Syllabus for LIN 637

Computational Linguistics 2 – Spring 2023

TuTh 2:00-3:20

Last Updated: January 26, 2025

COURSE WEBSITE: http://jeffreyheinz.net/classes/25S/637/

(clickable)

1 Course Outline

1.1 Description

This course is an introduction to theoretical foundations of computational linguistics. The course emphasizes the importance of algorithms, recursive data structures, formal language theory, finite-state machines, logic, and algebraic structures for describing and understanding natural languages, as well as their importance for the development of new tools and software applications. Empirical phenomena in phonology and syntax are sampled from a variety of languages to motivate and illustrate the concepts. Students will develop familiarity with the literature and tools of the field.

The course also serves as a bridge from introductory courses in linguistics (Syntax 1, Phonology 1, Phonetics) and computational methods (Statistics, Mathematical Methods in Linguistics, Computational Linguistics 1) to advanced courses and seminars in computational/mathematical linguistics. In contrast to the NLP courses offered by the department of computer science, this course focuses on studying the properties of natural language from a computationally informed perspective. The question is less about how to use computers for language-related tasks, but more about how language can be conceptualized as a computational problem.

1.2 List of topics

- *Open Problems and Questions*
  - Automata for recognizing and transducing unordered trees
  - Classifying complex morpho-phonological transformations
  - Classifying transducers used in NLP
  - Lifting subregular classes from string languages/relations to tree languages/relations
  - Learning string or tree acceptors or transducers from data

- *Formal Grammars and Languages*
  - The how, what and why of formalization
  - Recursive structures: Strings and Trees
  - Examples: Strictly Local, Strictly Piecewise, Tier-based SL, Regexp, Generalized Regexp, Star-free Regexp, MSO logic, Rewrite Grammars
- *String Languages*
  - Non-deterministic, 1-way, finite-state acceptors
  - Constructions
  - Deterministic, 1-way, finite-state acceptors
  - Powerset Construction
  - Deterministic, 2-way, finite-state acceptors
  - Algebraic Classes (including Definite, Reverse Definite, Generalized Definite, + Tier-ifications)
- *Tree Languages*
  - Non-Deterministic and Deterministic Top-Down Tree Acceptors
  - Non-Deterministic and Deterministic Bottom-Up Tree Acceptors
  - Constructions
  - Subregular Tree Languages
- *String Relations*
  - Non-Deterministic, 2-way, finite-state transducers (regular relations)
  - Non-Deterministic, 1-way, finite-state transducers (rational relations and functions)
  - Deterministic, 2-way, finite-state transducers (regular functions)
  - Deterministic, 1-way, finite-state transducers (sequential functions)
  - Composition
  - Subregular string relations
- *Tree Relations*
  - Non-Deterministic and Deterministic Top-Down Tree Transducers
  - Non-Deterministic and Deterministic Bottom-Up Tree Transducers
  - Subregular tree relations

Topics Covered and Anticipated Schedule

This schedule is tentative and subject to change.

Week	Dates	Topics
01	01/28 30	Defining Strings and Trees
02	02/04 06	rewrite grammars, regexps, intro to LTK
03	02/11 13	NFA for string languages
04	02/18 20	DFA for string languages
05	02/25 27	TD,BU NFA for tree languages
06	03/04 06	TD,BU DFA for tree languages
06	03/11 13	Algebraic Classifications of string languages
07	03/18 20	2NFT, 1NFT for string relations
08	03/25 27	SPRING BREAK
09	04/01 03	2DFT, 1DFT for string functions
10	04/08 10	TD,BU NFT
11	04/15 17	TD,BU DFT
12	04/22 24	TBD
13	04/29 05/01	Presentations
14	05/06 08	Presentations

1.3 Prerequisites

The only official prerequisite is Computational Linguistics 1 (Lin 537) or prior programming experience.

2 Graded Component

The final course grade depends on the following work.

Assignments	50%
Final Project Presentation	10%
Final Project	40%

- **Assignments**
  - Assignments will mostly be short and will begin during class time. If we finish them before class, they will checked for completeness. Otherwise, they are due at the beginning of next class.
  - There will be one longer assignment to classify a finite-state transducer for some aspect of natural language or NLP, using algebraic methods that have been implemented in software.
- **Project Presentation** You will present the current state of your research project in 20 minutes during the last two weeks of class. Your presentation should be accompanied by either a handout or slides.

- **Project**
  - You will conduct a project in this class and turn it in.
  - Projects may consist of theoretical research, program/software/tool development, a software-aided study of language data, or anything else I approve.
  - I want to encourage people to begin to work on this as soon as possible.

    Students are required to have a 1-1 meeting with me before February 24 to discuss potential final projects. Proposals can be submitted to me anytime after that meeting. I may return it with feedback for additional revision before approving it. Your project must be approved by me before the end of the day April 11.

    The proposal should: be 1-2 pages (300-500 words) in length, have a title, include references, provide background on the problem to be addressed (what it is and why it is important), provide an explanation or plan of how you approach the problem, provide an explanation of how you will measure success.

    Both the 1-1 meeting and the project proposal will count towards your Assignments grade.
  - The project itself is due Friday, May 16, 2025 at noon.

3 Additional Information

Readings Readings will be given regularly and made available on the course website or by email. It is presupposed in the lectures that you have done the required readings.

Workload per Credits

- *0 credits*: attend (but I highly recommend that you at least read the assigned papers as they will be important for following the lectures)
- *1 credit*: attend, participate, readings
- *2 credits*: attend, participate, readings, assignments
- *3 credits*: attend, participate, readings, assignments, project

Office Hours I will have office hours Tuesdays 11:30-13:00 and Wednesdays 1:30-3pm, and by appointment. You are encouraged to drop by even for short questions.

Policy on Collaborating Meeting with your classmates regularly to discuss course material and assignments is strongly recommended. Much learning occurs when working out problems with other people. However, each student must write-up their work (such as squibs) individually.

4 University Policies and Services

Attendance In the event of a short-term absence from class, students are encouraged to communicate immediately and work directly with instructors. However, if a student is

struggling with an extended absence due a hospitalization, family illness or death, they are encouraged to reach out to the Student Support Team.

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`. 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.

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. Faculty in the Health Sciences Center (School of Health Professions, Nursing, Social Welfare, Dental Medicine) and School of Medicine are required to follow their school-specific procedures. 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`

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. Faculty in the HSC Schools and the School of Medicine are required to follow their school-specific procedures. Further information about most academic matters can be found in the Undergraduate Bulletin, the Undergraduate Class Schedule, and the Faculty-Employee Handbook.