

Lecture 0

Syllabus

Course: Parsing and Processing
Course#: LIN630
Time: M 12:00–2:00, F 12:00–1:00
Location: CompLab SBS N250
Course Website: lin630.thomasgraf.net

Name: Thomas Graf
Email: lin630@thomasgraf.net
Office hours: M 11:15–12:00, F 11:15–12:00
Office: SBS N249
Personal Website: thomasgraf.net

1 Overview

- **Big Questions**

- What is the relation between competence and performance, grammar and parser?
- Are syntactic processing effects conditioned by the grammar?
- What qualifies as a parser as opposed to a recognizer or a parsing schema?
- Can we use insights from syntactic processing research to speed up current parsing technology?

The first two are common questions for any processing course. The third and fourth one hint at the special twist of this course: we approach these issues from a computational perspective! Parsing theory is a big (albeit messy) area of computer science, there's tons of parsing models on the market. So let's bring all these insights to bear on how humans parse natural language.

- **Teaching Goals**

At the end of this course you will

- be familiar with a variety of common parsing models (top-down, bottom-up, left-corner, Earley, CYK)
- know the most common syntactic processing effects (in particular those related to memory usage)
- be able evaluate claims in the psycholinguistic literature from a computational perspective

- **Prerequisites**

None beyond basic syntax skills — you should be able to draw a reasonable

tree for a sentence like *The fact that the employee who the manager hired stole office supplies did not go unnoticed by the janitor*. Knowledge of theoretical computational linguistics (e.g. as covered in Lin637) is helpful, but not necessary.

2 Course Requirements

- **Homeworks**

There will be weekly, peer-graded homeworks. Homeworks are essential if you want to learn anything in this course — you don't truly understand a parsing algorithm until you can carry it out yourself.

- **Peer Grading of Homeworks**

The best way to check your own understanding is to see if you can evaluate the solutions of your fellow students. That is why every week two students will be in charge of grading all the homeworks (with my help, of course).

- **Project Pitch**

At the end of the semester, you have to give me a 15 minute pitch for a follow-up project. This could be a psycholinguistic study, an implementation of a parsing algorithm, a proof of a specific theorem, whatever you find interesting. The pitch has to describe how the project intersects with your own research, explain its merit for linguistics or NLP, and draw from the material covered in this class. Ideally, you'll get me interested and we'll work on the project together.

- **Workload per Credits**

- *1 credit*: regular attendance, class participation
- *2 credits*: the above, plus doing all the homeworks
- *3 credits*: the above, plus peer grading and project pitch

Students who are retaking this course for credit can volunteer to give a guest lecture instead of a project pitch.

3 Outline

This outline assumes that we meet once a week for 110 minutes from 12:10 to 2:00pm. The time from 11:10 to 12:00 is reserved for discussion with the two peer-graders. If necessary, those 50 minute meetings can be moved to a different weekday so that students can get their corrected homeworks back before handing in the next batch.

A separate math & syntax primer will also be scheduled for week 1, and this time will be deducted from the very last meeting.

Wk	Chap	Topic
1	0	Organization, big picture and relevance
2	1,2	Parsing across disciplines, modular view of parsing
3	3	Top-down parsing
4	4	Top-down predictions for sentence processing
5	5	Bottom-up parsing
6	6	(Generalized) Left-corner parsing
7	same	same
8	Spring break	
9	7	Chart parsing overview, CKY
10	8	Earley parsing
11	10	Parsing with probabilities
12	12	Moving beyond context-free grammars
13	13,14	CFG-parsing of Minimalist syntax
14		guest lectures
15		guest lectures, Q&A

Depending on student interest, some of those chapters may be skipped to make room for other chapters, e.g. Dependency parsing, CCG parsing, or shallow parsing and partial parsing techniques.

4 Online Component

Rather than Blackboard, I use github to distribute lecture notes and readings (yes, you have to print them yourself). The (optional) readings are in a private repository, so you need a github account to access them. If you don't have one already, you can create one for free (github does not collect any user data). Make sure to email me your username asap so that I can give you access to the repository. If you do not want to use github for some reason, you can drop by my office to make an offline copy of the readings.

5 Policies

5.1 Contacting me

- Emails should be sent to lin630@thomasgraf.net to make sure they go to my high priority inbox. Disregarding this policy means late replies and is a sure-fire way to get on my bad side.
- 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 aside enough time and avoid collisions with other students.

5.2 Disability Support Services

If you have a physical, psychological, medical or learning disability that may impact your course work, please contact Disability Support Services, ECC (Educational Communications Center) Building, Room 128, (631) 632-6748. They will determine with you what accommodations, if any, 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 Disability Support Services. For procedures and information go to the following website: <http://www.stonybrook.edu/ehs/fire/disabilities>

5.3 Academic Integrity

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 are 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/uaa/academicjudiciary/>

5.4 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 Judicial Affairs 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.