LING 185A: Computational Linguistics I

Kevin Liang

Summer C, 2024

Email: kevinliang8@g.ucla.edu Pronouns: he/him/his
Lecture Hours: T/TR 2:00 -4:50PM
Section Hours: F 1pm-1:50pm
Office Hours: R 1pm-2pm (PDT), or by appointment

Zoom Link (Lecture, Section, Office Hours):
https://ucla.zoom.us/j/7197640935?pwd=ek1ESHFGZTRIK2taVE52ZG0wazR5QT09

Prerequisites: LING 120B
Program in Computing 10C or CS32

Course Description

Computational linguistics is a large, multifaceted and rapidly expanding field. There are many different kinds of work that might be classified as "computational linguistics", differing in

- goals (e.g., build a useful gadget, test a linguistic theory), and
- empirical domains (e.g., sounds, words, sentences),

but there are certain core analytical concepts, tools, and techniques that frequently appear throughout. Broadly speaking, the relevant foundational concepts concern the computational nature and properties of the kinds of systems that you have learned about in linguistics courses, namely grammars. This course aims to highlight these common ideas, starting with their simplest instantiations and gradually building up towards the more complex cases. The simple cases will probably closely parallel certain things you may have learned about in mathematics or computer science courses, and the more complex cases will hopefully look similar to things you have learned about in linguistics courses. An important goal is to highlight the connections between these areas. As a preview, some of the biggest ideas that will come up repeatedly are:

- recursive generation of infinitely many expressions by a finite machine,
- interchangeability/intersubstitutability of subexpressions within larger expressions, and
- an appreciation of the insights that can be gained by studying the human mind through the lens of formal/computational models

Course schedule

This tentative schedule is subject to change based on the progress of class discussions and student interests.

Week	*Date*	*Topic*	*comment*
1	8/6	Syllabus; *Haskell basics*; *math preliminaries*; *recursion*	
	8/8	*Introducing Finite-state automata (FSAs)*	Quizlet 1 due Friday
2	8/13	*More on FSAs: determinism, regular expressions*	HW 1 due Wednesday
	8/15	*Introducing context-free grammars (CFGs)*	Quizlet 2 due Friday
3	8/20	*Push-down automata*	HW2 due Wednesday
	8/22	*Stack-based parsing with CFGs*	Quizlet 3 due Friday
4	8/27	*Semirings*	HW3 due Wednesday
	8/29	*Semirings cont; Variants of formal grammars*	Quizlet 4 due Friday
5	9/3	*Subregular phonology*	HW 4 due Wednesday
	9/5	*Beyond CFGs: tree grammars*	Quizlet 5 due Friday
6	9/10	*Mildly context-sensitive grammars*	HW 5 due Wednesday
	9/12	*Spillover / Connections to NLP / Review*	Take-home exam due Friday

Course materials

Readings

There is no required textbook for this course. All materials (assignments, notes, readings) will be distributed through the course website on BruinLearn. But if you are interested in the topics covered in this class and want to dive deeper into them, here are some popular textbooks that would be fun and helpful:

- Hopcroft & Ullman (1979), *Introduction to Automata Theory, Languages and Computation*: chapters 2–5
- Sipser (2013), *Introduction to the Theory of Computation* (3rd ed.): chapters 1, 2
- Manning & Schütze (1999), *Foundations of Statistical Natural Language Processing*: chapters 6, 9–12
- Jurafsky & Martin (2009), *Speech and Language Processing* (2nd ed.): chapters 2, 4, 6, 12–14, 16
- Partee, Meuleun & Wall (1987), *Mathematical Methods in Linguistics*: chapters 16, 17, 18

Software

Many of the homework exercises will involve writing or modifying small programs. The programming language we will use is Haskell. You will need access to a computer with a text editor and

a Haskell compiler installed. There are many ways to get Haskell up and running on your machine, varying by your operating system. A good place to start is the Haskell Platform, available: https://www.haskell.org/downloads/.

On the role of Haskell and programming
Our use of Haskell in this course is a means to an end: it's a tool for helping to sharpen our understanding of certain concepts, and those concepts, rather than Haskell itself, are the real focus of the course. Completing this course will not necessarily make you a proficient Haskell programmer, although you'll be off to a good start.

Because of the secondary role of Haskell itself, I do not recommend consulting programming textbooks or tutorials that aim to teach you the language, at least until after the first few weeks, and my aim is to provide enough information for it not to be necessary at all. Haskell is a very "big" language and some of these other sources will begin with things that will not be relevant for us.

If and when you really want one, a good textbook for many purposes is Programming in Haskell by Graham Hutton (Cambridge University Press). The best way to use this would be as an occasional reference throughout the second half of the course, rather than as an initial learning aid. Another one is *Learn You A Haskell For Great Good!* by Miran Lipovača. It's free online (http://learnyouahaskell.com/), and contains way more than you'll need to know for this course.

Requirements and grading

Grade break-down

	Option 1	Option 2
Homework assignments	70%	56%
Quizlets	10%	10%
Final (take-home) exam	20 %	34%
In total	**100%**	**100**%

	+		-
A	$\geq$ 97%	93–96.99%	90–92.99%
B	87–89.99%	83–86.99%	80–82.99%
C	77–79.99%	73–76.99%	70–72.99%
D	67–69.99%	63–66.99%	60–62.99%
F		$\leq$ 59.99%	

Attendance

You will not be graded on lecture attendance, but it is crucial to your mastery of the course material and your success in the class. The lectures will not be recorded. Office hours will not be recorded. Labs on Friday have four main purposes:

- To go over the answers of past assignments and quizlets.
- General discussion of the course material.
- To allow you to ask specific questions about the current assignment.

