Course Syllabus (DRAFT – subject to change)
COMPSCI 690N, Spring 2017

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.

Schedule:

Non-structured language models	Week 1	Language Models. Information theory. Multinomials, EM. *[Assignment: Pereira 2001.]*
	Week 2	Machine Translation: IBM Models, noisy channel.
	Week 3	Log-linear Models. Loglin Model 2, LMs. *[Assignment: SGD LM/MT.]*
	Week 4	Neural Models.
Structured (linguistic) analysis	Week 5	Rules. Information extraction, shallow parsing.
	Week 6	Syntax. PCFGs. *In-class activity: competitive grammar writing. [Project proposal.]*
	Week 7	Structured Prediction. Parameter learning, sequence tagging. *[Assignment: StructSVMs.]*
	Week 8	Syntax. Dependencies. NN parsing.
	Week 9	Semantics. Argument realization, Davidsonian representations, relation extraction. *[Assignment: heuristic semantics.]* *[Project milestone report.]*
Discourse and documents	Week 10	Coreference.
	Week 11	Non-structured document models. Topic models, log-linear BOW. *[Assignment: paper summary]*
	Week 12	Contexted document models. Social networks, geolocation, political science.
	Week 13	Project presentations. *[Project final report due at end of finals.]*

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 may be helpful, but is not required.

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

* Eisenstein 2016, "A Technical Introduction to Natural Language Processing"
* Bender 2013, "Linguistic Fundamentals for Natural Language Processing"
* Goldberg 2015, "A Primer on Neural Network Models for Natural Language Processing"
* Manning and Schütze 1999, "Foundations of Statistical Natural Language Processing"
* Jurafsky and Martin, "Speech and Language Processing"

Coursework and grading:

* Homework assignments (40%). These include mathematical/written questions, as well as programming assignments to implement and analyze NLP methods.
* Final project (50%), including proposal, milestone report, presentation, and final report. Projects are done in groups of size one to three.
* Course participation (10%) in classroom or online discussions, including short in-class assignments.

The **final project** includes the components:

* Proposal: a 2-4 page document outlining the problem, your approach, possible datasets and/or software to use.
* Progress Report: a 5-10 page document describing preliminary work.
* Presentations: short in-class presentations.
* Final Report: a 12-20 page document describing the entire project and final results, as well as a related work section.

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 690N? We are recommending against it for most students. First, there is some overlap in topics, so time may be wasted if you already took 585. Second, there is considerably more advanced machine learning, which 585 does not entirely prepare students for. Third, space in the course is reserved

for PhD students (since 585 already exists), so we do not know if there will be space to accommodate all interested students. All that said, there may be some cases where taking both 585 and 690N makes sense; please ask us if you have questions.

Accommodation Statement:
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 Statement:
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/).