Course Syllabus
COMPSCI 685 (a.k.a. 690N), Spring 2018
http://people.cs.umass.edu/~brenocon/anlp2018/

Title: Advanced Natural Language Processing
Instructor: Brendan O'Connor

Description: This course covers a broad range of advanced level topics in natural language processing. It is intended for graduate students in computer science who have familiarity with machine learning fundamentals, but may be new to NLP. It may also be appropriate for computationally sophisticated students in linguistics and related areas. Topics include probabilistic models of language, computationally tractable linguistic representations for syntax and semantics, and selected topics in discourse and text mining. After completing the course, students should be able to read and evaluate current NLP research papers. Coursework includes homework assignments and a final project.

Rough Schedule:

Non-structured language models	Week 1	Language Models, EM.
	Week 2	Log-linear Models.
	Week 3	Neural Models.
	Week 4	Distributional semantics and word embeddings.
Structured linguistic models	Week 5	Sequence models: Inference.
	Week 6	Sequence models: Learning.
	Week 7	Syntax: PCFGs.
	Week 8	Syntax: Dependencies.
	Week 9	Semantics.
Discourse and documents	Week 10	Coreference.
	Week 11	Non-structured document models (e.g. topic models).
	Week 12	Contexted document models. Social networks, geolocation, political science.
	Week 13	
	Week 14	Project presentations.

Prerequisites: Familiarity with multivariate calculus, probability theory, dynamic programming, and implementation of machine learning algorithms, such as from COMPSCI 585, 688, 689, STAT 697ML, or equivalent. Previous experience in linguistics or natural language processing is suggested, but is not required.

Readings: There is no textbook. Readings will be provided. These will include research papers, and selections from:

- Eisenstein 2018, "Natural Language Processing" (draft text)
- Bender 2013, "Linguistic Fundamentals for Natural Language Processing"
- Goldberg 2017, "Neural Network Methods for Natural Language Processing"
- Manning and Schütze 1999, "Foundations of Statistical Natural Language Processing"
- Jurafsky and Martin, "Speech and Language Processing" (draft text)

Coursework and grading:

- Homework assignments (30%). These include mathematical/written questions, as well as programming assignments to implement and analyze NLP methods.
- In-class paper presentation (10%). Present a review of a research paper.
- Short, approximately weekly paper reviews (10%). One paragraph briefly summarizing and reacting to an assigned paper.
- Literature review (25%).
- Final project (25%).

The **literature review** is a paper that reviews a subfield of NLP of your choice. It must cover at least 15 resesarch papers, including at least 5 from before the year 2000. It should not merely describe the papers, but also synthesize, organize, and relate them to one another and the broader literature in NLP, and ideally also ML and linguistics. It can be done either individually or in a group of two. An excellent example is Das and Martins (2007), "A Survey on Automatic Text Summarization" (which itself was a class assignment, now highly cited!).

The **final project** can be done in a group of size one to three, and includes the components:

- Proposal: a 1-3 page document outlining the problem, your approach, possible datasets and/or software to use.
- Presentations: short in-class presentations.
- Final Report: an 8-20 page document describing the entire project, results and discussion, as well as all the components of a publishable research contribution: introduction, motivation, and related work.

Other UMass NLP Courses: Computer science also offers COMPSCI 585, "Introduction to Natural Language Processing," which is appropriate for advanced undergraduates and masters students. A list of related courses is at: http://people.cs.umass.edu/~brenocon/complang_at_umass/

Should I take both 585 and 685? Probably not – there's some overlap in topics. Please ask us if you have questions.

Accommodation:
The University of Massachusetts Amherst is committed to providing an equal educational opportunity for all students. If you have a documented physical, psychological, or learning disability on file with Disability Services (DS), you may be eligible for reasonable academic accommodations to help you succeed in this course. If you have a documented disability that requires an accommodation, please notify me within the first two weeks of the semester so that we may make appropriate arrangements.

Academic Honesty:
Since the integrity of the academic enterprise of any institution of higher education requires honesty in scholarship and research, academic honesty is required of all students at the University of Massachusetts Amherst. Academic dishonesty is prohibited in all programs of the University. Academic dishonesty includes but is not limited to: cheating, fabrication, plagiarism, and facilitating dishonesty. Appropriate sanctions may be imposed on any student who has committed an act of academic dishonesty. Instructors should take reasonable steps to address academic misconduct. Any person who has reason to believe that a student has committed academic dishonesty should bring such information to the attention of the appropriate course instructor as soon as possible. Instances of academic dishonesty not related to a specific course should be brought to the attention of the appropriate department Head or Chair. Since students are expected to be familiar with this policy and the commonly accepted standards of academic integrity, ignorance of such standards is not normally sufficient evidence of lack of intent (http://www.umass.edu/dean_students/codeofconduct/acadhonesty/).