Overview

Computational linguistics is broadly the study of how to enable computers to analyze language produced by people or even generate such language themselves. While the idea of machines that use language is quite old, today this technology is all around us, in our phone assistants, search engines, social media recommendations, YouTube captions, and more. In this class, we focus on core concepts and problems in computational linguistics, centered on how text can be represented, predicted, categorized, dismantled, and interpreted. We take a hands-on approach to exploring parts of this broad field and seeing how we can make computers work with language.

Prerequisites

Two 200-level CS courses. Recommended: CMPU 203 and CMPU 240.

Course goals

By the end of the semester, you should be able to:

- Implement standard algorithms for natural language processing, such as text normalization, statistical language models, vector representations of words, word sense disambiguation, part-of-speech tagging, and text classification,

- Read and analyze primary literature (that is, academic papers) in natural language processing (NLP) and related fields,

- Use and interpret appropriate metrics of evaluation of how NLP technologies perform in practice,)
• Critique the assumptions and approximations that are made about language when developing problems, algorithms, datasets, and evaluations, and

• Practice the art of writing and presenting research.

Schedule

The current course schedule can be found on the course website. Readings, assignments, and other information will be added throughout the semester.

Readings and worksheets

The textbook we will use is:

_Speech and Language Processing_, third edition (online draft)
Daniel Jurafsky and James H. Martin

We will also read academic papers and other relevant materials, which will be posted on the course website.

Reading discussions will happen on Tuesdays, so you will be expected to complete weekly readings before class on Tuesday.

In class, you’ll get a chance both to ask about the shared technical reading (usually from Jurafsky and Martin) and to discuss in small groups a modern reading on a subtopic you sign up for. Readings – and sign-ups for subtopics – will be posted the Friday prior to a Tuesday class.

When you complete the reading, you’ll fill out a short worksheet on Gradescope with questions about comfort with the concepts from the required reading and a check-in about how things are going. Worksheets are expected to be completed by the start of class on Tuesday. Worksheets are graded on completion to assess whether you’ve done the reading and thought about the content.

Additionally, during Tuesday class, you’ll have the opportunity to write some notes down about what you learned in discussion to turn in; these will also be graded based on completion as part of your class participation grade.
Assignments

The first half of this course features weekly assignments. These assignments will be posted just prior to class on Thursdays. Unless otherwise specified, assignments are due the following Wednesday at 10:00 p.m. on Gradescope.

These assignments account for the largest component of your course grade, and many of them involve substantial programming and analysis. You are strongly encouraged to start early. Do not underestimate how long it may take you to complete the assignments!

Even if you don’t fully complete an assignment, you should submit what you’ve done for partial credit.

Special topic presentation

In the second half of the semester, you will work as a small team of 3–4 students to lead a 35-minute class discussion of a topic in modern natural language processing, in which you will present the core problem and some ideas from recent work. Information about signing up for topics and expectations for these will be shared when we reach that point.

Final project

In the second half of the semester, you will also work on a final project in a group of 1–3 people. The final deliverables of this project will be (1) a 3–4 page ACL-style paper describing your project and (2) a 15-minute presentation of the highlights of your project. Your final project grade will be determined by these two deliverables, as well as several smaller milestone assignments in the second half of the semester.

Coursework and grades

The elements of the course will be weighted approximately as follows:
First half:
- Worksheets 10%
- Assignments 40%

Second half:
- Special topic presentation 10%
- Final project 40%

Remember that success in the course is more than just good grades. It means that you are being challenged to grow as a learner, that you are engaging actively with tasks that feed your growth, and that you are creating excellent work by completing challenging tasks with an appropriate level of support. It also means that you are building your lifelong learning skills so that once the course is over, you are better and stronger as a learner and can continue to learn new things independently.

Extensions

If you need a one-day extension in this class for any reason (lots of exams, busy with a family obligation, or just tired), please email me to let me know prior to the time the assignment is due. Any request for a one-day extension on something besides an exam or the final project submission will be honored, no matter the reason. If you’re working in a group, only one group member needs to request the late day.

If something comes up that will prevent you from completing classwork for multiple days (e.g., an illness or family emergency), please reach out to me as soon as you can so we can together determine a healthy plan for you to continue your coursework when it makes sense. My priority is to make sure you can focus on the situation at hand while ensuring you won’t spend the rest of the semester working to catch up again.

Academic integrity

Please read the CS department’s guide to academic integrity:

[cs.vassar.edu/integrity](http://cs.vassar.edu/integrity).

