Computational Linguistics 2

LING 4434 / CS 4745, Fall 2023
MW 1:25-2:40, (Uris 204) Morrill B07

Instructor: Marten van Schijndel
Email: mv443@cornell.edu
Office: 220 Morrill Hall
Office Hours: Tues 11:00-1:00
& By Appointment

Course Website: canvas.cornell.edu
Mailbox: 203 Morrill Hall

Course Description: An in-depth exploration of neural computational linguistic techniques. Whereas Computational Linguistics I (LING 4424) covers foundational techniques in symbolic computational modeling, this course covers the neural network methods widely used throughout computational linguistics and natural language processing as well as a number of techniques that can be used to probe the linguistic information and language processing strategies encoded in computational models. Through readings and discussion of a number of research papers from the literature, students will learn and use methods at the frontier of computational linguistics to explore what current computational models can represent about language and how they use that information.

The course will conclude with a large final project component, involving a final presentation, coding, data analysis, and a paper write up submitted after the end of classes on the date designated by the registrar.

Required Reading: We will read a number of research papers from the literature during this course. These are helpful places to learn more:
Published work: *ACL Anthology* `https://www.aclweb.org/anthology/`
Cutting edge unpublished work: *arXiv/cs.CL* `https://arxiv.org/list/cs.CL/recent`
Unpublished cognitive modeling work: *PsyArXiv* `https://psyarxiv.com/`
Programming tips: *StackOverflow* `https://stackoverflow.com/`

Homework: You should have completed Computational Linguistics 1 or Natural Language Processing before this and received a B+ or higher. Programming will be required. The homework assignments in this course will teach students to implement and use neural models and to conduct cutting edge neural analysis studies. Once you have developed familiarity with the models and methods from the literature, you will work in groups to conduct a replication of a paper from the literature. Avail yourself of office hours *before an assignment's due date* if you have trouble.

Final Project Scheduling and Deliverables: In contrast to the homework assignments, the final project will have students utilizing the models and methods discussed in the course to conduct a novel research study that will expand our knowledge of human language, human cognition, natural language processing tasks, and/or neural models. Students will prepare a final paper and presentation that puts into practice the core themes and arguments in this course. Several milestones of the assignment are due and students are expected to incorporate feedback in subsequent assignments. The final project should take at least 40 independent hours over the course of the semester and involves both a 20-minute presentation, a literature review, custom scripts to conduct

neural analysis, and a 5 page description of the project and its findings.

Oct 4: Meet with me before this date to discuss project plans

Oct 16: Submit a 2-page project proposal, which should summarize 3 related papers from the literature, describe a novel project, and relate your project to the prior papers.

Nov 1: Submit initial code pipeline and linguistic stimuli

Nov 15: Submit second code pipeline and stimuli

Nov 29: Give a 20-minute in-class presentation with slides describing the motivation, methodology, and results, followed by a question period

Final Project: Submit a 5-page paper in ACL format, describing your analyses and results. The analysis should include novel models or methods compared with previous work and should be novel in terms of the linguistic phenomena being studied. All code used to run the analyses must be submitted to canvas. The final project will be due after the end of classes at the time specified by the registrar.

Attendance/Participation: Lectures will be delivered for the first few weeks to provide background for the topics in the course. After the initial few weeks, we will discuss papers from the literature on neural network interpretability. You will be expected to read at least one of the core papers on each topic every week before Tuesday's class. Much class time will be devoted to discussion, working in groups, and answering questions on the lectures, so attendance is required because absence from in-class group work will hinder your classmates.

Grading: For each assignment, you must turn in any code along with your write-up. Each assignment builds to the next, so late work will hold up the rest. Therefore, I will accept assignments up to 48 hours after the due date with 10% off per extra 24-hour period (or partial 24-hour period). So if you submit an assignment 30 hours after the initial deadline, your grade for that assignment would be reduced by 20%.

Attendance	10%
Homework	3 x 20%
Final Project	30%

100–97	A+	89.9–87	B+	79.9–77	C+	69.9–67	D+
96.9–93	A	86.9–83	B	76.9–73	C	66.9–60	D
92.9–90	A-	82.9–80	B-	72.9–70	C-	Below 60	F

Special Accommodations: Please see Student Disability Services (SDS) about any accommodations you might need and they will convey that information to me. Email SDS at `sds_cu@cornell.edu`. SDS is located on level 5 of Cornell Health, 110 Ho Plaza, 607-254-4545, `https://sds.cornell.edu/`

Academic Integrity: Please don't cheat, and please cite people whose work you use. I am **required** to follow-up on suspected violations of Cornell's *Code of Academic Integrity*. I encourage students to watch this video to learn more about what constitutes cheating and why it matters: `https://cheatingvideo.provost.cornell.edu/`

If you have questions about issues of academic integrity, please see me.