behaviors that are in violation of this code and the possible sanctions. Furthermore, students are obligated to report any condition that facilitates

academic misconduct to appropriate personnel. If students have any questions or concerns, please consult with the Instructor or the TA.

Classroom Conduct

Students and faculty each have responsibility for maintaining an appropriate learning environment. Those who fail to adhere to professional behavioral standards may be subject to discipline. The Instructor pledges to treat each of the students with dignity, respect, and professional courtesy. Students are expected to do the same for the Instructor and for each other.

Attendance Policy, Class Expectations, use of AI, Make-Up Policy, Cell-phone use

- Students are required to submit all assignments and tests before the class period they are due. Please contact the instructor in advance if you need to skip a class, or cannot make a deadline.
- Attendance is mandatory. If you are absent for more than three classes, you will get a warning. If absences persist the instructor can prohibit further attendance and assign a failing grade for excessive absences.
- Requirements for class attendance and make-up exams, assignments, and other work in this course are consistent with university policies that can be found at: `catalog.ufl.edu/UGRD/academic-regulations/attendance-policies/`
- Students need to disclose any use of AI in their assignments, and need to indicate their prompts and their critical (content) edits of the AI output. AI output without prompts or critical edits will not be accepted. AI needs to be properly cited (`https://apastyle.apa.org/blog/how-to-cite-chatgpt`).
- Cell-phone use is not allowed during class unless this is part of the course assignments. Laptops and tablets can only be used to take notes and for in-class assignments.

Accommodation Policies

If a student qualifies for accommodations because of a disability, please submit their accommodation letter from the Disability Resource Center to the Instructor in a timely manner so that their needs can be addressed.

Religious Observances

A student should inform the Instructor of religious observances that will conflict with class attendance, tests or examinations, or other class activities prior to the class or occurrence of that test or activity. The Instructor is obligated to accommodate students' religious observances. See policy details here.

Course Evaluations

Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available here. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via here. Summaries of course evaluation results are available to students here.

Recording lecture content.

Students are allowed to record class lectures. However, the only allowable purposes of these recordings are (1) for personal educational use, (2) in connection with a complaint to the university, or (3) as evidence in, or in preparation for, a criminal or civil proceeding. Specifically, students may not publish recorded lectures without the written consent of the instructor. To "publish" means to share, transmit, circulate, distribute,

or provide access to a recording, regardless of format or medium, to another person (or persons), including but not limited to another student within the same class section. A recording, or transcript of a recording, is considered published if it is posted on or uploaded to, in whole or in part, any media platform, including but not limited to social media, book, magazine, newspaper, leaflet, or third party note/tutoring services. A student who publishes a recording without written consent may be subject to a civil action and/or discipline the Student Honor Code and Student Conduct Code.