Overview

Computational linguistics addresses questions at the intersection of human languages and computer science: How can computers acquire, comprehend, and produce natural languages, such as English? How can computational methods give us insight into observed human language phenomena? Because language is one of the most complex human abilities, building computational technologies that involve language is both challenging and important.

Course goals

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

- Comprehend the general formalisms underlying the language technologies you interact with,
- Understand the basis for many research avenues in natural language processing (NLP) and machine learning,
- Analyze ethical concerns about language technology,
- Appreciate the complexity of human language from a computational perspective,
- Apply theory to practice, implementing working NLP programs, and
• Innovate to improve the performance of a task or harness language-based data in creative new ways.

Prerequisites

Two 200-level Computer Science courses.

Recommended:

**CMPU 240: Theory of Computation**

Course format and expectations

The calendar with lecture notes, assigned readings, and exercises is on the course website and will be updated throughout the semester.

You are expected to attend and participate in class. However, you should not come to class if you are feeling sick; I’ll work with you to catch up after you are feeling better.

You are expected to seek help when needed. Sometimes students are embarrassed to seek help. I am happy to help with any course material, no matter when it was taught.

Readings

Our primary textbook is available for free online:

*Speech and Language Processing*, third edition (draft)

Dan Jurafsky and James Martin

Additional readings will be posted on the course website.

Coursework and grades

The elements of the course will be weighted approximately as follows:

- In-class exercises: 10%
- Assignments: 40%
- Special topics 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.

**Late work**

For flexibility, you have four “late days” that you can use to extend the deadline of any full homework assignment (not exercise) by 24 hours. You can use them all for one assignment or spread them out.

If you run out of late days, late assignments will incur a penalty of 10% for each day past the deadline. No late days can be used for the final project.

If you know you won’t be able to submit an assignment, talk to me. If you’re experiencing a major problem – medical, psychological, family, etc. – that is interfering with your ability to complete your class work, you should talk with the Dean of Studies, your class advisor, or Health Services, who will recommend appropriate accommodations to all of your professors, who genuinely want you to succeed.

**Collaboration and academic integrity**

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*, and you need to 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.

For more details of how this relates to programming, see the CS department’s guide:
Generative AI

Because this is an NLP class, we encourage you to use and interact with large language models (LLMs) like ChatGPT. However, if you use a LLM or other generative AI for help on classwork, you must acknowledge in writing what ideas/language/code are the LLM’s and what are yours. Be sure to note the specific LLM you used, the prompt(s) you gave, and whether you modified or edited the response.

Example: “I used ChatGPT with the prompt ‘What is the ethics of using a large language model?’ Then I edited the output as the first paragraph of my response. The second paragraph of the response is my own thoughts in response to the first paragraph.”

If the tool allows it – and most do! – you should submit a full transcript of your interactions along with your finished work.

Failure to properly cite your use of generative AI is a violation of academic integrity.

Collaboration

You may not share code from homework assignments with other students at any point in the course – even after the assignment deadline.

You can discuss assignments with other students in the class, but, if you do so, wait at least 30 minutes before writing your solution. This helps you know if you actually understand the solution.

Reference

You may consult external reference resources for general concepts and techniques, but you must cite them. You must not consult solutions to this or any similar assignment from other courses, books, or online resources.
Code reuse and adaptation

You must not copy, reuse, or adapt any code beyond what is given to you in the assignment and in class. 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!)

After the semester

Do not post homework solutions publicly (e.g., on GitHub or your website). You may show your work to potential employers or others who aren’t students.

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 Ed for announcements and online discussions. You can use it to ask questions about course concepts, assignments, and logistics. If you go 30 minutes working on an assignment without making progress, please ask for help! You can post questions on Ed, come to my office hours, or schedule a meeting.

Posts on Ed can be either public or private. Public posts are visible to everyone, while private posts are visible only to course staff. Any questions having to do with your particular solution to an assignment should be private; all other posts should be public. (If you have a question about something, it’s very likely that other students do, too!)

Because Ed 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 department website.

Ed is a discussion forum, so please feel free to respond to questions and comments — it’s great when students can learn from each other!
I typically only check email and other messages between 9 a.m. and 5 p.m. on weekdays. If you send a message that needs a response during those times, you can expect to get a response the same day. 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 experi-
ences 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 Nancy Ide (Vassar College) and Carolyn Anderson (Wellesley College). The course goals are adapted from Sravana Reddy (Wellesley College). The policy on generative AI is adapted from Katie Keith (Williams College). The policy on asking and answering questions and on communication on Ed is adapted from Laney Strange (Northeastern University). The statement on course conduct is adapted from Xanda Schofield (Harvey Mudd College).