Lecture 0

Syllabus

Course: Computational Syntax
Course#: Lin628
Time: TR 10:00–11:20am (changes tbd)
Location: SBS N117
Course Website: lin628.thomasgraf.net

Name: Thomas Graf
Email: lin628@thomasgraf.net
Office hours: tba
Office: SBS N249
Personal Website: thomasgraf.net

1 Course Outline

1.1 Bulletin Description

An in-depth survey of natural language syntax from a computational perspective. The primary focus is on combining state-of-the-art techniques from formal language theory with empirical insights from linguistic theory. Topics covered vary by year and may include tree transducers, logics for tree description, weak and strong generative capacity of natural language, lexicalized grammar formalisms, unification grammars, or the expressivity of probabilistic formalisms.

1.2 Full Description

The goal of this class is to develop a computational toolkit that is expressive enough to accommodate all of syntax without becoming computationally intractable. We will start out with n-gram models over strings and see that they fail this criterion even if probabilities are added to the model. From there we move up to (weighted) finite-state string automata, (probabilistic) context-free grammars, and so on, until we finally converge on first-order logic and restricted macro tree transducers as a lingua franca for syntax. While moving along this vertical axis of expressivity, we will also explore the range of options horizontally by looking at different encodings of syntactic dependencies, e.g. features vs constraints, slash-percolation vs movement, and phrase structure trees vs derivation trees. We will see that a computationally informed perspective provides the necessary degree of abstraction to develop a unified perspective of diverse grammar formalisms such as Minimalism, Government and Binding, GPSG, TAG, and Construction Grammar.

The course will benefit syntacticians who are curious about the connections between various formalisms (GPSG, CCG, MGs, TAG, linear LFG, Dependency Grammar, Relational Grammar) as well as natural language engineers that want to work with

Wk	*Topic*
1	Big picture, *n*-gram models
2	The insufficiency of finite-state string methods
3	Probabilities do not help
4	Dependencies and constituency as strictly local tree languages
5	Linear transductions from dependency trees to phrase structure trees
6	Adding AVMs without grammar blow-up
7	Constraints and their relation to subcategorization
8	(spring break)
9	Displacement with and without movement, multi bottom-up transductions
10	Extended tree transductions for idioms and CGx-style constructions
11	TAG-style adjunction, Late Merge, and their connection to lowering movement
12	Macro tree transducers and monadic second-order logic as a *lingua franca*
13	Weighted tree transductions and how to learn them
14	Towards weaker models
15	Summary, Q&A

Table 1: Tentative course outline

more sophisticated formalisms than context-free grammars. We will follow a lecture format with weekly homeworks and a large number of assigned readings; you should expect a larger time commitment than usual for a linguistics class. That said, proofs, theorems, and mathematics in general will be kept to a minimum, with an emphasis on hands-on modelling. Therefore the only hard prerequisite is a basic familiarity with syntax, comparable to what is covered in an undergraduate syntax class.

1.3 Prerequisites

This course presupposes basic familiarity with generative syntax as is usually acquired in *Syntax 1*. It is also useful to have some general mathematical maturity, e.g. by having taken *Mathematical Methods in Linguistics* or *Statistics*. Finally, it is recommended to have taken *Computational Linguistics 2* or to attend it in parallel with this course.

2 Learning Outcomes

- working knowledge of a variety of tree transducer types and their capabilities (top-down, bottom-up, linear, extended, multi bottom-up, macro tree transducer)
- ability to reinterpret linguistic concepts in computational terms
- assess linguistic phenomena from a computational perspective
- use computational concepts to identify new empirical generalizations
- formulate new syntactic analyses by combining existing computational techniques
- apply theory-heavy techniques to practical problems such as machine translation

3 Grading

- **Readings**
  Each week you have to prepare for the lecture by carefully going through the assigned readings. You should expect to spend about 3 hours a week on readings.
- **Homework**
  - weekly exercises
  - Selected exercises will be discussed in class.
  - Collaboration on homework problems is encouraged.
- **Final Paper**
  Write a computational syntax paper. You have to get my approval for your topic and your list of readings by the end of week 12. Possible topics include tree transducer implementations of specific pieces of syntactic machinery, computational analyses of empirical phenomena, a critical reply to an assigned paper, and much more (see the readings repository for other topics). Your paper must be written in LaTeX and adhere to the FG style requirements. Ideally, you will only write a final paper if you have a project that you want to present at a conference or turn into a journal paper.
- **Contributing to the Lecture Notes**
  I usually produce extensive lecture notes for my classes, but unfortunately this takes a lot of time that I simply do not have this semester. Instead, I will only produce rough drafts for the chapters and rely on you to flesh these out with nice prose, detailed examples, figures, tables, photos of researchers, jokes and so on. There will be a dedicated playground repository to allow the whole class to work on the lecture notes collaboratively.
- **Workload per Credits**
  - *1 credit*: regular attendance, readings, class participation
  - *2 credits*: the above, plus doing all the homeworks
  - *3 credits*: the above, plus writing a final paper or contributing to the lecture notes

4 Policies

4.1 Contacting me

- Emails should be sent to lin628@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 apart enough time and avoid collisions with other students.

4.2 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

4.3 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/

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 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.

5 Selected Bibliography

All readings materials are available for download in the readings repository of the course website.

5.1 Required

- Lecture notes, available at lin628.thomasgraf.net
- Müller, Stefan (in press): *Grammatical Theory: From Transformational Grammar to Constraint-Based Approaches*. Language Science Press

5.2 Supplementary

- Gécseg, Ferenc and Magnus Steinby (2015[2]): *Tree Automata*. http://arxiv.org/abs/1509.06233
- Comon, Hubert et al. (2008): *Tree Automata Techniques and Applications*. https://gforge.inria.fr/frs/download.php/file/10994/tata.pdf