University of Toronto / Department of Computer Science

CSC 2528, Fall 2018
Advanced Computational Linguistics

Meetings

Thursdays, 13:00–15:00, XX NNNN.

Web page

`http://www.cs.toronto.edu/~gh/2528/`

Instructor

Graeme Hirst:
gh@cs.toronto.edu; 416-978-8747; Room 283A, D.L. Pratt Building, 6 King's College Road.

Format

Advanced Computational Linguistics is a seminar-style discussion course. The class will meet weekly for presentations and discussions of recent research in computational linguistics. Participants will take it in turns to act as presenters.

There will be two presenters each week, who will usually each present a short (conference-length) paper on a single topic or together present one long (journal-length) paper. The presenters will act, in effect, as representatives or surrogates of the authors of the papers; they will present the research in a seminar-like talk, putting the work in its best light. The presentation will be followed by questions and discussion from the other participants. All participants are expected to prepare by reading the day's papers in advance of the class.

In most cases, when two papers are to be discussed, the presentation of each will be about 35–40 minutes, followed by 10–15 minutes of questions and discussion. For a single long paper, an arrangement appropriate to its structure would be made.

A draft of the presentation itself (slides, notes, etc) must be discussed with the instructor at least two days before the class. The presentation should be fully rehearsed before the class meeting.

Taking the course for credit

If you are taking CSC2528 for credit, your grade will be based on your best two presentations (25% each) and on a term paper on a topic relevant to the course (50%). One of your presentations, near the end of the term, will be on a paper of your choice relevant to your term paper.

How to prepare to present a paper in class

1. Read the paper carefully (see below).

2. Plan how to present the material. While you have to be faithful to the material that you are presenting, you are not obliged to present it the same way that the paper does. Perhaps you can see a better way to do it—improved organization, different examples, different placement of emphasis, an adaptation to this audience. (The author themselves would probably not just give a literal rendition of the paper in a live presentation.)

3. Prepare slides in Keynote or LaTeX with a slide macro package (or PowerPoint, if you really must). Simultaneously, develop notes on what you want to say with each slide.

4. Discuss the draft of your presentation with the instructor by the Tuesday preceding your talk or earlier. (Default time: 10:00 Tuesday.) Amend and polish it as necessary.

5. Rehearse your presentation privately. Make sure it's not too long or too short, and be sure that you know just how you'll present each point. Refine your presentation as necessary.

6. You may present the slides on your own laptop. Or email the final version (as a Keynote, PDF, or PowerPoint file) to the instructor at least two hours before class. Also, you may wish to prepare a paper copy with additional annotations as necessary for your own use during the presentation.

How to read a research paper

Reading a research paper carefully takes time. And you'll probably need to go through it at least twice. Take notes as you read, including notes on what the paper is about and notes on questions that arise and your thoughts about the paper. When you have finished, write up a summary of the paper (both the content and your opinions) in a form that will still mean something to you in a year's time.

Questions to think about as you read:

Classifying the paper and its context
- What kind of paper is this?
  - A stand-alone work or one of a series
  - Summary of a thesis
  - Summary of a large project
  - Completed work, work in progress, or incremental report
- What kind of research is this?
  - Analysis of a problem or phenomenon
  - Model of some aspect of language processing
  - Technique for some aspect of language processing
  - New synthesis of previous ideas
  - Analysis of properties of a system or formalism

- To whom is the paper addressed?
- Whom does the author reference? Whom does the author *fail* to reference? Whom does the author acknowledge?

The research
- What is the author trying to do in their research? What are they trying to do in this paper?
- What has been done that hasn't been done before? Is it really new, or is it just a small (but possibly useful) variant of something else?
- Why is this good or useful?
- Does the paper state precisely what the problem is that is being solved? Does it do so right up front, or at least early on?
- If the paper describes an algorithm or a system, does it give examples of the input and output?
- Is there enough information for other researchers to duplicate the results?
- How could this new knowledge or technique be used? What problems could it solve or help to solve?
- How general is the approach? Does it work on anything but the examples in the paper? Under what circumstances would it fail?
- What hasn't been done? What is still needed for a complete solution?
- What other research is similar to this? What are the similarities and differences?
- What are the strengths of this research? What are its weaknesses?

The paper
- Does the author know what her or she is talking about?
- Is the author using different words for familiar concepts or familiar words for different concepts?
- Is the paper well written and as easy to follow as it could be?
- Are the examples helpful?

