Computational Linguistics

Spring 2021

Tuesday & Thursday, 3:10–4:25 p.m.
Sanders Physics 309 and Zoom
Prof. Jonathan Gordon

cs.vassar.edu/~cs366

Overview

Computational linguistics addresses the fundamental 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?

In this course, you will learn how computers can do useful things with human languages, such as translate from English into Japanese, summarize a magazine article into a few sentences, find the main topics in the day’s news, and analyze prevailing opinions about a given topic on Twitter. You will also learn about how computational methods can help linguists explain language phenomena, including automatic discovery of different word senses and phrase structure.

Over the past decade, computational linguistics has been revolutionized by statistical and probabilistic methods which have been applied to analyze large bodies of language data (such as the web). You will learn the basics of probabilistic analysis and how to use them in machine learning.

Prerequisites

Any two 200-level Computer Science courses. CMPU 240 Theory of Computation is strongly recommended.

Course goals

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

• Comprehend the general formalisms underlying the language technologies you interact with,

• Be able to understand the basis for many research avenues in natural language processing (NLP) and machine learning,

• Appreciate the complexity of human language from a computational perspective, and

• Apply theory to practice, implementing working NLP programs.
Course format

This course uses a synchronous hybrid format: The course meetings will simultaneously take place face-to-face with social distancing and live-streamed on Zoom.

The class sessions are intended to augment and support the material in your readings by providing an opportunity for in-depth discussion, working through examples, and the chance for you to ask questions.

Meetings are recorded for asynchronous viewing, and you can also ask questions and discuss the material online. You are free to opt out of the face-to-face meeting and participate online, synchronously or asynchronously, at any time.

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

Readings

We will primarily read two books available for free online:

*Speech and Language Processing*, third edition (draft)
Dan Jurafsky and James Martin

*Natural Language Processing with Python*, second (online) edition
Steven Bird, Ewan Klein, and Edward Loper

Additional readings will be posted on the course website.

Software

In this course we will use Python, together with the Natural Language Toolkit (nltk). It is assumed that you are either familiar with Python already, or, due to your excellent preparation in the Vassar Computer Science program, you have the ability to learn it quickly. You can use the computers in the Asprey Lab in the Computer Science Department, which have all of the required software already installed. If you want to use your own computer, you should install the following:

- **nltk**: http://nltk.org/install
- **nltk data**: http://nltk.org/data

Python 3 is already installed on most macOS and Linux systems. If it’s not installed on your computer, it’s available from python.org
Coursework and grades

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

- Exercises 10%
- Assignments 40% (approx. four)
- Final project & presentation 50%

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 three “late days” that you can use to extend the deadline of any homework assignment by 24 hours. You may use up to two late days per homework assignment. If you plan to use a late day, please email the instructor before the original deadline.

If you run out of late days, late assignments will incur a penalty of 20% 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.

Academic integrity

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

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

2. 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!)

3. You may not post a public question that contains any part of your code.
You may consult online resources as part of your coursework, 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 Campuswire 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.

The quickest way to get feedback and help will be via Campuswire. We expect you to read Campuswire every day or two for announcements and clarifications to assignments; you are responsible for all clarifications made at least 48 hours before an assignment is due.

Posts on Campuswire 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 Campuswire 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. All student meetings this semester will be conducted online.

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.

Pandemic policies

Before coming to class, perform a self-evaluation and, if you feel even the slightest bit of sickness, stay home and participate remotely. You should be physically present in class only if you feel as healthy as normal.

Please know that at any time this semester, if you feel unable or unwilling to be present for a face-to-face meeting, you can opt out and participate remotely – no questions asked, no permission needed, no penalty incurred. Remember that face-to-face meetings are live-streamed, and you can participate in the class remotely through Campuswire, so you will not fall behind if you have to participate in class from home.

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.

Diversity and inclusion

In an ideal world, science would be objective. However, much of science is subjective and is historically built on a small subset of privileged voices. There may be both overt and covert biases in the material due to the lens with which it was written, even though the material is primarily of a scientific nature. Since integrating a diverse set of experiences is important for a more comprehensive understanding of science, please contact me if you have any suggestions to improve the quality of the course materials.

I would like to create a learning environment that supports diversity of thoughts, perspectives, and experiences, and honors your identities.

• If you have a name and/or set of pronouns that differ from those that appear in your official records, please let me know!

• If you feel that your performance in the class is being impacted by your experiences outside of class, please don’t hesitate to contact me. If you prefer to speak with someone outside of the course, you can contact your class advisor or the Dean of Studies.
• I (like many people) am still in the process of learning about diverse perspectives and identities. If something was said in class (by anyone) that made you feel uncomfortable, please talk to me about it.

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)

• Nicole Wong, SAVP (Sexual Assault and Violence Prevention) director (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. The course goals are adapted from Sravana Reddy, Wellesley College. The policy on asking and answering questions and on communication on Campuswire is adapted from Laney Strange, Northeastern University. The statement on diversity and inclusion is adapted from Monica Linden, Brown University.