LING 111/210: Advanced Computational Linguistics: Text Processing **Spring 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:** T 10-11am, W 1-2pm
	Classes:	MW 2:00-3:15pm on Zoom. Recordings will be made available on GauchoCast for students who are unable to attend synchronously.
	Sections:	On Zoom. *Undergrads:* T 2:00-2:50pm; *Grads:* W 12:00-12:50pm.
	Office hours:	At the times stated above, on Gather. See GauchoSpace for help.
	Forum:	You can get help through the GauchoSpace forum any time.
	Website:	All course information – including notices, assessment specifications, and weekly readings – is distributed through GauchoSpace.
Requirements	**Prerequisites:**	LING 110/297 in Winter 2021, or equivalent background.
	Textbook:	None; 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 modern techniques in computational linguistics and the ways they can be applied to linguistic research, and you will conduct your own computational linguistics research project.

Objectives

During this course, you will:

1. become familiar with modern approaches in computational linguistics;
2. develop a high-level understanding of the intuitions, algorithms, and math that are relevant to these approaches;
3. establish basic practical skills with industry-standard tools in version control, natural language processing, and deep learning;
4. develop strategies for engaging with scientific research, by reading, deconstructing, and critically evaluating computational linguistics papers;
5. gain hands-on experience in planning and conducting an original research project in computational linguistics.

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 / reviewing your notes	3	hours/week
quizzes	0.5	hours/week
other assessments / project-related work	5	hours/week

Note: guidelines are approximate; your hourly breakdown may differ.

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.

Attendance: This course is synchronous and will feature in-class activities. **You are expected to attend all classes and sections.** If time zones or other circumstances will prevent you from attending, please let us know as soon as possible.

Conduct Code: We follow UCSB's Student Conduct Code. You may discuss your work with other students, and consult information online for inspiration, but **all submitted work must be your own**. If you choose to complete your final project with a partner, you may submit joint work, provided it does not derive from external sources.

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 enrolled 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+.

This grading schedule may be changed for calibration purposes. **Any changes to the schedule can only improve your grade, not worsen it**.

Assessment

This course contains regular opportunities to learn new skills, practice the concepts you are learning in class, and build toward your final project.

Some assessments will be graded for **completion**: you will get all the points as long as you make a complete submission before the deadline. Other assessments will be graded for **quality**: your submission will be evaluated and you will get points based on this evaluation.

Assessment	**Weight**	**Total**	**How are they graded?**
Q quizzes (5)	3%	15%	completion (2% ea) + quality (1% ea)
PS paper summaries (3)	5%	15%	quality
TN training notebooks (2)	10%	20%	completion
PM project milestones (2)	10%	20%	completion
FP final project	30%	30%	quality

Quizzes. At the beginning of 5 different weeks of the quarter, a short multiple-choice quiz about the material of that week 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 your answers. You may change your answers as a result of this discussion, or based on further thought before the end of the week. At the end of the week, the remainder of the quiz credit will be awarded based on the number of questions you answer correctly.

Paper summaries. Three times during the quarter, you will be asked to submit a short summary of a paper that uses computational methods for linguistic research, following a provided template. One of these summaries will be about a paper we read as a class and discuss during class; one will be about a paper that you choose to read for your final project; and one may be about a paper from class or from your project. Each summary will be graded for quality, based on how well it demonstrates an understanding of the paper.

Training notebooks. In two sections during the quarter, a Jupyter notebook will be distributed. Each notebook will provide basic training in the use of an industry-standard tool in natural language processing or deep learning. You will receive full credit for a notebook as long as you complete and submit it by the end of the week after it is distributed.

Project milestones. In Weeks 4 and 8 of the quarter, you will submit a written milestone that ensures your project development is on track, and gives you an opportunity to seek out feedback from the instructor and/or TA. During the week a milestone is due, we will devote class and section time to helping you complete it. The first milestone will require you to define a topic and form of your project, and to develop a list of resources that you may use for it. The second milestone will require you to develop a full, actionable plan for the project, together with an indication of how it fits into a bigger picture of the field. You will receive full credit for a milestone as long as you submit it on time.

Final project. The course will culminate in your submission of a final project that focuses on using computational methods for linguistic research. The project may be written as a short academic paper or a scientific blog post, and may take any of the following forms:

(a) a comprehensive literature review that is targeted toward an original research project;

(b) the development and documentation of a new dataset, together with a plan for how it could be used in an original research project;

(c) a small-scale original research project (with a smaller literature review than in (a), using an existing dataset);

(d) an educational blog post that breaks down a computational linguistic method in an accessible manner and gives a concrete demonstration of how it could be used for linguistic research.

You will be expected to submit supplementary files containing code and data that underpin the project. The project will be graded for quality, based on a holistic evaluation that takes into account both the manner of presentation (academic paper or blog post) and the focus of the project (review, data, research, or methods education).

SCHEDULE

Monday's classes will generally be lectures. The primary exception is Weeks 4 and 8, which will be individual meetings (on Gather) to help you meet project milestones. **Wednesday's classes will generally be discussion**, of quizzes (if applicable) and papers. Paper discussions will be oriented toward the creation of paper summaries.

Assessment release. Quizzes will be released after Monday's class and training notebooks will be released in section; all other assessments will be available throughout the course.

Assessment due dates. All assessment due dates fall on a **Friday at 5pm**.

Week	Dates	Topic	Assessment out / due
1	M 3/29, W 3/31 *F 4/02*	Orientation, (pre-)processing with NLP Reading papers, equations, and figures –	
2	M 4/05, W 4/07 *F 4/09*	Introduction to neural networks **Q1 discussion**, paper discussion –	**Q1** out **TN1** out **Q1** due
3	M 4/12, W 4/14 *F 4/16*	Learning in neural networks **Q2 discussion**, paper discussion –	**Q2** out **Q2** due, **TN1** due
4	*M 4/19* W 4/21 *F 4/23*	*Project meetings: milestone 1* Paper discussion –	 **PS1** due, **PM1** due
5	M 4/26, W 4/28 *F 4/30*	Hidden Markov Models **Q3 discussion**, paper discussion –	**Q3** out **Q3** due
6	M 5/03, W 5/05 *F 5/07*	Recurrent Neural Networks **Q4 discussion**, paper discussion –	**Q4** out **TN2** out **Q4** due, **PS2** due
7	M 5/10, W 5/12 *F 5/14*	Memory and attention in RNNs **Q5 discussion**, paper discussion –	**Q5** out **Q5** due, **TN2 due**
8	*M 5/17* W 5/19 *F 5/21*	*Project meetings: milestone 2* Paper discussion –	 **PS3** due, **PM2** due
9	M 5/24, W 5/26 *F 6/28*	Transfer learning Paper discussion –	
10	*M 5/31* *W 6/02* *F 6/04* *F 6/11*	*(No class: Memorial Day)* *Project help office hours* – –	 **FP** due