

Natural Language Processing CAP 4641

Instructor: Bonnie J. Dorr (Other past/present instructors: Shlok Gilda, Yuanyuan Lei, Sangpil Youm)

Teaching Assistants: Contact through Canvas or Slack (include all TAs)

Office Hours: Zoom office hours w/ TAs (see Canvas). Both F2F and UOnline students use Slack and other interactive tools. Flipped F2F sessions provide opportunities for open interaction/QA time.

Course Description: This course examines how machines understand, process, and generate human language through computer science, linguistics, and AI. You'll learn core concepts and modern tools, embeddings, and transformers, with a focus on real-world applications. You'll learn how to build NLP systems and understand how and why they work, so you can critique them as well as create them.

Course Pre-Requisites / Co-Requisites: Prerequisite: COP 3530

Course Objectives

Students learn the theory and practice of Natural Language Processing. Among the theoretical topics included are: parts of speech, n-gram models, context-free grammars, parsing and ambiguity, dependency parsing, sequence labeling, hidden markov models, vector semantics and embeddings, and semantic role labeling. Practically, we are motivated to study natural language to better understand how to develop a range of applications, including machine translation systems, conversational agents, and techniques for evaluating NLP output. The course also covers the application of machine learning solutions to natural language data. **Python** will be used to illustrate concepts and implement project solutions. Students will be able to apply their knowledge to:

- implement and evaluate their own programmatic solutions to natural language problems
- apply symbolic and non-symbolic techniques to natural language problems
- evaluate the performance of natural language applications

Materials and Supply Fees There are no materials or supply fees for this course.

Relation to Program Outcomes (ABET):

Outcome	Coverage*
1. An ability to identify, formulate, and solve complex engineering problems by applying principles of engineering, science, and mathematics	High
2. An ability to apply engineering design to produce solutions that meet specified needs with consideration of public health, safety, and welfare, as well as global, cultural, social, environmental, and economic factors	High
3. An ability to communicate effectively with a range of audiences	
4. An ability to recognize ethical and professional responsibilities in engineering situations and make informed judgments, which must consider the impact of engineering solutions in global, economic, environmental, and societal contexts	Medium
5. An ability to function effectively on a team whose members together provide leadership, create a collaborative and inclusive environment, establish goals, plan tasks, and meet objectives	
6. An ability to develop and conduct appropriate experimentation, analyze and interpret data, and use engineering judgment to draw conclusions	
7. An ability to acquire and apply new knowledge as needed, using appropriate learning strategies	High

*Coverage is given as high, medium, or low. Unspecified coverage indicates this outcome is not assessed.

Required Textbooks and Software

All readings are online and freely available:

- The third edition of *Speech and Language Processing* is published here: <https://web.stanford.edu/~jurafsky/slp3/>
Note: One reading from slp2 on "Chunking" will be provided in Week 6 (see Canvas)
- The second edition of *Natural Language Processing with Python* is published here: <http://www.nltk.org/book/>
- A primer on spaCy tool for NLP is published here: <https://course.spacy.io/en>

Supplemental/Recommended Materials

Recommended online primers, tutorials, resources, and other texts may include:

- Python: <http://www.python.org/>
- Python refresher: <https://wiki.python.org/moin/BeginnersGuide>
- Learn and Practice Python: <http://inventwithpython.com/>

Required Computer

UF student computing requirement: <https://news.it.ufl.edu/education/student-computing-requirements-for-uf/>

Course Schedule

Below is a general schedule for the semester. Please refer to the Canvas *Announcements* and *Modules* for final details about each topic, exact dates, and the specific schedule we implement this semester.

Module	Topic	Assessment Due	Activity & Scaffolding	Instructional Materials
0	Orientation: What is NLP? What are LLMs?	Orientation Videos and Quiz SLACK & BLADE Enrollment	Orientation Videos and Quiz, SLACK & BLADE Enrollment	SLP 1 intro, 1.1-1.4(draft) <i>(not online after 2023)</i>
1	Language & Intelligence, NLP Pipeline, Ambiguity, Applications	Quiz 1	Module 1 Videos, Quiz 1, Project 0	SLP 2 intro, 2.1-2.2, 18 intro, 18.1, 19 intro, 19.1, 19.1.1; spaCy 101, NLTK, Simple ELIZA.ipynb
2	Text Processing in Python, Corpus Analysis	Quiz 2	Module 2 Videos, Quiz 2, Project 0	SLP 2.3-2.7, spaCy 101, NLTK, CAP 4641 NLP Tools, Simple-Tokenizer.ipynb
3	N-gram Language Models	Quiz 3 Project 0	Module 3 Videos, Quiz 3, Project 0	SLP 3.1-3.4, 3.6
4	Context Free Grammars (CFGs), Constituency Trees, CFG Parsing	Quiz 4	Module 4 Videos, Quiz 4	SLP 18.2, 18.5, D.1-4
5	Grammar Conversion, CKY, Dependency Parsing	Quiz 5	Module 5 Videos, Quiz 5, Project 1	SLP 18.4, 18.6, 19.2, D.5

