LING 110/297: Foundations of Computational Linguistics **Winter 2021**

Logistics	**Instructor:**	Prof. Simon Todd (sjtodd@ucsb.edu) *pronunciation: SIGH-min TAHD; pronouns: he/him/his* **Office hours:** Th 2-3pm
	TA:	Phillip Rogers (phillip@ucsb.edu) *pronunciation: FIL-ip RAH-jurz; pronouns: he/him/his* **Office hours:** M 11am-12pm, W 9-10am
	Classes:	MW 2:00-3:15pm on Zoom. Recordings will be made available for students in other time zones who are unable to attend synchronously.
	Sections:	*Undergrads:* M 4:00-4:50pm or 5:00-5:50pm; *Grads:* T 3:30-4:20pm. Starting in Week 2, on Zoom.
	Office hours:	At the times stated above, on Gather. See GauchoSpace for help.
	Forum:	You can get help through the GauchoSpace forum any time.
	JupyterLab:	You will use the course JupyterLab to retrieve, complete, and submit assignments. See the tutorial on GauchoSpace for setup help.
	Website:	All course information is distributed through GauchoSpace.
Requirements	**Prerequisites:**	LING 194/297 in the Fall, or ability to write code in Python.
	Textbook:	We will use freely-available chapters from the 3rd edition draft of Jurafsky & Martin's *Speech and Language Processing*.
	Computing:	You will need a computer and a stable internet connection.
	GitHub:	You must create a free GitHub account to retrieve and submit materials for this course. See the tutorial on GauchoSpace for help.

Description

Computational linguistics concerns the question of how computers can represent and process human language. Computational linguists develop, analyze, and apply computational models of language, both to create language technologies, and to inform the scientific understanding of language structure and usage. In this course, you will learn about, practice, and critically evaluate some of the most foundational ideas in computational linguistics.

Objectives

During this course, you will:

1. become familiar with foundational ideas in computational linguistics;
2. develop an understanding of the intuitions behind these ideas and methods that utilize them, including relevant algorithms and math;
3. develop the skills to apply your understanding in hands-on programming tasks;
4. think critically about the positive and negative effects of computational linguistic technologies in the real world;
5. identify ways that computational methods can inform the scientific understanding of language structure and usage.

Policies

Workload: This is a 4-unit class, with an expected workload of 12 hours/week:

classes	2.5	hours/week
section	1	hour/week
reading / review	3	hours/week
quizzes / check-in surveys	0.5	hours/week
assignments	5	hours/week

Note: each assignment is expected to take 10 hours, across 2 weeks.

Pandemic: The physical, mental, and emotional health of you and those around you is of utmost importance. **If you are unable to meet deadlines due to factors related to the COVID-19 pandemic, we will work with you to make alternative arrangements.** Please seek appropriate care and contact us as soon as you are able.

Late work: To make sure we can give you feedback and help in a timely manner, all deadlines are hard deadlines. Exceptions can be considered on a case-by-case basis, if you let us know in advance. **Late assignments will have their maximum grade capped by 10% for each day they are late, and will receive no credit if submitted more than 5 days late. Late submissions are not permitted for quizzes, check-in surveys, or reflection essays.**

Attendance: This course is synchronous. **You are expected to attend all classes and sections.** If time zones prevent you from attending, please let us know at the beginning of the quarter, so we can set you up with access to recordings of classes.

Conduct Code: We follow UCSB's Student Conduct Code. You may discuss your assignments with other students, and consult information online for inspiration, but **all submitted work must be your own**.

Accommodations: The course is designed to be maximally equitable and inclusive of students with learning disabilities; there are no exams, and quizzes can be completed outside of class with no time limits. If you require additional accommodations, please place a request online through the Disabled Students Program (DSP) as early as possible.

Zoom etiquette: **Please disable your video and microphone at the start of class**, to prevent overloading other students' connections. You may ask questions in the chat or by unmuting your microphone. If possible, **please enable your video and/or microphone for breakout rooms and section**, to facilitate group communication.

Consent to record: Recordings of class will be made available to approved students only. These recordings can only be streamed, not downloaded. **By participating in class, you give consent for any video and/or audio captured from you to be included in the recording.** If you do not consent to being recorded, please leave your video and microphone off and only ask questions in the chat.

