

Introduction to Computational Linguistics Ling 581 Spring 2021

Jean Mark Gawron
San Diego State University
gawron@sdsu.edu
<https://gawron.sdsu.edu/compling>

2021 Jan 21

1 Introduction

This course will serve as an introduction to the field of computational linguistics, which includes aspects of speech recognition, natural language processing, information retrieval, and information extraction.

The course begins with an introduction to finite-state automata and some basic natural language applications; this is extended to finite-state transducers with applications in morphology (word structure). Other topics covered: ngram language models, classifiers (Naive Bayes and Logistic Regression), sentiment analysis, part of speech tagging, context-free grammars and context-free parsing (with statistical extensions), and distributional semantics.

2 Goals

The primary goal of the course is to acquaint students with a basic set of computational techniques that have proved useful in a variety of natural

language applications, with a particular emphasis on probabilistic methods. The principles and mathematics behind these techniques often overlap with those used in other fields in which machine learning has been successfully applied, such as computer vision. However, the problems, in particular the relevant statistical properties, are quite different. Thus, this class should provide a nice complement to other classes you may be taking which use machine learning.

Students should acquire enough facility with the concepts and tools so that they can use them to construct well-specified solutions to simple computational linguistic problems. A well-specified solution is one that a programmer can use to write a program.

3 Student Learning Outcomes

1. Students will be able to identify and use three different incarnations of Finite-State Natural Language processing methods, Regular Expressions, Finite-State Automata (as language recognizers and generators), and Finite-State Transducers (as morphological analyzers).
2. Students will be able to apply basic laws of probability to derive three different kinds of conditional probability language models, ngram models, HMM taggers, and Naive Bayes Classifiers (applied to sentiment analysis).
3. Students will learn the main probabilistic ideas embodied in Naive Bayes and Maximum Entropy Classifiers.
4. Students will be able to identify the major components in a CKY context-free parser, to sketch the proof that it is an n^3 algorithm, and to understand how an exponential number of parses can be stored in n^2 space.
5. They will also learn how to convert an all paths CKY parser into a statistical best-parse parser.
6. Students will be able to identify dynamic programming aspects of three major algorithms, Minimal Edit Distance, the Viterbi Best Path Algorithm (for HMMs), and the CKY context-free parsing algorithm.

7. Students will be able to explain the geometric justification for a variety of semantic similarity models for words.

4 Course Materials

The course will use the textbook:

Jurafsky, Daniel and Martin, James H. 2000. *Speech and Language Processing*. Prentice-Hall. (2nd Edition only!)

There will be exercises for most of the chapters covered. Readings will generally be assigned weekly as posted in the course outline.

5 Course Policies

Canvas Class Website	Your instructor will use Canvas to record and report your assignment grades. Course assignments will also be housed on Canvas and you will need access to Canvas to complete them. For detailed guidelines for using Canvas as a student, see the Canvas Student Guide. The course outline and course schedule, however, will be maintained on the class website .
Zoom and Class Meetings	<p>This class will feature synchronous class meetings each week at the class's regularly scheduled times as listed in the class schedule and at the beginning of this syllabus. These class meetings will be held on Zoom using Zoom link. [Use at class time!]</p> <p>the URL and room ID available on Canvas. All students are strongly encouraged to attend all Zoom sessions. That said, attending synchronous sessions may be difficult or impossible for some students. For this reason, attendance is not required. Similarly, any graded assignments offered during a synchronous class meeting will have asynchronous alternatives available via Canvas. All class meetings will be recorded, hosted online, and posted to Canvas for the benefit of students who could not attend. Because sessions are recorded, you are entitled to certain privacy considerations. First, you may change your Zoom name to a pseudonym. Please communicate this pseudonym to your instructor before attending a class meeting to ensure that you are admitted to the meeting. Second, you are not required to use a webcam or mic during class meetings, though you are certainly encouraged to do so.</p>

Zoom and
Office
Hours:

Your instructor will hold virtual office hours via Zoom using the link and meeting ID available on Canvas. You are strongly encouraged to visit your instructor during office hours if you need additional help. Do not wait until the last minute to get the help you need. Please come to office hours prepared with specific questions.

COVID-19
Considerations

Given how highly disruptive the ongoing pandemic can be, I want to acknowledge that I understand some of us may require special accommodations relating to illness or caring for loved ones. You do not need to give any personal information you are uncomfortable sharing, but if you require special arrangements for these reasons, let me know and I will do my best to assist you or to find someone on campus who can help.

