CSC2611 (W2019): Computational Models of Semantic Change

Date/Time: Thursday, 10am-12pm

Location: BA1230

Instructor: Yang Xu
Contact: yangxu@cs.toronto.edu
Office Hours: By appointment

This syllabus may be adjusted as the course progresses.

Course Description: Words are fundamental components of human language, but their meanings tend to change over time, e.g., *face* ('body part →'facial expression), *gay* ('happy→'homosexual), *mouse* ('rodent →'device). Changes like these present challenges for computers to learn accurate representations of word meanings—a task that is crucial for natural language systems. This course explores data-driven computational approaches to word meaning representation and semantic change. Topics include latent models of word meaning (e.g., LSA, word2vec), corpus-based detection of semantic change, probabilistic diachronic models of word meaning, and cognitive mechanisms of word sense extension (e.g., chaining, metaphor). The course involves a strong hands-on component that focuses on large-scale text analyses and seminar-style presentations.

Note: This graduate course presumes extensive knowledge of Python programming and big data analytics. Undergraduates who are interested in enrolling should obtain special permissions from the instructor. Preferred preparatory courses include CSC108, CSC148, COG260, COG403, and courses in computational linguistics and natural language processing.

Objectives: This course is aimed at the following three objectives.

1. Develop a broad foundation for the interdisciplinary study on semantic change.
2. Develop technical skills in the computational analysis of longitudinal textual data.
3. Develop essential communicative skills in scientific presentation and writing.

Recommended background readings:

- Traugott, E.C., & Dasher, R.B. *Regularity in semantic change*. CUP. 2001.
- Sweetser, E. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. CUP. 1991.
- Hopper, P.J., & Traugott, E.C. *Grammaticalization*. CUP. 2003.
- Lakoff, G. *Women, fire, and dangerous things: What categories reveal about the mind*. UCP. 1987.

Deliverables and Assessments:

Paper presentation	25%
Lab assignment	15%
Project proposal	10%
Project milestones	5%
Project final report	20%
Project final presentation	10%
Code repository	15%

Letter Grade Scale:

90 - 100%	A+	77 - 79%	B+
85 - 89%	A	73 - 76%	B
80 - 84%	A-	70 - 72%	B-
		0 - 69%	Fail

Course Policies:

- **General**
  - Students are expected to present and lead discussion on at least 1 technical paper.
  - Students with scheduled presentations are required to send the PDF slides to the instructor two days before the presentations.
  - Late submissions will receive a 1 point deduction per delayed hour until no point can be further deducted.
- **Attendance**
  - Attendance is expected in general and required on days of presentation.
  - Students are responsible for all missed assignments due to absence, unless they notify the instructor at least two days prior to the due date.
- **Project**
  - Students are expected to work independently on projects.
  - Students may obtain the instructor's permission to work on their own research projects, provided that the projects are relevant to the course.
  - Students may proceed with their projects only if the initial proposals have been approved by the instructor. Otherwise they may do so until the revised proposals have been approved.

Schedule (see course webpage for reading/project/presentation information):

Date	Content
Jan 10	• Overview
Jan 17	• Distributed representations of word meaning • Lab assignment
Jan 24	• Automatic detection of semantic change • Project announcement (lab assignment due)
Jan 31	• General laws of semantic change
Feb 7	• Probabilistic models of semantic change • Project proposal due
Feb 14	• Novel word sense identification
Feb 28	• Cognitive mechanisms of word sense extension • Project milestone 1
Mar 7	• Research topic 1: Children's overextension
Mar 14	• Research topic 2: Cross-linguistic polysemy • Project milestone 2
Mar 21	• Research topic 3: Language evolution
Mar 28	• Research topic 4: Lexical dark matters • Project final report due
Apr 4	• Project final presentation

Resources:

- Python:

  Jupyter: https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/

  Natural Language Processing with Python: http://www.nltk.org/book/

  Natural Language Toolkit: http://www.nltk.org/

  Bare essentials: http://www.cs.toronto.edu/~yangxu/PythonBookletV4.pdf

- GitHub:

  Creating a repo: https://help.github.com/articles/create-a-repo/

  Common commands: https://gist.github.com/jedmao/5053440

- Word embeddings:

  Word2vec: https://code.google.com/archive/p/word2vec/

  GLOVE: https://nlp.stanford.edu/projects/glove/

  Lda2vec: https://github.com/cemoody/lda2vec

  tSNE: https://github.com/paulorauber/thesne

  HistWords: https://nlp.stanford.edu/projects/histwords/

- Longitudinal text corpora:

  Project Gutenberg: https://www.gutenberg.org/

  Google N-grams: http://storage.googleapis.com/books/ngrams/books/datasetsv2.html

  Syntactic N-grams: http://commondatastorage.googleapis.com/books/syntactic-ngrams/index.html

  Helsinki Corpus of English: http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/

  Early English Books Online: https://corpus.byu.edu/eebo/

  CHILDES: https://childes.talkbank.org/

- Lexical resources:

  WordNet: https://wordnet.princeton.edu/

  MetaNet: https://metanet.icsi.berkeley.edu/metanet/

  Metaphor Map of English: http://mappingmetaphor.arts.gla.ac.uk/

Historical Thesaurus of English: http://historicalthesaurus.arts.gla.ac.uk/

Dictionary of Old English: https://www.doe.utoronto.ca/pages/index.html

- Benchmark data:

  WordSimilarity-353: http://www.cs.technion.ac.il/~gabr/resources/data/wordsim353/

  SimLex-999: https://www.cl.cam.ac.uk/~fh295/simlex.html

  SemEval-2017: http://alt.qcri.org/semeval2017/index.php?id=tasks

  Stanford Question Answering: https://rajpurkar.github.io/SQuAD-explorer/

- Human behavioural data:

  University of South Florida Free Association Norms: http://w3.usf.edu/FreeAssociation/

  Human Brain Cloud: http://www.humanbraincloud.com/

  Word concreteness ratings: http://crr.ugent.be/archives/1330

  Word affectiveness ratings: http://crr.ugent.be/archives/1003

  Word age-of-acquisition norms: http://crr.ugent.be/archives/806