Computational Phonology

Course	Info
Course number	lin626
Time	see Meeting Schedule below
Location	SBS N-250 (CompLab)
Website	lin626.thomasgraf.net
Instructor	Thomas Graf
Email	[coursenumber]@thomasgraf.net
Office hours	M 2:30-4:30, F 2:20-3:20
Office	SBS N-249

Attention: To get access to the readings repository, you must email me your github username.

Course Outline

Bulletin Description

An in-depth survey of natural language phonology from a computational perspective. Topics vary by year and may include formal language theory (subregular hierarchy, finite-state transductions), computational modeling (maximum entropy grammars, Hidden Markov Models), and machine learning.

Full Description

The seminar covers two distinct traditions in computational phonology: subregular phonology and finite-state optimality theory. The former will build on a recent book manuscript (Heinz et al 2016), whereas the latter draws directly from the primary literature (Frank & Satta 1998, Karttunen 1998, Jäger 2002, Riggle 2004).

The primary goal of the seminar is not to merely familiarize participants with a certain type of computational phonology. Rather, students will get to apply the computational perspective to their own research. Consequently, the emphasis will be on hands-on modelling and the use of computational tools for empirical analysis rather than mathematics and computational theory. Ideally, the work that students do in this seminar will result in presentations at computational conferences or workshops (e.g. NECPhon, Sigmorphon, MOL) and possibly even journal publications (Computational Linguistics, Transactions of the ACL, JoLLI, Journal of Language Modelling).

To reflect this goal, we will have two very different meetings per week. One (the *seminar session*) is run like a normal seminar and focuses on the critical discussion of assigned readings. The other one (*advisee session*) functions more like an advisee meeting, with

students reporting on their research progress, providing feedback on drafts of papers or abstracts, and so on.

Prerequisites

Students are expected to have some previous experience with phonology at the undergraduate or graduate level. For students who lack sufficient expertise in phonology, additional background readings will be made available. No prior mathematical or computational experience is required.

Meeting Schedule

I consider classes a dynamic process where activities and workload vary throughout the semester. This isn't easily accommodated by a rigorous fixed schedule, which is why I would like to propose a number of changes to the current template of MW 5:30–6:50.

At the beginning of the semester, it will be important to quickly bring you up to speed so that you have enough of a background to explore the primary literature on your own and start your quest for a research project. So for the first three weeks we will meet multiple times for a longer than usual period. After the first three weeks, the seminar time gets split between seminar sessions of 80 minutes (for everybody) and advisee sessions of 60 minutes (for students that are enrolled for at least 2 credits). The schedule *en detail*:

Day	Date	Time	Activity
M	Aug 29	5:30 - 6:50	organizational meeting
W	Aug 31	5:45 - 8:00	extended 2h seminar session with 10min break
F	Sep 2	3:00 - 5:15	extended 2h seminar session with 10min break
M	Sep 5		Labor day
W	Sep 7	5:45 - 8:00	extended 2h seminar session with 10min break
M	Sep 12		no class
W	Sep 14	5:45 - 8:00	extended 2h seminar session with 10min break
F	Sep 16	3:05 - 5:15	final extended 2h seminar session with 10min break
M	week 4+	1:15 - 2:15	60min advisee session
		5:30 - 6:50	regular 80min seminar session
M	Nov 21	no class	early Thanksgiving break

Given a 15 week semester with one week of vacation (Labor day + Thanksgiving), this adds up to `80+4*125+120+11*(60+80) = 2240`, which is exactly the official amount of `14*2*80 = 2240`.

I also intend to organize a one-day workshop during finals week where students present the results of the research they did in this seminar. The workshop will be announced on LINGO-L, and there might even be one or two invited speakers.

Teaching Goals

- master the primary literature in computational phonology
- understand technical concepts from formal language theory (subregular languages and transductions)
- apply computational ideas to the analysis of empirical data
- assess and characterize the typological diversity and unity of natural language phonology
- relate the formal complexity of phonological patterns to issues of language learning
- develop essential skills for academic work
  - critical analysis of primary literature
  - giving colleagues constructive feedback
  - writing peer reviews, abstracts, and short papers
  - conference presentations

Grading

Overview: Workload per Credits

Credits	Duties
0	attendance of seminar session
1	do the readings
	moderate discussion of a paper, with short handout
	review of paper
2	attendance of advisee session
	discussing your work in advisee session
	two (2) page abstract for mainstream linguistics conference
3	final paper, ACL style
	final presenation at workshop

Readings and Discussion

Each week will have one or two readings assigned (check the readings repository!), which are discussed in the seminar session. The expected load is about 25 to 50 pages per week, depending on the difficulty of the material. Each seminar session will have a student act as the moderator of the discussion. This requires

- preparing a short handout with the essential talking points, and
- walking seminar participants through the reading, and
- asking questions to test comprehension of the material, and
- bringing unclear points to everybody's attention, and
- critically evaluating the assigned reading.

