



Introduction to Computational Linguistics

Course Number	L44 Ling 317	Instructor	Dr. Nick Danis
Semester	Spring 2019	Contact	nsdanis@wustl.edu
Time	Tu./Th. 2:30–4:00pm	Office Location	Psychology 402F
Location	Eads 003	Office Hours	Wed 1–3pm or by appt.

1 Description

This course introduces the computational tools, both practical and theoretical, to describe and analyze natural language. We will touch upon most of the major aspects of the field of computational linguistics and natural language processing, such as analysis of raw corpora; modeling phonological, morphological, & syntactic patterns; learnability; and speech processing. We will use Python 3 throughout the course. This course is an introduction to these topics, and therefore no programming experience is required. (In fact, this course is *not* meant for students with a strong background in computer science.) Students, however, are encouraged to complete optional exercises and readings on basic programming techniques to aid in the completion of assignments and discussions. Prerequisite: L44 Ling 170D and either L44 Ling 258 or CSE 131.

2 Course Goals

- Learn practical, computational tools (e.g. Python NLTK) to manipulate natural language data
- Understand the computational properties of natural language, independent of whether it is being computed by machines or humans (i.e. treating grammars as purely mathematical objects)
- Apply these techniques to solve problems in phonology, morphology, and syntax

3 Required Materials

There are no required purchases for this course. Any readings will be made available online. Additionally, all software necessary is free as well. This course assumes you have (or have access to) a working computer with a desktop operating system (macOS, Windows, Linux, etc.).

3.1 Readings

A majority of the readings will come from the following source:

- Steven Bird, Ewan Klein & Edward Loper. 2014. *Natural Language Processing with Python*. 2nd edn. <http://www.nltk.org/book/>

This book is available for **free** via the URL above. (You do **not** have to purchase anything.) Note that we are using the 2nd edition, which is different from the printed 1st edition. All other readings will be posted as PDFs when necessary.

3.2 Software

The majority of work will be completed in Jupyter Notebook. These documents incorporate rich text and Python code that can be run in the document itself. We will go over installation instructions in class. Installation includes a Python install as well. This course uses Python 3.

- Jupyter Notebooks: <http://jupyter.org/>
- Python Natural Language Toolkit (NLTK): <https://www.nltk.org/>

I recommend **not** installing anything until class begins, unless you feel comfortable doing so. If you already have a Python build installed, you may use it; however, all assignments will be based on Jupyter Notebooks, so the final version of your homework should be submitted in this format.

4 Grade

Your grade breakdown is shown below.

Item	Pct.
Participation & Discussion	10%
Weekly Assignments	40%
Midterm Exam	20%
Final Project	30%

4.1 Participation & Discussion

Attendance is required. If you have multiple unexcused absences, it will be impossible to receive full credit in the category above. There will be approximately five (5) in-class assignments (short group work/problem sets). If there end up being more than 5 in the semester, only your highest 5 will be counted.

Additionally, as part of this category, all students are expected to participate in this course's Piazza site, which is available via the link below. Rather than emailing content questions to me directly, I encourage you to post your questions on Piazza. Both myself and other students can then collaborate to answer. Find our class page at: <https://piazza.com/wustl/spring2019/144ling317/home>

4.2 Weekly Assignments

The bulk of the coursework consists of weekly, take-home assignments. These will be distributed as Jupyter notebooks that you must complete, and they must be submitted online. The lowest scoring assignment will be dropped. Unexcused late submissions cannot be accepted.

4.3 Midterm Exam

The in-class midterm will test the concepts covered in class to that point. You will not be asked to create novel code for the midterm, but you should be familiar with the basic programming techniques that we use in class (e.g. loops, functions). There will be a review session before the exam.

Schedule (tentative)

Date	Day	Comment	Reading	Due
1/15/19	Tu.	Introduction, programming basics	Stillings et al - Natural Language Processing	
1/17/19	Th.	Mathematical basics, installing software	NLTK ch 1	Notebook 0
1/22/19	Tu.	Chatbots	Searle - Mind, brains, and programs	
1/24/19	Th.			Notebook 1
1/29/19	Tu.	Regular Expressions	Introducing Regular Expressions	
1/31/19	Th.			Notebook 2
2/5/19	Tu.	Raw Text/Corpora	NLTK ch 2-3	
2/7/19	Th.			Notebook 3
2/12/19	Tu.	Ngrams	NLTK 4-5	
2/14/19	Th.			Notebook 4
2/19/19	Tu.	Finite State Automata (FSAs)	Sipser 2012 ch 1	
2/21/19	Th.			Notebook 5
2/26/19	Tu.	Building FSAs in Python	Jurafsky and Martin 2007 ch 3	
2/28/19	Th.			Notebook 6
3/5/19	Tu.	Review		Proposal
3/7/19	Th.	Midterm		
3/12/19	Tu.	<i>No class (Spring Break)</i>		
3/14/19	Th.	<i>No class (Spring Break)</i>		
3/19/19	Tu.	Context-Free Grammars (CFGs)	NLTK ch 9 Chomsky - Syntactic Structures ch 1-4	
3/21/19	Th.			Notebook 7
3/26/19	Tu.	CFGs	Sipser 2012 ch 2	
3/28/19	Th.			
4/2/19	Tu.	Learning	Heinz and Idsardi 2013	
4/4/19	Th.		Heinz - Learning Long-Distance Phonotactics (optional)	Notebook 8
4/9/19	Tu.	CLASS CANCELED		
4/11/19	Th.	Speech Processing/Synthesis	Ladefoged 2005 ch 7-8	Notebook 9
4/16/19	Tu.	Presentations		
4/18/19	Th.	Presentations		
4/23/19	Tu.	Presentations		
4/25/19	Th.	Presentations		

4.4 Final Project

The final project is broken down into three parts: the proposal, the presentation, and the project itself. More detailed information will be distributed in class closer to the due date; below is simply a rough guide.

For the project, you must choose one of the following:

- **Paper:** Write a 5-6 page research paper on some topic related to this course, either theoretical or applied.
- **Program:** Write an annotated Python program that tackles or expands on some linguistic problem related to the class.

The **proposal** must be between 300–500 words, describing which type of project you choose and the preliminary description of what it will entail. The **presentation** will be around ten minutes (including questions from the class) where you reveal the problem or topic you are focusing on as well as any preliminary solutions or conclusions you have at that point. The proposal and the presentation are each worth 5% of your final course grade. The final project, whether it is a paper or a program, will be due during finals period (exact date TBD).

5 Academic Integrity

This course adheres to the university's Academic Integrity Policy (<https://studentconduct.wustl.edu/academic-integrity/>), and takes cheating and plagiarism very seriously.

6 ADA Compliance

Washington University is committed to providing accommodations and/or services to students with documented disabilities. Students who are seeking support for a disability or a suspected disability should contact Disability Resources at 935-4153. Disability Resources is responsible for approving all disability-related accommodations for WU students, and students are responsible for providing faculty members with formal documentation of their approved accommodations at least two weeks prior to using those accommodations. I will accept Disability Resources VISA forms by email and personal delivery. If you have already been approved for accommodations, I request that you provide me with a copy of your VISA within the first two weeks of the semester. Please see more information at <http://cornerstone.wustl.edu/>.