6	Part of Speech (POS) & Named Entity (NE) Tagging, Sequence Labeling	Quiz 6 Exam 1 Review Quiz	Module 6 Videos, Quiz 6, Project 1, Exam 1 Review Quiz	SLP 17.1-17.3, 18.3, SLP2 13.5 (<i>not online after 2022</i>) (Exam 1 covers Modules 1-7; Exam 1 Review quiz covers Modules 1-5)
7	Hidden Markov Models (HMMs), Viterbi, Conditional Random Fields (CRFs); Exam 1 Review	Quiz 7 Project 1 Exam 1 Review Videos Mid-Course Survey	Module 7 Videos, Quiz 7, Project 1, Exam 1 Review Videos, Exam 1 out (6 Day Window)	SLP 17.4-17.5, A.1-4, NLTK (Exam 1 covers Modules 1-7; Exam 1 Review quiz covers Modules 1-5; Review Videos cover Modules 1-7)
8	Lexical Semantics, Word Meaning, Vector Semantics; Exam 1	Exam 1 (<i>Modules 1-7 only</i>) Quiz 8	Module 8 Videos, Quiz 8, Exam 1 out, Project 2	SLP 6.1-6.3 Mathematical Explanation Using Cosine Similarity
9	Word Similarity, Vector Embeddings, Basic Metrics	Quiz 9	Module 9 Videos, Quiz 9, Project 2	SLP 6.4-6.8, SLP 4.7
10	Semantic Role Labeling (SRL), Research Talk	Quiz 10 Project 2	Module 10 Videos, SRL Talk & Project 3 Video (Youm), Quiz 10, Project 2	SLP 21.1-21.6
11	Contextualized Embeddings, BERT	Quiz 11	Module 11 Videos, NN Video (Lehman), Quiz 11, Project 3	SLP 9.1, 10.1, 10.3, 11.1-11.3
12	Machine Translation, Divergences, Encoder-Decoders, Gen AI Research Talk	Project 3 Exam 2 Review Quiz	Module 12-13 Videos, NLP/Gen AI Research Video (Dorr), Quiz 12, Project 3, Exam 2 Review Quiz	SLP 13.1-13.3, 13.6.1-13.6.2 (Exam 2 covers 1-13 with emphasis on 8-13; Exam 2 Review quiz covers 8-13)
13	NLP Eval Methodologies, Metrics; Exam 2 Review	Quiz 12 Exam 2 Review Videos	Module 12-13 Videos, Quiz 12 (cont), Exam 2 Review Videos; Exam 2 out (5 Day Window)	Review materials (Exam 2 covers 1-13 with emphasis on 8-13; Exam 2 Review videos cover 8-13)
14	Exam 2	Exam 2 (<i>Modules 1-13, emphasizing 8-13</i>)	Exam 2	Review materials

Format and Expectations, Attendance and Makeup Policy, Asking Questions

(This does not apply to UOnline students.) Class time occurs in our assigned face-to-face classroom [F2F] and virtually [online] via Zoom. If you're feeling unwell, please prioritize your recovery and stay home. (More on this point under "General Attendance" below.)

Format: For F2F students, this course is taught as a "flipped classroom." (*UOnline students are entirely online*). **For all students:** Formal lectures are pre-recorded and delivered within Canvas modules. Each module contains 3–5 short video snippets with embedded PlayPosit "thought questions" to reinforce key concepts, with multiple attempts allowed. A multiple-attempt quiz is also included at the end of each module to check your understanding. You are responsible for completing all videos, thought questions, and quizzes by their posted due dates, starting with those released on the first day of the semester (see the course schedule). Additionally, four projects provide hands-on experience applying course concepts using Python, and two online (single-attempt) exams assess your understanding of multiple modules.

F2F Class time is dedicated to discussion of class themes and preparation for upcoming assessments, based on students' questions about course content. You **must come prepared with questions**, as these sessions will be highly interactive and driven by students. (You may be asked to login to the system and share your screen for specific questions.) Do not expect a formal lecture.

