

# Linguistics 120: Language & Technology

## Syllabus

August 23, 2021

## 1 Basics

Course: LIN 120

Lectures: Tuesday Thursday 1.15pm-2.10pm

Course Location: Frey Hall, Room 100

Instructor: Owen Rambow

Email: [owen.rambow@stonybrook.edu](mailto:owen.rambow@stonybrook.edu)

Office hours: Tuesday 3-4pm, Wednesday 2.30-4.30pm

Office: SBS N257

Sections: Friday 9.15-10.10 (Chemistry 128, Frey3128, or Melville N4072) or  
Friday 10.30-11.20 (Humanities 1023, Melville N4006, or Melville W4530)

TAs: Salam Khalifa [salam.khalifa@stonybrook.edu](mailto:salam.khalifa@stonybrook.edu),

Kalina Kostyszyn [kalina.kostyszyn@stonybrook.edu](mailto:kalina.kostyszyn@stonybrook.edu),

Yang Liu [yang.liu.11@stonybrook.edu](mailto:yang.liu.11@stonybrook.edu)

Undergraduate teaching practicants: Alana Gill [alana.gill@stonybrook.edu](mailto:alana.gill@stonybrook.edu),  
Michelle Wang

Technology required: CoCalc, <https://cocalc.com/app>

## 2 Bulletin Description

This is the current description of this course in the Stony Brook Bulletin.

An introduction to how computers process language and solve language-related tasks. This course discusses the language technologies of our daily life – spam filtering, machine translation, and many more – and shows how they work under the hood. The course explores a variety

of issues: Why do computers do well in some areas (spell checking) yet fail miserably in others (essay grading)? Will we ever have perfectly fluent AIs as depicted in science fiction? And how will these technological advances impact the role of language in our society? Students will also acquire basic programming skills and write scripts for simple language tasks. No previous training in mathematics or computer science required.

- SBC: TECH
- Prerequisites: none
- Credits: 3

## **3 The Course in a Nutshell**

### **3.1 The Big Take-Home Message**

We all use language technology on a daily basis (Google Translate, Siri or other conversational agents, email spam filters, Google). It has improved massively over the last 20 years. Some tasks are now performed quite well, but other language tasks remain elusive for language technology.

### **3.2 Major Questions**

- How do computers process language?
- Why do computers perform better than humans at some tasks (spell checker, spam filter), near-human at some (speech recognition, translation), but fail at deep language understanding?
- Will we ever have systems that have perfectly natural conversations with humans as depicted in science fiction (2001, Star Trek, Blade Runner, Her, Ex Machina, System Shock, Super Intelligence)?
- What is the role of machine learning?
- What are ethical considerations of language technology?

### **3.3 Teaching Goals of the Course**

The teaching goals relate to understanding language technology, and to learning to program in Python. These goals are related: only if you have a hands-on understanding of programming can you understand language technology.

- Acquire basics of programming in Python and of computer science.

- Understand the importance of algorithms and data structures, and learn to conceptualize linguistic problems in computational terms.
- Acquire basic programming skills in Python.
- Understand basic notions in cognitive science and natural language processing (NLP).
  - Develop familiarity with notions of natural language processing, a subfield of artificial intelligence.
  - Understand the role of machine learning and big data in NLP.
  - Understand how and why humans and computers differ in their linguistic abilities.
  - Develop awareness of the risks of language technology to society (surveillance, language death).
- Understand applications of natural language processing to various problems. The goal is also to enable you to use NLP in your future research, even if it is not directly related to language or NLP.
  - Learn to work with text and text corpora.
  - Learn about existing technologies in NLP.

### 3.4 SBC: TECH

The SBU TECH requirement is defined as follows. It should be clear from the above teaching goals how this course (if successful) fulfills the TECH requirements.

1. Demonstrate an ability to apply technical tools and knowledge to practical systems and problem solving.
2. Design, understand, build, or analyze selected aspects of the human-made world. The “human-made world” is defined for this purpose as “artifacts of our surroundings that are conceived, designed, and/or constructed using technological tools and methods.”