6 Course Design

Reading Assignments:	This course follows the textbook closely. Much of what we discuss in class will mirror the content contained in the assigned readings. Because we will be moving fairly quickly through a number of complex topics this semester, I strongly suggest that you complete each assigned reading prior to the day it is due so that you are prepared for the discussions we have as a class. Readings will be posted on the course outline . on the course website.
Graded Homework Exercises	These assignments are available on Canvas and are meant to be completed after the class meeting addressing their chosen topic. The activities contained on these assignments typically mirror the content of the textbook's discussion exercises fairly closely or else are modeled after our in-class activities (those activities held during synchronous Zoom meetings).
Class Participation:	During many class meetings, we will be completing graded exercises. Students can submit these exercises to Canvas for credit once the meeting ends. Students unable to make the class meeting can also submit the exercises once they have had a chance to review the recorded lecture.
Final Exam:	The final exam will be a cumulative test covering topics from throughout the semester. The exam will be takehome, and you are encouraged to utilize all of our class texts (the textbook, slides, worksheets, etc.) while completing the exam.

7 Grading Policies

Grading will be based on assignments, quizzes, midterms and finals.

- Assignments: 40%
- Midterm 30%
- Final 30%

Complete your reading and homework assignments before class.

There are two reasons to complete the readings and homework assignments before each class:

1. Midterm/Final: Not only is homework worth 40% of your grade, but questions on your midterm and final will be modeled on your homework problems.
2. Homeworks will generally be due on Tuesday, and will be handed back on Tuesday of the next week. This will be the only opportunity to discuss homework solutions, and will include discussion of typical student issues and knowledge gaps. Students can hardly benefit from those discussions if they haven't done the assignment or have tried to do the assignment.

Many homeworks are graded loosely, especially at the beginning of the course, since they are regarded as learning experiences. But midterms and finals will be graded as evaluations of knowledge. Basically, the instructor will let many mistakes go on assignments; the instructor will deduct for mistakes on the midterms and finals.

Attendance and Participation: While attendance is not part of your grade, all students are strongly encouraged to attend all class meetings if able. This class covers a lot of complicated topics and moves fairly quickly. Attending our synchronous class meetings is an excellent way to interact with your instructor and your classmates and provides opportunities to ask questions in real time.

Late Work: All assignments in this class come with deadlines. Meeting these deadlines is important. That said, your instructor recognizes that some of you may be in different timezones, and keeping track of deadlines becomes much more complicated when time zone conversions are involved. For this reason, all students will be given a 24 hour grace period for assignment submission. Your assignment will not receive a late penalty so long as it is submitted within 24 hours of the original deadline. Once 24 hours have passed, your assignment will receive a 30% grade penalty. Once a week has passed from the original deadline, assignments may no longer be submitted for credit.

8 Pre-requisites

At least two linguistics courses or at least two programming or CS courses. Students with no programming background will find this course challenging.

9 Group Work

Group work is encouraged on the assignments. The midterm and final should be completed without any help. To be clear on this, collaboration or group work on the midterms and finals will be considered cheating.

When turning in collaborative assignments, your collaborators should be identified on your paper.

10 University Policies

- Accommodations: If you are a student with a disability and are in need of accommodations for this class, please contact Student Ability Success Center at (619) 594-6473 as soon as possible. Please know accommodations are not retroactive, and I cannot provide accommodations based upon disability until I have received an accommodation letter from the Student Ability Success Center.
- Student Privacy and Intellectual Property: The Family Educational Rights and Privacy Act (FERPA) mandates the protection of student information, including contact information, grades, and graded assignments. I will use Canvas to communicate with you, and I will not post grades or leave graded assignments in public places. If I would like to retain your assignment as an example to share with future classes, I will not do so without first obtaining your permission. Any assignments retained in this way will have your name and other personally-identifiable information removed. Students maintain intellectual property rights to work products they create as part of this course unless they are formally notified otherwise.
- Religious observances: According to the University Policy File, students should notify the instructors of affected courses of planned absences for religious observances by the end of the second week of classes.
- Student Email Addresses: Students are provided with an SDSU Gmail account for their official use. This SDSU email address will be used for all communications. Per university policy, students are responsible for checking their official university email once per day. For more information, see the Student Official Email Address Use Policy [here](#).

Academic
Honesty:

The University adheres to a strict policy prohibiting cheating and plagiarism. Examples of academic dishonesty include but are not limited to:

- copying, in part or in whole, from another’s test or other examination;
- obtaining copies of a test, an examination, or other course material without the permission of the instructor;
- collaborating with another or others in work to be presented without the permission of the instructor;
- falsifying records, laboratory work, or other course data;
- submitting work previously presented in another course, if contrary to the rules of the course;
- altering or interfering with grading procedures;
- assisting another student in any of the above;
- using sources verbatim or paraphrasing without giving proper attribution (this can include phrases, sentences, paragraphs and/or pages of work);
- copying and pasting work from an online or offline source directly and calling it your own;
- using information you find from an online or offline source without giving the author credit;
- replacing words or phrases from another source and inserting your own words or phrases.