General Attendance: If you are in the F2F section and need to miss class, *never* email the professor or TA asking what was covered. Check inside Canvas for posted zoom recordings and/or ask fellow students for their notes on that day's class period. (*UOnline students are entirely online.*) Here are additional points of significance:

- Class attendance will not be taken (in-person attendance is not mandatory).
- The F2F class time will be recorded (see the disclaimers and notice that follow).

Disclaimer #1: Zoom recordings of F2F sessions will be posted in the corresponding Canvas module. It can take a bit for the recording to process, so check back about 2–3 hours after class.

Disclaimer #2: The Zoom platform records these sessions and I cannot guarantee they will not experience errors in the recording process.

Notice to F2F Students: Those who choose to participate with their camera engaged or utilize a profile image are agreeing to have their video or image recorded. If you are unwilling to consent to have your profile or video image recorded, be sure to keep your camera off and do not use a profile image. Likewise, students who un-mute during class and participate orally are agreeing to have their voices recorded. If you are not willing to consent to have your voice recorded during class, you will need to keep your mute button activated. A text "chat" feature is also available during F2F sessions. "Chats" sent through the public channel may be recorded. If you do not wish for your "chat" to be recorded, you must send it through the private channel directly and only to the professor. Since the F2F session is in a public setting, private chats may or may not always be answerable. As in all courses, unauthorized recording and unauthorized sharing of recorded materials is prohibited.

Excused absences: In this course, exceptions for missed graded activities are rare and apply only to events you could not have anticipated before or during the multi-day window of the activity (quizzes, projects, and exams). The Dean of Students Office (www.dso.ufl.edu) provides rules, guidance, and approval for excuse documentation. An instructor notification letter from the DSO must directly state that an **"absence has been excused"** and the letter must specify the dates of the entire multi-day duration of the assessment. If the letter does not meet these requirements, the student must provide

the instructor with documentation that complies with UF's official excused absence policy (<https://catalog.ufl.edu/UGRD/academic-regulations/attendance-policies/#absencetext>). Without this, the absence will not be considered excused, and accommodations cannot be provided.

Makeups or extensions: Exam makeups or extensions are provided if appropriate documentation is presented in a timely fashion (see above). However, flexibility is built into the schedule: exams are 90-120 minutes and may be taken anytime within a 5-6 day window. This usually avoids the need for makeups or extensions and lets you plan ahead for a timeframe that fits scheduling constraints.

Late Assignments: A **50% penalty will be applied** if your submission arrives after the deadline. Any quiz, exam, or project that is submitted **more than 24 hours after the deadline cannot be accepted**. For example, a project that is due on Sunday at 11:59pm is **late** starting after 11:59 Sunday night (even if just one second late), and will be graded 50% off up until 11:59pm on Monday night. No credit is awarded starting after 11:59pm Monday.

Protocol for questions about course content (F2F students): Start by taking advantage of the flipped classroom format. If you raise a question outside of class that is better suited for the flipped setting, the professor or TA may ask you to present it to the whole class. The weekly interactive meetings are designed for problem solving, discussion of class themes, preparation for upcoming assessments, and review of key elements of graded activities. **Arrive each week with questions in hand** so they can be answered during class—and so your classmates can benefit from the discussion.

Protocol for questions about course content outside class (ALL students): Be sure you have thoroughly watched the lecture videos and completed the associated thought questions—it's likely you'll find answers to your questions there. If questions remain, **first** post them in the **#general** Slack channel—without revealing potential solutions—so others can benefit from the exchange. Many students have similar questions, and peer interaction often leads to shared assistance. This approach reduces redundancy and ensures common questions are addressed for everyone. You will also be introduced to BLADE, a new AI-powered, 4641-specific chat agent to help answer course-related questions. TAs and graders may hold weekly Zoom office hours, which are another way to ask questions, but whenever possible, use methods that allow the whole class to benefit.

Note: Private matters that cannot be addressed in a general channel (e.g., an emergency situation) may be sent via Canvas—**please include the lead TAs, or you are likely not to receive an answer**. Do not write to request additional points or extra credit. Also, see *Re-grades* under "Grading Policy".

Evaluation of Grades: All assessments are scheduled for multi-day time windows. Total Points and Grade Percentage for each assessment type are provided below.