Personal consequences
- What does this paper mean for my own work?

Term paper

The term paper should be a critical review of the research literature on some problem, approach, or technique in computational linguistics. It need not include any novel research of your own (but it may). The topic may be quite narrow (defined by perhaps six relevant papers) or somewhat broader (up to 20 relevant papers), but shouldn't be too unfocused.

Typically, you'll first choose a topic area in which to start reading, and will narrow in on a precise topic as the reading progresses. The topic area must be submitted for approval at (or before) the third class meeting, to make sure that it is appropriate and feasible. The instructor might suggest some starting points for reading, but finding relevant literature is part of the task.

A good paper will include a description of the problem, a thoughtful, novel synthesis of the research literature, analysis of the current work with discussion of its strong and weak points, and suggestions for the directions in which research should proceed. The broader the topic, the more

important synthesis is; the narrower the topic, the more important analysis is. The body of the paper should be around 5000 words long (about 20 pages). A paper shorter than that suggests insufficient work; a longer paper suggests an inability to synthesize.

The paper should have a title page and abstract, and should be typeset in 12pt type, with line-spacing of 1.5 or more and margins of at least 25mm. Citations in the text should be in the name–year style ("Harvard style"). Clarity, neatness, and linguistic well-formedness all count towards the grade of the paper.

You must submit a "half-way report" in the ninth week of term in order to get feedback on your progress. A half-way report will normally include the precise topic of the paper, a list of all the papers that have been read, a draft of the introductory section, and an outline of the remainder of the paper.

Schedule

Thursday 27 September, 13:00: Submit choice of topic area.
Thursday 8 November, 13:00: Submit half-way report.
Monday 7 January, 13:00: Submit term paper.†
†Improve your holiday break: submit your paper by Friday 21 December.

Some current topics in computational linguistics

For your term paper, a "topic" is operationally defined as a fairly narrow research question or goal that is being addressed by more than a single group of researchers. A topic could be (or might have been) the subject of an ACL workshop. You are not limited to topics on this list. You can also browse through the computational linguistics literature for ideas. You should not choose a paper that has already been discussed recently in this or another CL graduate course.

- Bilingual and multilingual corpora: alignment, applications.
- Text summarization and related applications: current methods and issues.
- Text classification.
- Information retrieval, question-answering, and information extraction: current methods and issues.
- What is context, and how can it be formalized and used in computational language understanding?
- Machine translation: Current methods and issues.
- Coreference and anaphor resolution: knowledge-poor and knowledge-rich methods; Centering Theory.
- Issues in discourse structure and cohesion. Determining rhetorical relations between sentences.
- Viewpoint and subjectivity in text: separating opinions from facts in news text.
- Statistical and hybrid statistical–linguistic methods of parsing.
- Automated acquisition of lexical knowledge, both semantic and syntactic.

- Computational models of human sentence processing, including: psychologically plausible methods of parsing; models of ambiguity resolution; models of difficulties in processing; and models of lexical access.
- Computational models of human language acquisition.
- Surface realization in natural language generation; lexical choice.
- Stochastic text generation.
- Text-to-speech systems; learning letter-to-sound rules and prosody rules.
- Applying machine learning to discourse processing; models of dialogue and conversation; instructional dialogues.
- Recovery from misunderstanding; detection of user misunderstanding and misconception.

Where to find the CL research literature

Much or most of the important CL literature is published in the journals *Computational Linguistics* and *Transactions of the ACL*, the conferences and workshops of the Association for Computational Linguistics (ACL), and the International Conference on Computational Linguistics (COLING). These are all archived at *aclanthology.info*. Most other relevant journals require paid-subscription access through the UofT Library website.

Journals:
  Computational Linguistics
  Transactions of the Association for Computational Linguistics
  Computer Speech and Language
  Language Resources and Evaluation
  Digital Studies in the Humanities
  Machine Translation
  Natural Language Engineering

Conference proceedings:
  Association for Computational Linguistics:
    — annual meetings (ACL)
    — European meetings (EACL)
    — North American meetings (NAACL, HLT/NAACL)
    — Empirical Methods in Natural Language Processing (EMNLP)
    — Workshops at any of these conferences
  International Conference on Computational Linguistics (COLING)
  International Joint Conference on Natural Language Processing (IJCNLP)