



## LING/CSC 439/539: Statistical Natural Language Processing

Communication #113, Tue/Thu 2:00 – 3:15

*Last modified: August 21, 2017*

### Description of Course

This course focuses on building statistical models of natural language. We do this with two aims. First, these models have tremendous value in the practical/computational domain and are widely used in human language technology applications. Second, these models have significant appeal as theoretical models of how language is processed, or how grammars are organized. This is a highly interdisciplinary course, bringing together elements of both linguistics and computer science. Natural Language Processing (NLP) has a large applied component, and as such this course will have a considerable focus on project-based assignments rather than written ones.

### Course Prerequisites or Co-requisites

The students taking this course must know how to program, and have a decent understanding of data structures such as hash maps and trees. Ideally, the students should have taken a calculus course. We will, however, cover the necessary math background in class.

- Prerequisites: **Ling 438/538**, or **CSC 483/583**.
- Recommended: **Math 129 (Calc II)**

**Programming:** Programming skills are required for this course. We will be using Python 2. Students unfamiliar with Python must have a working Python 2.x environment up and running and read through Chapter 1 in Natural Language Processing with Python (see below) within the first week.

### Instructor and Contact Information

Instructor: Mihai Surdeanu

Email: [msurdeanu@email.arizona.edu](mailto:msurdeanu@email.arizona.edu)

Web: <http://surdeanu.info/mihai>

Office: Gould-Simpson 746

Office hours: Tue 12:30 – 2

Teaching assistant: Gustave Hahn-Powell

Email: [hahnpowell@email.arizona.edu](mailto:hahnpowell@email.arizona.edu)

Office: Gould-Simpson 903

Office hours: Wed 2 – 3

Teaching assistant: Patricia Lee

Email: [pllee@email.arizona.edu](mailto:pllee@email.arizona.edu)

Office: TBA

Office hours: TBA

## Course Format and Teaching Methods

The course will be delivered using in-person lectures. No lab sections will be offered but the instructor encourages additional discussion on the topics introduced in the lecture materials. These discussions will be managed on a Piazza site controlled by the instructor.

The Piazza site is available here: <https://piazza.com/arizona/fall2017/ling439539/home>

## Course Objectives and Expected Learning Outcomes

At the conclusion of this course students should understand fundamental statistical methods for the processing of natural language, including: (a) text classification, (b) sequence modeling and its applications to part-of-speech tagging, (c) algorithms for structured learning such as shift-reduce and applications to syntactic parsing, and (d) cross-lingual and mono-lingual alignment algorithms such as IBM Model 1 and their applications to machine translation and question answering.

Graduate students are expected to have an in-depth understanding of these techniques. For example, graduate students are expected to know how to code the underlying machine learning framework necessary for text classification such as logistic regression.

## Absence and Class Participation Policy

UA's policy concerning Class Attendance, Participation, and Administrative Drops is available at <http://catalog.arizona.edu/policy/class-attendance-participation-and-administrative-drop>

The UA policy regarding absences for any sincerely held religious belief, observance or practice will be accommodated where reasonable: <http://policy.arizona.edu/human-resources/religious-accommodation-policy>.

Absences preapproved by the UA Dean of Students (or dean's designee) will be honored. See <https://deanofstudents.arizona.edu/absences>

Participating in the course and attending lectures and other course events are vital to the learning process. As such, attendance is required at all lectures and discussion section meetings. Students who miss class due to illness or emergency are required to bring documentation from their health-care provider or other relevant, professional third parties. Failure to submit third-party documentation will result in unexcused absences.

## Course Communications

Please use the email addresses above to contact the instructor or the TA. All course materials will be posted in D2L. Please use the Piazza site above to ask clarification questions about the material.

## Required Texts or Readings

This course follows the following textbook:

- Christopher D. Manning and Hinrich Schütze. 1999. Foundations of Statistical Natural Language Processing. 6th printing with corrections, 2003. The MIT Press. <http://nlp.stanford.edu/fsnlp/> (available for free electronically through UA library)
- Additional research articles covered in class will be distributed by the instructor.

**Highly recommended:** For students not comfortable with natural language processing in Python, a companion reference such as the following is also highly recommended:

- Steven Bird, Ewan Klein, and Edward Loper. 2009. Natural Language Processing with Python. <http://nltk.org/book/> (currently available for free electronically through the NLTK website)

## Required or Special Materials

No special tools or supplies needed.

