



RUTGERS

School of Arts and Sciences

Computational Linguistics

01:615:455

Spring 2023

Instructor: Natalia Tyulina

Email: TBD

Lecture: Friday 12:10-3:10, CS 115 (Scott Hall Room 115)

Student Support Hours: Online by appointment

Course Description

Main goal of this course is to introduce linguistics majors to programming as a tool for analyzing various language phenomena. No prior knowledge of programming is required.

Course Learning Goals

By fully participating in this course, students will be able to:

- Understand basic techniques in designing programs (using the Python programming language) that manipulate natural language
- Count frequencies of different categories in text
- Understand and apply computational techniques to the analysis of empirical phenomena in different areas of linguistics
- Understand the computational properties of natural language and basic formalisms used in language technologies

Department learning goals met by this course:

- Students will reason about language; identify how incorrect or irrational assumptions and prejudices distort understanding of language; demonstrate knowledge about language in the world including a sophisticated understanding of linguistic and cultural variation, and evaluate popular views on the nature of human languages and their speakers.
- Majors and Minors will also demonstrate technical mastery over the tools of linguistic analysis in syntax, phonology and semantics and apply linguistic theory in these areas. They will investigate linguistic data and analyze it; demonstrate strong problem-solving skills; extend their understanding of theoretical linguistics into other domains of linguistic research; apply the techniques of linguistics that they have learned in the core courses to new topics; and access current research in the field. Some students will investigate language in a broader context, where it can be systematically and rationally explored using their sophisticated understanding how language works.

Prerequisites

01:615:201 Introduction to Linguistic Theory

Materials

There is no textbook for this course. Some required readings will be assigned, and links to the readings and/or pdf documents will be posted on Canvas. Students are strongly encouraged to bring a laptop to all lectures; we will devote some time to practice at the end of each class.

For coding we will be using Google Colab, which is an online Jupyter Notebooks environment from Google.

Assignments

Assignments will take the form of small coding projects accompanied by a write-up describing the general approach taken and any challenges encountered. Students will usually be able to verify the technical correctness of their code by running a provided unit test. Students will also be graded on the readability of their code, and the quality of the write-up.

The final assignment will be an open-ended mini-project which will involve either writing a Python program using the techniques learned in class to analyze some linguistic phenomenon, or a research paper on one of the relevant problems in computational linguistics. During the last class students will give short presentations describing their projects.

Grading

80% of students' grades will be derived from the assignments and the final mini-project (roughly the size of two homeworks); the remaining 20% will be reserved for participation and attendance. Assignments must be submitted on time or will receive a 0 grade.

Letter Grade Distribution

At the end of the semester the final grade will be converted to a letter grade according to the standard Rutgers system:

A	$\geq 90\%$
B+	85–89.9%
B	80–84.9%
C+	75–79.9%
C	70–74.9%
D	60–69.9%
F	$\leq 59.9\%$

Attendance and Absences

Attendance is expected and will be taken each class. You are allowed to miss 1 class during the semester without penalty. Any further absences will result in point and/or grade deductions. However, students who have reason to believe they may be contagious for COVID-19 or other infectious diseases should not come to class, and they need to notify the instructor ahead of time. Students are encouraged to report absences using the University self-reporting absence system: <https://sims.rutgers.edu/ssra>.

Students are responsible for all missed work, regardless of the reason for absence. It is also the absentee's responsibility to get all missing notes or materials.

Academic Integrity

All students are required to abide by [Rutgers Student Code of Conduct](#). Students are expected to work independently and accepting solutions (e.g., lines of code or write-up) from others is an act of plagiarism.

Discussion amongst students is encouraged, but when in doubt, direct your questions to the instructor.

Student Support and Mental Wellness

Below is the list of useful resources to support students in their academic success and mental wellness:

- Student Success Essentials: <https://success.rutgers.edu>
- Student Support Services: <https://www.rutgers.edu/academics/student-support/>
- The Learning Centers: <https://rlc.rutgers.edu>
- Rutgers Libraries: <https://www.libraries.rutgers.edu>
- Bias Incident Reporting: <https://studentaffairs.rutgers.edu/bias-incident-reporting>
- Office of Veteran and Military Programs and Services: <https://veterans.rutgers.edu>
- Student Health Services: <http://health.rutgers.edu>
- Counseling, Alcohol and Other Drug Assistance Program Psychiatric Services (CAPS): <http://health.rutgers.edu/medical-counseling-services/counseling/>
- Office for Violence Prevention and Victim Assistance: www.vpva.rutgers.edu

Disability Accommodations

Rutgers University welcomes students with disabilities into all of the University's educational programs. In order to receive consideration for reasonable accommodations, a student with a disability must contact the appropriate disability services office at the campus where you are officially enrolled, participate in an intake interview, and provide documentation:

<https://ods.rutgers.edu/students/documentation-guidelines>. If the documentation supports your request for reasonable accommodations, your campus's disability services office will provide you with a Letter of Accommodations. Please share this letter with your instructors and discuss the accommodations with them as early in your courses as possible. To begin this process, please complete the registration form: <https://webapps.rutgers.edu/student-ods/forms/registration>.

Week	Week of	Topic	Notes
1	1/20	Syllabus & motivations; intro. to Python	
2	1/27	Literals, variables, operators, types	Joyner §2, Shaw §1-14
3	2/03	Control flow	Joyner §3-3.3
4	2/10	Functions; imports	Joyner §3.4
5	2/17	Classes; unit tests	Joyner §5.1, unittest
6	2/24	Regular expressions	HOWTO , re , Bird et al. §3.4
7	3/03	Text encoding	Bird et al. §3.3 , Gorman , chardet , unicodedata
8	3/10	Structuring data; dataframes	Project proposal due
9	3/17	No class: spring break	
10	3/24	N-gram models	J&M §3
11	3/31	Formal language preliminaries; set theory	
12	4/07	Finite-state automata (FSA)	
13	4/14	Context-free grammars (CFG)	J&M §17
14	4/21	NLTK	
15	4/28	WordNet	J&M §23
16	5/05	Presentations	Final project due

(Please note that this is subject to change.)

References:

Bird, S., Klein, E. and Loper, E. (n.d.) *Natural Language Processing with Python*. <https://www.nltk.org/book>

Joyner, D. 2016. *Introduction to Computing*. McGraw-Hill Education.

Jurafsky, D., and Martin, J. H. 2023. *Speech and Language Processing*. [3rd edition draft](#)

Shaw, Z. A. 2017. *Learn Python 3 the Hard Way: A Very Simple Introduction to the Terrifyingly Beautiful World of Computers and Code*. Addison-Wesley Professional.