## 4 Prerequisites

### 4.1 Formal Prerequisites

None

## 4.2 What You Need to Succeed

- Ability to operate a computer (use a web browser, install software, edit text files)
- Willingness to play around with open-ended problems

## 4.3 What You Do NOT Need to Succeed

You do not need any of the following:

- Programming experience
- Math (except for addition, multiplication and fractions)
- Linguistics (LIN 101 helps a bit, though)

## 4.4 Technology Requirements

You have to sign up for a CoCalc account for the semester, which is \$14. You must sign up with your `stonybrook.edu` email address. We will then bring you in to the CoCalc project for this course.

Instructions for our situation (you pay for course): <https://doc.cocalc.com/teaching-students.html#student-pay-for-course>

Class materials and discussions will be hosted on Blackboard, where you should have an account through Stony Brook.

This class requires you to have access to some device with a web browser and a keyboard (for access to CoCalc and Blackboard). If you do not have one, consider a laptop loan: <https://www.stonybrook.edu/commcms/studentaffairs/student-support/> . For additional IT support: <https://it.stonybrook.edu/services/itsm> .

## 5 Schedule by Week

The weekly schedule consists of two lectures and a recitation. Usually, one lecture will focus on the theory behind language technology, and the other will explore programming with Python.

The approximate schedule for topics and assignments is as follows.

Week	Date	Topic Theory	Topic Python
1	8/24	Syllabus	CoCalc tutorial
2	8/31	Language Technology	Python Basics
3	9/7	Dialog Systems	Strings
4	9/14	Word Based Models	Control Flow
5	9/21	Word Based Models, ctd.	Lists & Loops
6	9/28	N-grams	Practice
7	10/5	N-grams, ctd.	Midterm
8	10/12	Fall Break	String Cleaning
9	10/19	Language and Structure	Functions
10	10/26	What is Meaning?	Tokenizing
11	11/2	Machine Learning	N-grams
12	11/9	Machine Learning, ctd.	Frequencies
13	11/16	Deep Learning	Frequencies, ctd.
14	11/23	Ethical issues in Natural Language Processing	Thanksgiving Break
15	11/30	Wrapping up	Final

## 6 CoCalc

We will be using CoCalc for this class, a cloud-based service for collaborative programming and data science. Your assignments will be automatically collected and graded on CoCalc after the deadline. All you need for CoCalc is a device with a recent browser (Windows, OS X, or Linux computer, Chromebook, Android tablet; iOS devices may run into some issues, but will retain general functionality). While CoCalc can be used for free, we will upgrade to a premium tier for this class, which costs you \$14 per semester. There are no other costs for this class (no clickers or mandatory textbooks to purchase).

## 7 Mode of Instruction

This course will be an in-person lecture course. Lectures on Tuesdays will be about natural language processing and language technologies; lectures on Thursdays will be about the Python programming language.

Course materials for the Tuesdays lectures will be available through Blackboard. Course materials for the Thursday lectures will be available through CoCalc (see Section 6).

Recitation sessions will also be in-person. You sign up for one of six recitation sessions.

## 8 Grading

This course can only be taken for 3 credits. The overall grade reflects the split of the course into a general introduction to language technology, and an introduction to Python. Student grades are determined by the following components.

### 8.1 Python exercises (20%)

These are small Python homeworks that make sure you are keeping up with learning Python.

- Once per week.
- Programming in Python.
- Assigned on Wednesdays at 11:59pm.
- Due the following Wednesday at 11:59pm.
- No late hand-ins (automatically collected by CoCalc when due). Submission is mandatory.
- Each assignment will have one of three grades: 0 (completely wrong); 1 (right idea but does not work as done); 2 (working solution).
- The worst three assignment grades will be dropped (so there will be 10 grades from the homeworks).

Why these regular homeworks? Learning programming is like learning a new language, it needs constant practice. The assignments are mandatory to force you into practicing on a regular basis. They barely count towards your letter grade because this makes it easier for you to focus on mastering the material and stop obsessing about your grade.

### 8.2 In-Class Python Midterm/Final (20% + 10%)

- You have to take both a midterm and a final.
- They are in-class exams.
- Each one tests the material from the previous half semester.
- If you have a higher grade on your midterm than on your final, then the midterm is 20% of your grade and the final is 10%.
- If you have a lower grade on your midterm than on your final, then the midterm is 10% of your grade and the final is 20%.