## Assignments and Examinations: Schedule/Due Dates

Grading will be based on four assignments, two exams (midterm and final), a programming project, and overall in-class participation. Please note that all four assignments will have a considerable programming component. For the final programming project, the students may propose a NLP topic that interests them, or implement one of the topics suggested by the instructor.

As a rule, work will not be accepted late except in case of documented emergency or illness. You may petition the professor in writing for an exception if you feel you have a compelling reason for turning work in late.

The due dates are as follows:

<b>Task</b>	<b>Deadline</b>
<b>HW 1</b>	August 27
<b>HW 2</b>	September 24
Midterm review	October 10
<b>Midterm</b>	October 12
<b>HW 3</b>	October 29
<b>HW 4</b>	November 26
Final review	December 5
<b>Project</b>	December 7

This course will have a comprehensive written final examination. Information on the final exam regulations and schedule:

<https://www.registrar.arizona.edu/courses/final-examination-regulations-and-information>

<http://www.registrar.arizona.edu/schedules/finals.htm>

## Grading Scale and Policies

The grading scheme is as follows:

<b>Component</b>	<b>Weight</b>
<b>Assignments</b>	300 pts
<b>Midterm exam</b>	200 pts
<b>Final exam</b>	275 pts
<b>Programming project</b>	200 pts
<b>In-class participation</b>	25 pts
<b>Total</b>	<b>1000 pts</b>

<b>Grade</b>	<b>Point Range</b>
<b>A</b>	900 – 1000
<b>B</b>	800 – 899
<b>C</b>	700 – 799

<b>D</b>	600 – 699
<b>E</b>	0 – 599

### Undergraduate vs. Graduate Requirements

This course will be co-convened. To differentiate between graduate and undergraduate students, the instructor will require graduate students to implement more complex algorithms for the programming project, which might require additional reading of research articles. The instructor will provide the additional reading material and will guide the research process. Similarly, assignments and exams will have additional requirements/questions for graduate students. The overall grading scheme will be the same between graduate and undergraduate students (see the two tables above).

**Requests for incomplete (I) or withdrawal (W)** must be made in accordance with University policies, which are available at <http://catalog.arizona.edu/policy/grades-and-grading-system#incomplete> and <http://catalog.arizona.edu/policy/grades-and-grading-system#Withdrawal>, respectively.

### Scheduled Topics/Activities

The course will cover the topics listed below ("MS x" indicates the corresponding chapter in the Manning/Schutze textbook):

<b>Week</b>	<b>Topics</b>	<b>Readings</b>
<b>1</b>	Introduction, text categorization	MS 1, 16
<b>2</b>	Crash course in ML for text categorization: kNN, perceptron, logistic regression, feed forward neural networks	Materials provided by instructor
<b>3</b>	Crash course in ML, part 2	
<b>4</b>	Distributional similarity: count-based methods, word embeddings	Materials provided by instructor
<b>5</b>	Probability theory	MS 2 + materials provided by instructor
<b>6</b>	Probability theory, part 2	
<b>7</b>	N-gram models	MS 5, 6 + parts of MS 7, 8
<b>8</b>	Midterm review and exam	
<b>9</b>	Sequence models: HMM, MEMM, LSTM, and applications to part-of-speech tagging and information extraction	MS 9, 10 + materials provided by instructor
<b>10</b>	Sequence models, part 2	
<b>11</b>	Structured learning: shift-reduce algorithms, PCFG, tree LSTM, and applications to syntactic parsing	MS 11, 12 + materials provided by instructor
<b>12</b>	Structured learning, part 2	
<b>13</b>	Alignment models and applications to machine translation and question answering	MS 13 + materials provided by instructor
<b>14</b>	Alignment models, part 2	
<b>15</b>	Advanced techniques: question answering, reading comprehension, summarization	Materials provided by instructor

### Classroom Behavior Policy

To foster a positive learning environment, students and instructors have a shared responsibility. We want a safe, welcoming, and inclusive environment where all of us feel comfortable with each other and where we can challenge ourselves to succeed. To that end, our focus is on the tasks at hand and not on extraneous activities (e.g., texting, chatting, reading a newspaper, making phone calls, web surfing, etc.).

