

CS 490A: Applications of Natural Language Processing

Fall 2022

Instructor: Laure Thompson, laurejt@cs.umass.edu

Course Staff:

- Pengshan Cai, pengshancai@umass.edu
- Junda Wang, jundawang@umass.edu
- Chloe Eggleston, ceggleston@umass.edu
- Mudit Chaudhary, mchaudhary@umass.edu
- Alok Kumar, alokkumar@umass.edu

Lecture: TuTh 4:00-5:15pm in Goessmann Laboratory Room 20

Office Hours: See [Moodle](#) for details.

Course Description. This course will introduce Natural Language Processing (NLP) methods and applications, such as text classification, sentiment analysis, machine translation, and other applications to identify and use the meaning of text. During the course, students will (1) learn fundamental methods and algorithms for NLP; (2) become familiar with key facts about human language that motivate them, and help practitioners know what problems are possible to solve; and (3) complete a series of hands-on projects to use, implement, experiment with, and improve NLP tools. 3 credits.

Course Prerequisites. COMPSCI 220 and COMPSCI 240. An alternate prerequisite of LINGUIST 492B is acceptable for Linguistics majors.

Course Platforms. We will use

- **Moodle** to (1) host all course content including readings, in-class notes, and assignment details; and to (2) post course announcements
- **Echo360** for lecture recordings
- **Gradescope** for collecting and tracking grades
- **Piazza** for Q&A. Please use private posts for general communication with course staff. For private/sensitive issues, email Prof. Thompson directly.
- **Zoom** for office hours.

Links for each of these will be shared via [Moodle](#).

Course Texts. For this course, we will primarily use [Dan Jurafsky and James Martin, Speech and Language Processing \(3rd ed. draft\)](#). All readings for this course will be posted to Moodle. Required readings should be read before class.

Below are several textbooks that might be useful to consult.

- Bender, [Linguistic Fundamentals for NLP](#). Focuses on linguistic issues relevant to NLP.
- Bird et al., [NLP with Python](#). Aimed at a more introductory level than this course, but the book is a good gentle introduction to NLP with a Computational Linguistics emphasis.

- Einstein, [NLP Notes](#).
- Goldberg, [Neural Networks Methods for NLP](#).
- Manning and Schütze, [Foundations of Statistical NLP](#).
- Murphy, [Machine Learning: A Probabilistic Perspective](#). Excellent, though advanced, coverage of most machine learning methods we will use.
- Smith, [Linguistic Structure Prediction](#). Focuses on structured prediction inference methods for NLP.

Assignments & Grading.

Grading Scale. F = <63.50, D = 64-66, D+ = 67-69; C- = 70-73, C = 74-76, C+ = 77-79; B- = 80-83, B = 84-86, B+ = 87-89; A- = 90-93, A = >93.49

Final grades will have a granularity of two decimal places and will be rounded to reach a letter grade for the D through A-range.

Assignment Categories & Weights.

- Exercises – 10%
- Homework – 25%
- Midterm – 25%
- Annotation Project – 25%
- Survey Paper – 15%

Exercises.

Exercises will be assigned approximately weekly. Typically, these exercises will be (or start as) in-class activities. All exercise materials will be made available electronically. They will be graded on a scale of ✖, ✓-, ✓.

Homework.

Homework assignments will contain a mixture of written and programming problems. These assignments must be completed independently. Typically, homework will be released Thursdays after class and be due the following Wednesday at 11:59pm.

Midterm.

The midterm will be in-class and is scheduled for Thursday, October 27th.

Annotation Project.

In this project you will gain exposure to the entire NLP design process for creating a new annotation-based NLP task. In groups, you will (1) design a new document-level labeling task that requires *human judgment*, (2) create annotated data and annotation guidelines for this task, and (3) build predictive models using your annotated data.

Survey Paper.

You will write a 2000-word survey paper on an NLP topic or subfield of your choice. For this assignment, you will need to read and synthesize the contents of published NLP

papers in order to describe the history and current state of the art for your selected topic/subfield. Survey papers are due on the last day of finals, December 16th.

Policies.

Attendance.

There is no attendance requirement for this course. Lectures will be recorded, and in-class exercises can be made up electronically. **Do not come to class if you are feeling unwell.** Attendance is encouraged (if you are healthy and able) for educational experiences but there is no automatic grade deduction for missing class.