The handout will be shared in the readings repository. Ideally, it should be written in markdown or LaTeX, but a doc file is also acceptable if absolutely necessary. Both the source file and a pdf must be emailed to me, I will then upload them to the readings

repository.

Paper Review

A central (and sometimes irritating) part of academic life is reviewing papers for conferences and journals. Review writing has some unique challenges because it is rarely discussed in grad school, there are few opportunities to read reviews, and expected review formats and criteria differ widely between subfields.

To give you some initial experience with the peer-reviewing process, you must write a peer-review of the paper you were the moderator for. Like the handout, both the source file and the pdf will be made available in the readings repository. The length of a review depends greatly on the quality of the paper, but unless you are dealing with truly abysmal work that deserves a good riffing, around 4 pages is usually enough. However, if you decide to be a kind reviewer that points out every minor typo, the review may easily hit 10 pages. How detailed you want to be is up to you, this is an opportunity for you to develop your own reviewing style: meticulous and detailed, or focussed on the big picture? You should follow the general template for peer reviews:

1. One paragraph summary of work to demonstrate that you understood it and are thus qualified to review it
2. Overall evaluation of work in 2 paragraphs, with a clear recommendation to the editor:
   - reject
   - revise and resubmit
   - accept with major revisions
   - accept with minor revisions
   - accept as is

   You should also give a justification for your recommendation (novelty of work, soundness and relevance of results, and so on).
3. 1 page discussion of major shortcomings (in presentation, evidence or argumentation) as well as suggestions for improvement
4. optionally a final section with page by page discussion of minor points

Reviews are due 2 weeks after the discussion of the paper (that's slightly faster than the usual turn-around of 4 weeks for computational linguistics conferences; journals, in particular in theoretical linguistics, usually give you about 2 months).

Caution: For manuscripts I will make the reviews available to the authors so they can revise the current draft. For this reason, you should put serious effort into your review. Also, you do not need to put your name on the review if you wish to stay anonymous.

Research Project: Abstract, Paper, and Presentation

One major goal of this seminar is to get you started on a computational research project of your own. It is my hope that none of the work you do for this seminar will be a wasted effort that ends up shelved away in some drawer without any public exposure. To maximize the chances that your work will result in a poster, talk, or maybe even a

peer-reviewed publication, the output you have to produce for this seminar is deliberately chosen to be suitable for immediate submission. Once you have decided on a research project, you will have to produce

1. a 2 page abstract (2 credits),
2. an ACL-style conference paper (3 credits),
3. a conference presentation for our final workshop (3 credits).

The abstract should follow the usual guidelines for linguistics conferences:

- at most 2 pages, including figures and references
- A4 paper with 2.5cm margins or letter paper with 1in margins (A4 gives you slightly more space)
- 12pt Times New Roman
- examples interspersed with text, not collected at end

For the paper, you have the choice between a short paper or a long paper, which is a common distinction at computational conferences like the ACL. Short papers are similar to squibs, whereas long papers are closer to a full research paper (but still very short for linguistic standards). In the words of the ACL:

> Please note that a short paper is not a shortened long paper. Instead short papers should have a point that can be made in a few pages. Some kinds of short papers are:
>
> A small, focused contribution
>
> Work in progress
>
> A negative result
>
> An opinion piece
>
> An interesting application nugget

Short papers are limited to four (4) pages excluding references, long papers to eight (8). For formatting guidelines (and a Latex template that takes care of everything for you), see the official ACL style files. It is important that your papers conform to these requirements because computational conferences expect camera-ready submissions that require no further typesetting.

Students that write a paper must present it at the seminar workshop, which will be held at the end of the semester (date still to be determined). Presentations for long papers will be 20+10 minutes, whereas we may experiment with different formats for short papers.

Timeline for Requirements

Requirement	Deadline
paper discussion	depends on your choice of paper
paper review	2 weeks after discussion
abstract	asap, but week 10 at the latest

Requirement	Deadline
paper	week 15
presentation	some day of finals week

Policies

Contacting me

- Emails should be sent to [coursenumber]@thomasgraf.net. 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.

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

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/

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.

Link List

Using git

- Github app for Windows; supports only Windows 7 or later
- Github app for Mac; supports only OS X 10.9 or later
- List of alternative GUI clients for git
- Tutorials for using git via the command line
- Official documentation for git

Markdown

- Interactive tutorial to markdown basics
- Complete markdown syntax
- Overview of Github's markdown dialect

LaTeX

- Overleaf (formerly writeLaTeX) is an online LaTeX editor with live preview
- List of commonly used math symbols
- Andrew Roberts' Getting to Grips with LaTeX