The California State University system requires instructors to report all instances of academic misconduct to the Center for Student Rights and Responsibilities. Academic dishonesty will result in disciplinary review by the University and may lead to probation, suspension, or expulsion. Instructors may also, at their discretion, penalize student grades on any assignment or assessment discovered to have been produced in an academically dishonest manner.

Resources
for Students:

A complete list of all academic support services—including the Writing Center and Math Learning Center—is available on the Student Affairs’ Academic Success website. Counseling and Psychological Services (619-594-5220) offers confidential counseling services by licensed therapists; you can Live Chat with a counselor on this page between 4:00pm and 10:00pm, or call the San Diego Access and Crisis 24-hour Hotline at (888) 724-7240.

Sexual Violence/
Title IX mandated
reporting:

As an instructor, one of my responsibilities is to help create a safe learning environment on our campus. I am a mandated reporter in my role as an SDSU employee. It is my goal that you feel able to share information related to your life experiences in classroom discussions, in your written work, and in our one-on-one meetings. I will seek to keep the information you share private to the greatest extent possible. However, I am required to share information regarding sexual violence on SDSU's campus with the Title IX coordinator, Jessica Rentto (619-594-6017). She (or her designee) will contact you to let you know about accommodations and support services at SDSU and possibilities for holding accountable the person who harmed you. Know that you will not be forced to share information you do not wish to disclose and your level of involvement will be your choice. If you do not want the Title IX Officer notified, instead of disclosing this information to your instructor, you can speak confidentially with the following people on campus and in the community. They can connect you with support services and discuss options for pursuing a University or criminal investigation. Sexual Violence Victim Advocate (619-594-0210) or Counseling and Psychological Services (619-594-5220, psyc-serv@sdsu.edu). For more information regarding your university rights and options as a survivor of sexual misconduct or sexual violence, please visit titleix.sdsu.edu or sdsutalks.sdsu.edu.

Classroom
Conduct
Standards:

SDSU students are expected to abide by the terms of the Student Conduct Code in classrooms and other instructional settings. Prohibited conduct includes:

- Willful, material and substantial disruption or obstruction of a University-related activity, or any on-campus activity.
- Participating in an activity that substantially and materially disrupts the normal operations of the University, or infringes on the rights of members of the University community.
- Unauthorized recording, dissemination, or publication (including on websites or social media) of lectures or other course materials.
- Conduct that threatens or endangers the health or safety of any person within or related to the University community, including 1. physical abuse, threats, intimidation, or harassment. 2. sexual misconduct. Violation of these standards will result in referral to appropriate campus authorities.

Medical-related
Absences:

Students are instructed to contact their instructor in the event they need to miss class, etc. due to an illness, injury or emergency. All decisions about the impact of an absence, as well as any arrangements for making up work, rest with the instructor. Student Health Services (SHS) does not provide medical excuses for short-term absences due to illness or injury. When a medical-related absence persists beyond five days, SHS will work with students to provide appropriate documentation. When a student is hospitalized or has a serious, ongoing illness or injury, SHS will, at the student's request and with the student's consent, communicate with the student's instructors via the Vice President for Student Affairs and may communicate with the student's Assistant Dean and/or the Student Ability Success Center.

Student Economic
Crisis Response
Team:

If you or a friend are experiencing food or housing insecurity, technology concerns, or any unforeseen financial crisis, it is easy to get help! Visit sdsu.edu/ecrt for more information or to submit a request for assistance. SDSU's Economic Crisis Response Team (ECRT) aims to bridge the gap in resources for students experiencing immediate food, housing, or unforeseen financial crises that impacts student success. Using a holistic approach to well-being, ECRT supports students through crisis by leveraging a campus-wide collaboration that utilizes on and off-campus partnerships and provides direct referrals based on each student's unique circumstances. ECRT empowers students to identify and access long term, sustainable solutions in an effort to successfully graduate from SDSU. Within 24 to 72 hours of submitting a referral, students are contacted by the ECRT Coordinator and are quickly connected to the appropriate resources and services. For students who need assistance accessing technology for their classes, visit our ECRT website (sdsu.edu/ecrt) to be connected with the SDSU library's technology checkout program. The technology checkout program is available to both SDSU and Imperial Valley students.

11 Office Hours

Office [Zoom Office Hours Link](#)
Tu,Th 3:30-4:30
Wed 11:00-12:00
by appointment

12 Contact Information

Dept: Department of Linguistics and Oriental Languages
Uni: San Diego State University
Address: 5500 Campanile Drive
Email: gawron@sdsu.edu