Communication.

Please use Piazza private posts to contact course staff. For private/sensitive issues, email Prof. Thompson directly. Please include “CS490A” in your email subject line.

COVID-19.

Everyone must comply with the official [university policies](#). Although not required, wearing masks in this classroom is strongly encouraged.

Collaboration.

All individual (non-group) assignments must be completed independently; **all submitted code and writing must be your own.** For group assignments, all submitted work must be produced solely by the members of your group. If you use or build off external material in your writing or code, you must properly cite their sources. See this [infographic](#) for more on written plagiarism.

You are encouraged to have high level discussion of assignments with your classmates. However, these collaborative discussions *must* be disclosed; list the names of all discussants on your homework submission. Do not share code or written materials.

Here are some example scenarios of the collaboration policy:

- Acceptable: Alice and Bob discuss alternatives for storing large, sparse vectors feature counts, as required by a problem set.
- Unacceptable: Alice and Bob pair-program code for storing feature counts.
- Acceptable: Bob is confused about how to approach a coding problem and asks Alice for a conceptual description of her approach.
- Unacceptable: Alice and Bob divide a homework assignment into parts, and then share their solutions with each other to complete the assignment.
- Acceptable: Alice asks Bob if he encountered a failure condition at a “sanity check” in a coding problem, and Bob explains at a conceptual level how he overcame that failure condition.
- Unacceptable: Alice or Bob obtain a solution to a previous year’s assignment or to a related assignment from another class and use it as the starting point for their own solution.

Late Work Policy.

Listed below are the blanket late policies for this course. All deadlines have a 2-hour grace period. If you should need an extension, please contact Prof. Thompson in advance to discuss your situation. In case of unforeseen circumstances (e.g., illness, internet issues, personal emergency), contact Prof. Thompson as soon as possible.

- *Exercises.* Submissions will be accepted up to a week late, but for reduced credit. Late submissions of ✓ quality will receive a grade of ✓-.
- *Homework.* You will have **five** slip days for homework assignments. A slip day provides an extra 24 hours to submit an assignment. At most two slip days may be used for a single assignment (for a 48-hour extension). Slip days are applied immediately and cannot be changed retroactively. Once slip days are exhausted, homework may be submitted up to 48 hours late for 50% credit. Homework will not be accepted beyond 48 hours late.
- *Annotation Project.* Beyond the grade period, project deliverables will not be accepted late.
- *Survey Paper.* Beyond the grace period, late submissions will not be accepted.

Inclusion.

In this course, each voice has something of value to contribute. Please take care to respect the different experiences, beliefs, and values expressed by students and staff involved in this course. We support UMass Amherst's commitment to diversity, and welcome individuals of all ages, backgrounds, citizenships, disability, sex, education, ethnicities, family statuses, genders, gender identities, geographical locations, languages, military experience, political views, races, religions, sexual orientations, socioeconomic statuses, and work experiences.

Accommodation Statement.

The University of Massachusetts Amherst is committed to providing an equal educational opportunity for all students. If you have a documented physical, psychological, or learning disability on file with Disability Services (DS), you may be eligible for reasonable academic accommodations to help you succeed in this course. If you have a documented disability that requires an accommodation, please notify me within the first two weeks of the semester so that we may make appropriate arrangements. For further information, please visit Disability Services (<https://www.umass.edu/disability/>)

Academic Honesty Statement.

Since the integrity of the academic enterprise of any institution of higher education requires honesty in scholarship and research, academic honesty is required of all students at the University of Massachusetts Amherst. Academic dishonesty is prohibited in all programs of the University. Academic dishonesty includes but is not limited to: cheating, fabrication, plagiarism, and facilitating dishonesty. Appropriate sanctions may be imposed on any student who has committed an act of academic dishonesty. Instructors should take reasonable steps to address academic misconduct. Any person who has reason to believe that a student has committed academic dishonesty should bring such information to the attention of the appropriate course instructor as soon as

possible. Instances of academic dishonesty not related to a specific course should be brought to the attention of the appropriate department Head or Chair. Since students are expected to be familiar with this policy and the commonly accepted standards of academic integrity, ignorance of such standards is not normally sufficient evidence of lack of intent (https://www.umass.edu/dean_students/codeofconduct/acadhonesty/).

I reserve the right to modify this syllabus as needed to account for current events and to better support student learning.