In particular, note that:

1. You may not copy code written by anyone else (e.g., a classmate, a friend, an online source, a book).
Using code or other material from sources as “inspiration” and submitting highly derivative solutions is considered copying. (That is, you can’t “paraphrase” someone else’s work!)

You may not post a public question that contains any part of your code.

You may consult online resources as part of your course work, but you may not copy code from online source. If you get an idea of how to solve a problem from an online source, include a citation near the top of your source code.

If you haven’t already done so, you should read “Going to the Source”, available from the Dean of the College website. Note that the guidelines that apply to writing in general apply equally to the writing of computer programs. Copying someone else’s code without attribution is plagiarism. Give proper attribution for the help you receive.

Quoting from Chapter X, “In suspected cases of plagiarism, the instructor prepares a written statement of complaint to the Academic Panel.” Please don’t put yourself or your professor in that position. When in doubt, stop and ask me first.

Communication and getting help

I expect all students to participate in class discussions. In return, I will make myself available to answer questions, listen to concerns, and talk to any student about topics related to the class. I welcome your feedback throughout the semester about how the course is going.

Our course uses Slack for online discussions. You can use it to ask questions about course concepts, assignments, and logistics. It is also used for discussions among students, including for any work to be done in groups.

Because Slack is an extension of our classroom discussion, we expect everyone to behave accordingly: No disrespect, rudeness, or abuse will be tolerated.

Instructor availability

Office hours will be listed on the instructor’s website. If you need to meet at a different time, send the instructor a message.

Please come to office hours! Coming to office hours does not send a signal that you are behind or need “extra help”; on the con-
trary, the most successful students are usually those who come to office hours early and often.

I typically only check email and other messages between 8 a.m. and 8 p.m. on weekdays. If you send a message that needs a response during those times, you can expect to get a response within six hours (often much sooner). Otherwise, you can expect one when I’m back online.

Academic accommodations

Academic accommodations are available for students registered with the Office for Accessibility and Educational Opportunity (AEO). Students in need of disability (ADA/504) accommodations should schedule an appointment with me early in the semester to discuss any accommodations for this course that have been approved by the Office for Accessibility and Educational Opportunity, as indicated in your AEO accommodation letter.

Course conduct

Natural language processing as a field is prone to touch on topics related to culture and identity. The unique experiences you bring to the class from your own experiences not only strengthen our community but also actively contribute to the learning of everyone in the classroom. On the flip side, false cultural assumptions and negative comments about others both causes harm and makes our collective work as conscientious NLP scholars harder.

As your instructor, I am committed to creating a classroom environment that welcomes all students, regardless of race, gender, social class, religious beliefs, etc. We all have implicit biases, and I will try to continually examine my judgments, words and actions to keep my biases in check and treat everyone fairly. I expect that you will do the same with respect to me and the other members of the class, and that you will let me know if there is anything I can do to make sure everyone is encouraged to succeed in this class.

Title IX

Vassar College is committed to providing a safe learning environment for all students that is free of all forms of discrimination and sexual harassment, including sexual assault, relationship abuse, and
stalking. If you (or someone you know) has experienced or experiences any of these incidents, know that you are not alone. Vassar College has staff members trained to support you in navigating campus life, accessing health and counseling services, providing academic and housing accommodations, helping with legal protective orders, and more.

Please be aware all Vassar faculty members are “responsible employees,” which means that if you tell me about a situation involving sexual harassment, sexual assault, relationship abuse, or stalking, I must share that information with the Title IX Coordinator. Although I have to make that notification, the Title IX office will only provide outreach by email. You will control how your case will be handled – you don’t have to read or respond to the email, and it is completely up to you whether to pursue a formal complaint. Our goal is to make sure you are aware of the range of options available to you and have access to the resources you need.

If you wish to speak to someone privately, you can contact any of the following on-campus resources:

- Counseling Service (counselingservice.vassar.edu, 845-437-5700)
- Health Service (healthservice.vassar.edu, 845-437-5800)
- Rachel Gellert, Director of Support Advocacy and Violence Prevention (savp.vassar.edu, 845-437-7863)
- SART (Sexual Assault Response Team) advocate, available 24/7 by calling the CRC at 845-437-7333 and asking for SART

The SAVP website (savp.vassar.edu) and the Title IX section of the EOAA website (eoaa.vassar.edu/title-ix) have more information, as well as links to both on- and off-campus resources.

Acknowledgments

This course – and this syllabus – is based in large part on work by other professors, especially Xanda Schofield of Harvey Mudd College. The policy on asking and answering questions and on communication on Slack is adapted from Laney Strange, Northeastern University.