

Mathematical Methods in Linguistics

Fall 2022

Course	Info
Course#	LIN 361/539
Time	TueThu 9:45-11:05am
Location	Humanities 2047
Website	lin361.thomasgraf.net lin539.thomasgraf.net
Instructor	Thomas Graf
Email	[coursenumber]@thomasgraf.net
Office hours	W 9:30-10:15am (Seminar room) W 11:45-1:00pm (N-249) F 12:00-1:00pm (N-249)
Office	SBS N-249

A friendly plug: If you are interested in this class, also consider attending the department's Mathematical Linguistics Reading Group (<http://mlrg.thomasgraf.net>).

1 Course Outline

1.1 Bulletin Description

An overview of the mathematical foundations of theoretical and computational linguistics. Topics covered include set theory, morphisms, logic and model theory, algebra, lattices, lambda calculus, probability theory, information theory, and basics of formal language theory. A strong emphasis is put on the linguistic application of the mathematical concepts in the study and analysis of natural language data.

1.2 Full Description

This course is an introduction to mathematics in linguistics. It aims to help students familiarize themselves with mathematical concepts and applications that are widely relevant to theoretical

and/or computational linguistics. This covers a wide range of topics, mostly from *discrete mathematics*. The course is also very different from what you did in high school, there's precious few numbers here and we don't care much about trigonometry or calculating compound interest. In contrast to a proper mathematics course, we also focus more on techniques and tools rather than theorems and proofs. This means that you will learn how to work with things like matrices, semirings, and lattices, but you won't have to prove things about them. So this is more like a CS methods course than a proper math class.

For more information about the content, see the Selected Topics section. You will see that the schedule for this class is very ambitious. It has to be: this class serves an integral function of our Computational Linguistics MA and must get students to a level where they can take courses and read textbooks on mathematically demanding topics such as machine learning.

1.3 Teaching Goals

- master essential concepts and techniques in mathematics and theoretical computer science
- apply mathematical techniques to the study of language
- formalize linguistic ideas in mathematical terms
- develop learning autonomy and the ability to expand your mathematical knowledge through self-study

1.4 Prerequisites

No prior mathematical or computational experience is required.

I have run this course as an independent study with high school students who had no prior experience in linguistics, and they could follow along for the most part and only needed some help in specific areas.

1.5 Textbook

None, but there are detailed lecture notes that will eventually become a textbook. The lecture notes will be made available as PDFs in the course repository (check the pdf folder, please ignore the rest). Detailed instructions will follow via email.

2 Selected Topics

A brief selection of the topics to be covered (we will probably deviate from this order):

1. Basics of mathematics
 - Topics: sets, multisets, tuples, functions
 - Applications: bag of words model of text, n-gram models of grammaticality
2. Types of infinity
 - Topics: bijections, function inverse
 - Applications: is language infinite?

3. Relations and orders
 - Topics: properties of orders, posets, lattices, antimatroids
 - Applications: mereology, string extension learners, OT, feature systems, adjunct algebras, syntactic relations, linguistic universals
4. Graph theory
 - Topics: (un)directed graphs, connectedness, components
 - Application: parse forest representation, autosegmental phonology, AVMs, unification grammars
5. Automata theory
 - Topics: finite-state automata and transducers, regular expressions, push-down automata
 - Application: complexity of phonology & morphology VS syntax
6. Logic
 - Topics: propositional logic and first-order logic, types, lambda calculus
 - Application: semantics, model-theoretic syntax, subregular linguistics, CCG
7. Linear algebra
 - Topics: vectors and vector spaces, matrices, tensor product
 - Application: vector space semantics, spatial semantics, inflectional morphology
8. Abstract algebra
 - Topics: monoids, groups, semirings
 - Application: violation semirings in OT, semiring parsing
9. Probability theory
 - Topics: calculating probabilities with addition and multiplication
 - Application: weighted context-free grammars, corpus-based techniques
10. Information theory
 - Topics: entropy, cross-entropy
 - Application: probabilistic machine learning, surprisal for processing

3 Grading

This course can only be taken for 0 or 3 credits. Student grades are determined by the following components:

1. Pre-assessment (P/F; 5%)

At the beginning of the semester, students are asked to take a survey to assess their prior knowledge of mathematical linguistics. It is perfectly normal not to know a single answer. Bring the completed survey to the first session of week 2. Performance is P/F depending on whether a filled-out survey was submitted (answering “Don’t know” on each question is perfectly fine).

2. Feedback on weekly assignments (10% each time you volunteer, max 70%)

A list of exercises from the lecture notes is assigned every Thursday and your anonymous answers are due the following Tuesday. You should make a reasonable effort to complete the exercises, but your answers are not graded.

Instead, full solutions are distributed the same day. You can volunteer to help me go through the student answers and send me an email with your observations: where did folks struggle, what did they do well, which exercises caused confusion, and so on. You then return the assignments, completely unaltered, on Thursday.

- Thursday, week n : assignment n
- Tuesday, week $n + 1$: hand in your answers for assignment n , may volunteer to look at some of the assignments, official solutions distributed
- Thursday, week $n + 1$: answers are returned

Caution: In order to maintain anonymity among your peers, don't put identifying information on your assignments!

3. **Post-assessment (P/F; 5%)** Exactly the same as the pre-assessment.

4. **Take-home mid-term + final (A-F; remainder)**

We will have a midterm and final exam, but they won't be taken in class. Instead, they take the form of longer assignments and you are given more time than usual to complete them. These will be graded. The mid-term can only improve your grade:

1. If $\text{grade}(\text{mid-term}) \leq \text{grade}(\text{final})$: $\text{grade} = \text{grade}(\text{final})$
2. If $\text{grade}(\text{mid-term}) > \text{grade}(\text{final})$: $\text{grade} = (\text{grade}(\text{mid-term}) + \text{grade}(\text{final}))/2$

4 Policies

4.1 Contacting me

- Emails should be sent to [coursenumber]@thomasgraf.net. Disregarding this policy means late replies and might easily make me cross.
- Reply time < 24h in simple cases, possibly more if meddling with bureaucracy is involved.
- If you want to come to my office hours and anticipate a longer meeting, please email me so that we can set apart enough time and avoid collisions with other students.

4.2 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.

4.3 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 Technology & Management, 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

4.4 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/studentaffairs/geretogether/latest.php>