Activity	Total Points	Percentage of final grade
PlayPosit Videos and Thought Questions	40 (40 videos): 1 pt per video, <u>multiple</u> attempts.	14%
Quizzes	24 (12 quizzes): 2 pts per quiz, <u>multiple</u> attempts.	16%
Projects: programming tasks coupled with programming quizzes	97 (4 projects): 15 pts, 20 pts, 27 pts, 35 pts. <u>Multiple</u> attempts for programming tasks. <u>Single</u> attempt only for programming quizzes.	40%
Exams	60 (2 exams): 27 pts, 33 pts. <u>Single</u> attempt only.	30%
Total	221	100%

Your grade is based on four components—PlayPosit Videos/thought questions, quizzes, exams, and projects—all of which are subject to the Excused Absence (Missed Graded Activities) and Late Assignment policies described above. For each category, the combined score for all assessments determines the percentage of your final grade, as specified in the course grading table above.

Videos/PlayPosit (14%) - Each module includes short lecture videos with embedded PlayPosit “thought questions” to reinforce key concepts. There are 40 videos in total across Modules 0–13. Your score on these thought questions determines your grade for this category. Multiple attempts are allowed. The Orientation (Module 0) videos must be completed within the first few days of the semester. Doing so will unlock the remaining modules, which you may visit at any time thereafter.

Quizzes (16%) - Each module includes a Canvas quiz (2 points each) that tests what you learned from the videos. Your score in this category is the total number of points earned across all quizzes. Quizzes are open-book and allow multiple attempts. There are 12 quizzes in total across Modules 1–13. Canvas scoring* applies to Module Quizzes. (See Grading Policy.)

Projects (40%) - Four programming projects use Python in a Jupyter notebook format, combining automatically graded programming snippets (which can be executed iteratively until correct) with single-attempt quiz questions about key aspects of the assignment. Your score in this category is based on the total number of correct responses to both the programming and quiz questions. Instructions specify permitted resources and submission details. Projects generally span 2–3 weeks, with more time allotted for heavier-weighted projects later in the semester. **Start early.** Canvas scoring* applies to Project Quizzes. (See Grading Policy.)

Exams (30%) - Two open-resource online exams will be administered (textbook and Canvas materials). Your score in this category is based on the total number of correct answers across both exams. Exam 1 covers Modules 1–7; Exam 2 covers Modules 1–13 (with an emphasis on Modules 8–13). Each exam is timed (90 or 120 minutes) and may be taken at any chosen day/time within a pre-assigned 5–6 day window. Each exam draws on all course materials provided up to that point. Canvas scoring* applies to Exams. (See Grading Policy.)

Grading Policy: The range used to calculate your final letter grade in our course will be no harsher than the grading scale provided in this table. Your final point total will be calculated using the percentages given in the *Evaluation of Grades* section. The percent you earn on each activity will be multiplied by the grade points associated with that activity. More information on UF grading policy may be found at: <https://catalog.ufl.edu/UGRD/academic-regulations/grades-grading-policies/>

***Canvas Scoring:** Canvas awards partial credit on multiple-answer questions—but not if too many wrong answers are chosen. Points are added for correct answers and deducted for incorrect ones. For example, if a 9-point question has 3 correct answers, each correct choice earns 3 points and each wrong choice loses 3 points. Selecting all options will lower your score if wrong answers are chosen.

Grade Points	Letter Grade	Highest	Lowest
4.00	A	100+	92.00
3.67	A–	91.99	89.00
4.00	A	100+	92.00
3.67	A–	91.99	89.00
3.33	B+	88.99	86.00
3.00	B	85.99	82.00
2.67	B–	81.99	79.00
2.33	C+	78.99	76.00
2.00	C	75.99	72.00
1.67	C–	71.99	69.00
1.33	D+	68.99	66.00
1.00	D	65.99	62.00
0.67	D–	61.99	59.00
0.00	E	58.99	0.00

Re-Grades

All assessments are electronic and automatically graded, with a complete record of attempts, saves, final answers, and assigned grades. If you believe there is a technology error or disagree with the answer to a specific question, **first** post your concern in the **#general** Slack channel—without revealing potential solutions—clearly stating the issue and providing a justification. This allows the class to confirm whether others have the same concern. If the **#general** discussion does not resolve the issue, contact the [entire instructional team](#) via Canvas. Regrades typically affect the entire class, and any necessary adjustments will be applied electronically to all students at once.

Right to change information

Although every effort has been made to be complete and accurate, unforeseen circumstances arising during the semester may require adjustments. Consequently, given due notice to students, the instructor reserves the right to change any information on this syllabus or in other course materials.

The UF grading policy is at: <https://catalog.ufl.edu/ugrad/current/regulations/info/grades.aspx>

Visit <https://go.ufl.edu/syllabuspolices> for Academic policies and resources—including student accommodations—as well as Campus Health and Wellness Resources.