Inclusive Excellence is a fundamental part of the University of Arizona's strategic plan and culture. As part of this initiative, the institution embraces and practices diversity and inclusiveness. These values are expected, respected and welcomed in this course.

Students are asked to refrain from disruptive conversations with people sitting around them during lecture. Students observed engaging in disruptive activity will be asked to cease this behavior. Those who continue to disrupt the class will be asked to leave lecture or discussion and may be reported to the Dean of Students.

Some learning styles are best served by using personal electronics, such as laptops and iPads. These devices can be distracting to other learners. Therefore, students who prefer to use electronic devices for note-taking during lecture should use one side of the classroom.

### **Threatening Behavior Policy**

The UA Threatening Behavior by Students Policy prohibits threats of physical harm to any member of the University community, including to oneself. See <http://policy.arizona.edu/education-and-student-affairs/threatening-behavior-students>.

### **Elective Name and Pronoun Usage**

This course supports elective gender pronoun use and self-identification; rosters indicating such choices will be updated throughout the semester, upon student request. As the course includes group work and in-class discussion, it is vitally important for us to create an educational environment of inclusion and mutual respect.

### **Accessibility and Accommodations**

Our goal in this classroom is that learning experiences be as accessible as possible. If you anticipate or experience physical or academic barriers based on disability, please let me know immediately so that we can discuss options. You are also welcome to contact the Disability Resource Center (520-621-3268) to establish reasonable accommodations. For additional information on the Disability Resource Center and reasonable accommodations, please visit <http://drc.arizona.edu>.

If you have reasonable accommodations, please plan to meet with me by appointment or during office hours to discuss accommodations and how my course requirements and activities may impact your ability to fully participate.

Please be aware that the accessible table and chairs in this room should remain available for students who find that standard classroom seating is not usable.

### **Code of Academic Integrity**

Students are encouraged to share intellectual views and discuss freely the principles and applications of course materials. However, graded work/exercises must be the product of independent effort unless otherwise instructed. Students are expected to adhere to the UA Code of Academic Integrity as described in the UA General Catalog. See <http://deanofstudents.arizona.edu/academic-integrity/students/academic-integrity>.

The University Libraries have some excellent tips for avoiding plagiarism, available at <http://www.library.arizona.edu/help/tutorials/plagiarism/index.html>.

Selling class notes and/or other course materials to other students or to a third party for resale is not permitted without the instructor's express written consent. Violations to this and other course rules are subject to the Code of Academic Integrity and may result in course sanctions. Additionally, students who use D2L or UA e-mail to sell or buy these copyrighted materials are subject to Code of Conduct Violations for misuse of student e-mail addresses. This conduct may also constitute copyright infringement.

### **UA Nondiscrimination and Anti-harassment Policy**

The University is committed to creating and maintaining an environment free of discrimination; see <http://policy.arizona.edu/human-resources/nondiscrimination-and-anti-harassment-policy>

Our classroom is a place where everyone is encouraged to express well-formed opinions and their reasons for those opinions. We also want to create a tolerant and open environment where such opinions can be expressed without resorting to bullying or discrimination of others.

### **Department of Computer Science Code of Conduct**

The Department of Computer Science is committed to providing and maintaining a supportive educational environment for all. We strive to be welcoming and inclusive, respect privacy and confidentiality, behave respectfully and courteously, and practice intellectual honesty. Disruptive behaviors (such as physical or emotional harassment, dismissive attitudes, and abuse of department resources) will not be tolerated. The complete Code of Conduct is available on our department web site. We expect that you will adhere to this code, as well as the UA Student Code of Conduct, while you are a member of this class.

### **Additional Resources for Students**

UA Academic policies and procedures are available at <http://catalog.arizona.edu/policies>

Student Assistance and Advocacy information is available at <http://deanofstudents.arizona.edu/student-assistance/students/student-assistance>

Office of Diversity information is available at <http://diversity.arizona.edu/>

Campus Health information may be found here: <http://www.health.arizona.edu/counseling-and-psych-services>

OASIS Sexual Assault and Trauma Services  
[http://oasis.health.arizona.edu/hpps\\_oasis\\_program.htm](http://oasis.health.arizona.edu/hpps_oasis_program.htm)

### **Subject to Change Statement**

Information contained in the course syllabus, other than the grade and absence policy, may be subject to change with advance notice, as deemed appropriate by the instructor.