Grading

Letter grades will be assigned according to the following schedule (upper limits exclusive):

excellent	99–100%	A+	93–99%	A	90–93%	A–
good	87–90%	B+	83–87%	B	80–83%	B–
adequate	77–80%	C+	73–77%	C	70–73%	C–
barely passing	67–70%	D+	63–67%	D	60–63%	D–
not passing			<60%	F		

For undergraduate students taking the course P/NP, the minimum score for P is 73% (C). For graduate students taking the course S/U, the minimum score for S is 83% (B).

Extra credit, if offered, can be applied toward any grade *except* A+.

Assessment

This course contains regular opportunities to practice and reflect on the things you are learning, which allow us to identify areas we need to focus on more in class and section.

Assessment	**Weight per**	**Total weight**
assignments (4)	12.5%	50%
quizzes (10)	2%	20%
check-in surveys (10)	2%	20%
reflection essays (2)	5%	10%

Assignments. There will be 4 assignments, distributed on Wednesdays of weeks 2–8 and due 2 weeks later. **All assignments count toward the grade.** Each assignment will revolve around a different foundational idea in computational linguistics, and will give you direct opportunities to address the course learning goals with respect to that idea.

Quizzes. At the beginning of each week, a short multiple-choice and/or short-answer quiz will be distributed. You will gain partial credit simply for attempting this quiz prior to Wednesday's class. At the beginning of Wednesday's class, you will be randomly paired with someone else in the class, and you will have 10 minutes to discuss and justify your answers. You may change your answers as a result of this discussion; after the discussion, the remainder of the quiz credit will be awarded based on the correctness of your responses. **In week 1, a placement test will take the place of the quiz.**

Check-in surveys. At the end of each week, you will complete a short survey that asks you how the week has gone, what you've learned, and what you have outstanding questions or concerns about. Credit will be awarded for completing each survey on time.

Reflection essays. At the beginning of the quarter, you will write a short essay describing how you feel going into the course, what you expect from the course and hope to accomplish, and why. At the end of the quarter, you will write another essay reflecting on how your feelings have developed, how the course matched your expectations, and how well you accomplished what you'd hoped. Credit will be awarded for completing each essay on time.

SCHEDULE

In the first class, there will be a **placement test** that asks you to write a function that performs a simple task in Python. If you are unable to complete this placement test, you will struggle to complete assignments, and you are consequently advised to drop the course.

Assessment timing. Assignments will be released and due by 2pm every other Wednesday. Quizzes will be released after Monday's class and should be attempted before Wednesday's class. Check-in surveys are due by 5pm each Friday. Reflection essays are due by 5pm.

Week	Dates	Topic	Assessment out / due
1	M 1/04,	Orientation, **placement test (Quiz 1)**	
	W 1/06	Probability, notation	
	F 1/08	–	**Essay 1** due, **Survey 1** due
2	M 1/11,	N-gram language models	**Quiz 2** out
	W 1/13	**Quiz 2 discussion**, smoothing	**A1** out
	F 1/15	–	**Survey 2** due
3	*M 1/18*	*(No class: MLK Jr. Day)*	**Quiz 3** out
	W 1/20	**Quiz 3 discussion**, adv. smoothing	
	F 1/22	–	**Survey 3** due
4	M 1/25,	Distributional semantics	**Quiz 4** out
	W 1/27	**Quiz 4 discussion**, PPMI vectors	**A1** due, **A2** out
	F 1/29	–	**Survey 4** due
5	M 2/01,	Dense vectors	**Quiz 5** out
	W 2/03	**Quiz 5 discussion**, training data effects	
	F 2/05	–	**Survey 5** due
6	M 2/08,	Noisy channel models	**Quiz 6** out
	W 2/10	**Quiz 6 discussion**, autocorrect	**A2** due, **A3** out
	F 2/12	–	**Survey 6** due
7	*M 2/15*	*(No class: Presidents' Day)*	**Quiz 7** out
	W 2/17	**Quiz 7 discussion**, other noisy channels	
	F 2/19	–	**Survey 7** due
8	M 2/22,	Supervised learning, precision and recall	**Quiz 8** out
	W 2/24	**Quiz 8 discussion**, logistic regression	**A3** due, **A4** out
	F 2/26	–	**Survey 8** due
9	M 3/01,	Multinomial logistic regression	**Quiz 9** out
	W 3/03	**Quiz 9 discussion**, generative models	
	F 3/05	–	**Survey 9** due
10	M 3/08,	Unsupervised learning	**Quiz 10** out
	W 3/10	**Quiz 10 discussion**, Morfessor	**A4** due
	F 3/12	–	**Survey 10** due
	M 3/15	–	**Essay 2** due