- To prepare you for the next homework/quizlets.

It is strongly recommended that you attend these labs at the scheduled times as well.

LING 185A is a difficult course, and summer classes move very quickly. If you do not attend lectures and labs, you will fall behind.

Homework assignments (70% in total; 14% each)

Assignments will be posted to BruinLearn after class on Tuesdays, and will be due on BruinLearn **by next Wednesday (11:59 PM PDT)**. Please do not email me your assignments. Late assignments will be accepted up to two days late with a penalty of 10% for each late day used (i.e., assignments submitted 1 minute to 23 hours and 59 minutes late will be able to receive a maximum of 90% credit). **Assignments more than two days late will not receive course credit** (except in cases of legitimate disruptions, e.g., verified illness, jury duty, bereavement, religious observances). **If you fall behind in the course, it will be *very difficult* to catch up!** Homework grades will be posted on BruinLearn by the Wednesday of the following week.

Students are permitted (encouraged, even!) to collaborate on homework assignments. However:

- Each person must hand in their own assignment that is reflective of their own understanding (no direct copies or jointly authored assignments are allowed).
- You must list at the top of your assignment all of the people you've collaborated with.

Your discussions should abide by (both the letter and spirit of) the "whiteboard policy": you may work together on a whiteboard and discuss things for as long as you wish and in as much detail as you wish, but then you must erase the whiteboard and not take any written notes away from this discussion. The idea is that being able to write up your solution individually establishes that you understand what you submit.

Quizlets (10% in total; 2% each)

There will be five short take-home quizzes (quizlets) which will be uploaded to BruinLearn on Tuesdays after lecture and will be due on BruinLearn **by the end of Friday (11:59 PM PDT)**. Please do not email me your quizzes. The aim of these is to keep you thinking about the course material, check your understanding of something from the previous lecture, and sometimes to wrap up the week and prepare you for the new HW. **No late quizlets will be accepted.** Quiz grades will be posted on BruinLearn by the Monday of the following week.

Final take-home exam (20%)

During the summer, there is no designated finals week. There will be a cumulative take-home exam. The final exam will be posted to BruinLearn after class on Tuesday of the last week (9/10) and will be due **that Sunday 11:59 PM PDT (9/15)**. **No late exams will be accepted.**

Another grading option

You are allowed to drop one lowest hw grade, and then the hw will be 56% of the grade. For this option, the final will be 34% of your final grade. When I assign the final letter grade, I will automatically choose the option that yields a higher grade.

Getting help

- You can email me at the email address above. I will respond to emails within 24 hours.

Campus-wide policies

- All requests for academic accommodations based on qualifying disabilities or medical conditions must be made through the Center for Accessible Education. If you are already registered with the Center for Accessible Education (CAE), please request your Letter of Accommodation in the Student Portal. If you are seeking registration with the CAE, please submit your request for accommodations via the CAE website. Students with disabilities requiring academic accommodations should submit their request for accommodations as soon as possible, as it may take up to two weeks to review the request. For more information, please visit the CAE website (http://www.cae.ucla.edu), visit the CAE at A255 Murphy Hall, or contact CAE by phone at (310) 825-1501.

- Please read the academic dishonesty section of the UCLA Student Conduct Code:

  http://www.deanofstudents.ucla.edu/Portals/16/Documents/UCLACodeOfConduct_Rev030416.pdf

  and the Dean of Students' associated syllabus insert:

  http://www.deanofstudents.ucla.edu/Portals/16/Documents/Syllabus.pdf

  Incidents of suspected academic dishonesty (including copying someone else's work or allowing someone else to copy your work) will be referred to the Dean of Students.

- **Title IX** prohibits gender discrimination, including sexual harassment, domestic and dating violence, sexual assault, and stalking.

  - If you have experienced sexual harassment or sexual violence, you can receive confidential support and advocacy at the CARE Advocacy Office for Sexual and Gender-Based Violence, 1st Floor Wooden Center West, CAREadvocate@careprogram.ucla.edu, (310) 206-2465. In addition, Counseling and Psychological Services (CAPS) provides confidential counseling to all students and can be reached 24/7 at (310) 825-0768.
  - You can also non-confidentially report sexual violence or sexual harassment directly to the University's Title IX Coordinator, 2241 Murphy Hall, titleix@conet.ucla.edu, (310) 206-3417. Reports to law enforcement can be made to UCPD at (310) 825-1491. Faculty and TAs are required under the UC Policy on Sexual Violence and Sexual Harassment to inform the Title IX Coordinator should they become aware that you or any other student has experienced sexual violence or sexual harassment.

- You are encouraged to familiarize yourself with the full range of student resources and services listed here: https://www.studentincrisis.ucla.edu/Resources

Learning outcomes

- Knowledge outcomes:
  - Students should know what finite-state grammars and context-free grammars are, and know the limitations of each.
  - Students should understand the relationship between grammatical structure, recursion and dynamic programming.
- Skills outcomes:
  - Students should be able to read descriptions of grammatical systems in traditional mathematical notation and write corresponding programs using recursion and/or dynamic programming.
- Attitudes and values outcomes:
  - Students should come to appreciate the kind of understanding of the human mind that can come from trying to express its workings in a completely formalized system.
- Behavioral outcomes:
  - By combining the skills and knowledge outcomes above, students should be able to construct programmed implementations of simple grammatical models.

Acknowledgements

Much of the materials used in the lectures, assignments and quizlets in this course comes from previous iterations designed/taught by Dylan Bumford, Tim Hunter, Connor Mayer, Laurel Perkins, and Yang Wang.