LING 185A: Computational Linguistics I

Connor Mayer

Summer C, 2020

E-mail: connormayer@ucla.edu
Office Hours: M/W 5pm-6pm
Office Hours Zoom ID: 984 8063 5594

Class Hours: T/Th 10am-12:50 pm
Lecture Zoom ID: 918 4928 0692

Lab Hours: F 11-11:50 am
Lab Zoom ID: 974 8786 3727

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

Website: **https://ccle.ucla.edu/course/view/201C-LING185A-1**

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
- the relationship between discrete structures and probabilistic models.

Course schedule

This schedule may be adjusted based on the progress of class discussions and student interests.

Week	*Date*	*Topic*	*Comment*
1	8/4	*Python; recursion; Regular expressions (REs)*	
	8/6	*Introducing finite-state automata (FSAs)*	Quizlet 1 due
2	8/11	*More on FSAs: determinism, relation to REs*	HW1 due
	8/13	*Probabilistic FSAs; finite-state transducers*	Quizlet 2 due
3	8/18	*NLP: Part-of-speech tagging using PFSAs*	HW2 due
	8/20	*Introducing context-free grammars (CFGs)*	Quizlet 3 due
4	8/25	*Probabilistic CFGs*	HW3 due
	8/27	*Stack-based parsing with CFGs*	Quizlet 4 due
5	9/1	*Beyond CFGs: tree grammars*	HW4 due
	9/3	*Subregular phonology*	Quizlet 5 due
6	9/8	*Multiple context-free grammars*	HW5 due
	9/10	*Spillover / Connections to NLP / Review*	Take-home exam due

Course Materials

Readings

There is no required textbook for this course. All materials (assignments, notes, readings) will be distributed through the course website on CCLE.

There are two well-known textbooks that are often used for introductory computational linguistics courses: Jurafsky & Martin's *Speech and Language Processing* (Prentice Hall) and Manning & Schütze's *Foundations of Statistical Natural Language Processing* (MIT Press). Both take an approach that is more oriented towards specific applications than we will be taking in this course, but you may find them interesting and useful.

Hopcroft, Motwani & Ullman's *Automata Theory: Languages and Computation* is an excellent introduction to many of the formalisms we will explore in this course, and Partee, Meuleun & Wall's *Mathematical Methods in Linguistics* shows how these can be applied to linguistic problems.

For a general introduction to neural networks, I recommend Goodfellow, Bengio and Courville's *Deep Learning*. For specific applications to language, Goldberg's *Neural Netowrks for Natural Language Processing* is an excellent and accessible resource.

Software

Many of the homework exercises will involve writing or modifying small programs. The programming language we will use is Python 3. You will need access to a computer with a text editor and Python 3 installed (**https://www.python.org/downloads/**).

Requirements and grading

Recording of lectures and labs

You are *strongly* encouraged to attend lectures and labs at the scheduled times so you can ask questions, but the department has also requested that recording of lectures and labs be uploaded to CCLE to accommodate schedule conflicts that arise from remote learning. Office hours will not be recorded. Only the instructor will be able to record lectures, and lectures will only be accessible to members of the class via CCLE. If you want to opt out of being recorded, you are welcome to keep your camera and mic disabled throughout lecture and ask any questions via email or during office hours. For more information on UCLA's Zoom privacy policies, see **https://www.adminvc.ucla.edu/covid-19/academic-continuity/protecting-privacy-and-data-during-remote-working-and-using-zoom**.

Lectures will assume that students have read and largely understood any assigned readings. In many cases the assigned readings will treat a subject in more detail than could ever be done in class. For this reason, it is crucial that you complete assigned readings before the lecture.

Labs on Friday have three main purposes:

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

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 (75% of final grade)

Assignments will be posted to CCLE after lecture on Tuesdays, and will be due on CCLE by the beginning of class on the following Tuesday. Please do not email me your assignments. **Late assignments 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!

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 (5% of final grade)

There will be five short take-home quizzes (quizlets) which will be uploaded to CCLE on Tuesdays after lecture and will be due on CCLE by the beginning of lecture on Thursday. Please do not email me your assignments. 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 get the ball rolling for the new day's content. **No late quizlets will be accepted.**

Final take-home exam (20% of final grade)

There will be a cumulative take-home exam, which will constitute 20% of your final grade. The final exam will be posted to CCLE after class on Tuesday of the last week (9/8) and will be due at the beginning of the final lecture on that Thursday (9/10). **No late exams will be accepted.**

Optional extra SONA credit (1% bonus to final grade)

You have the opportunity to earn an extra 1% extra towards your final grade by participating in an online experiment through the Psychology Department Subject Pool. Serving as a subject in an experiment provides students with direct exposure to psychological research. By participating in experiments, you will have the opportunity to contribute to on-going research at UCLA while getting an inside glimpse of how research studies are conducted.

One credit is given for every hour of experiment participation. If you complete 1 hour(s) of experiment(s), you will have 1% added to your final grade at the end of the quarter.

The posting and scheduling of experiments is handled via the Psychology Department Subject Pool system at **http://ucla.sona-systems.com/**. More information on using the system can be found at **http://www.psych.ucla.edu/undergraduate/subject-pool-experiment-participation**.

NOTE: Before signing up for experiments, you MUST select the course for which you want your experiment credits to count. All experiments for Summer Session C 2020 must be completed by September 8th, 2020 (Tuesday of 6th Week).

As an alternative to participation in a SONA experiment, you can also read a research article and write a one-page summary to receive equivalent credit. I will provide a list of suggested articles, but you are free to choose your own as well so long as it is relevant to the material in this course. This will also be due on September 8th.

Getting help

- You should use the CCLE forum for discussions about the content of lectures or about homework assignments. Please don't reveal the answers to assignments!
- You can email me at the address above, with "185A" somewhere in the subject line for any logistical/administrative matters. I will respond to emails within 24 hours.

Campus-wide policies

- Students needing academic accommodations based on a disability should contact the Center for Accessible Education (CAE) at (310) 825-1501 or in person at Murphy Hall A255.

When possible, students should contact the CAE within the first week of the term as reasonable notice is needed to coordinate accommodations. For more information visit

http://www.cae.ucla.edu.

- 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.

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 material used in the lectures and assignments in this course comes from previous iterations designed by Tim Hunter and Ethan Poole.