### 8.3 Class participation (25%)

We have created a discussion board for you on CoCalc. Use it to:

- Ask questions (about Python, language technology, linguistics, computers, or anything else).
- Help fellow students by answering their questions.
- Link to online materials you found that are related to material covered in the lectures.

Additionally, recitation with your TAs is a perfect opportunity to discuss applications of your code, debugging, and many of the same topics we'll ask you to talk about online. Why? We want you to ask many questions, all the time. You can learn a tremendous amount by helping others. By pointing out interesting material you have come across, you make the class more enjoyable for everybody, including your TAs and the instructor.

### 8.4 Debates and Small Homeworks (25%)

Each week (except two), there will be one of two types of small weekly exercises:

- Technology: A task to explore a specific online language technology, and to see if you can break it. You will need to interpret the result.
- Debate: You will be asked to discuss a specific topic on Blackboard with your peers. Typically, this will involve reading an article from the popular press and reflecting on it.

These will be assigned just after the Tuesday lecture, and will be due Sunday 11.59pm.

You have to submit evidence that you have explored the technology or made meaningful contributions to the debate. You will be given explicit instructions for how to do each task. For the debate, the general structure will be that a prompt will be given on a specific topic related to the theoretical aspects of the class, and students will be asked to share their thoughts and comment on other students' posts.

## 9 Final Remarks

### 9.1 Getting Help

For questions related to class materials outside of office hours:

- By default: use discussion forums on CoCalc; fellow students may have the same questions, or perhaps one of your colleagues can answer your question. This will help facilitate discussion, and everyone will learn something about the problem at hand.
- If this doesn't address the question or brings up additional problems, email your TA.
- If your TA can't help, contact the instructor.

For personal and confidential issues, reach out to your TA or the instructor

## 9.2 Mini-Homework before the First Thursday Lecture

Here are your tasks for the lecture on Thursday August 26:

- Carefully reread this syllabus.
- Sign up for CoCalc (create a new account using your Stony Brook email address – do not use any other email address!).
- Familiarize yourself with CoCalc's interface.
- Check out the *How to Ace This Class* document on CoCalc.

## 9.3 Supplementary Textbook

There are many Python textbooks; most are a bad fit for this course. We recommend:

Al Sweigart (2015): *Automate the Boring Stuff with Python*

The online version is free at <https://automatetheboringstuff.com>, while digital versions and hardcopy versions are around \$25. There are supplementary videos on Youtube: [https://www.youtube.com/playlist?list=PLGoJzB271\\_7r-iLYuEHEPJ5pSIYxXjJEn](https://www.youtube.com/playlist?list=PLGoJzB271_7r-iLYuEHEPJ5pSIYxXjJEn)

# 10 Policies

## 10.1 Student Accessibility Support Center Statement

If you have a physical, psychological, medical, or learning disability that may impact your course work, please contact the Student Accessibility Support Center, Stony Brook Union Suite 107, (631) 632-6748, or at [sasc@stonybrook.edu](mailto:sasc@stonybrook.edu).



They will determine with you what accommodations are necessary and appropriate. All information and documentation is confidential.

Students who require assistance during emergency evacuation are encouraged to discuss their needs with their professors and the Student Accessibility Support Center. For procedures and information go to the following website: <https://ehs.stonybrook.edu//programs/fire-safety/emergency-evacuation/evacuation-guide-disabilities> and search Fire Safety and Evacuation and Disabilities.

## **10.2 Academic Integrity Statement**

Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person's work as your own is always wrong. Faculty is required to report any suspected instances of academic dishonesty to the Academic Judiciary. For more comprehensive information on academic integrity, including categories of academic dishonesty please refer to the academic judiciary website at [http://www.stonybrook.edu/commcms/academic\\_integrity/index.html](http://www.stonybrook.edu/commcms/academic_integrity/index.html) .

## **10.3 Critical Incident Management**

Stony Brook University expects students to respect the rights, privileges, and property of other people. Faculty are required to report to the Office of Student Conduct and Community Standards any disruptive behavior that interrupts their ability to teach, compromises the safety of the learning environment, or inhibits students' ability to learn. Until/unless the latest COVID guidance is explicitly amended by SBU, during Fall 2021, "disruptive behavior" will include refusal to wear a mask during classes.

For the latest COVID guidance, please refer to: <https://www.stonybrook.edu/commcms/strongertogether/latest.php> .