

# LING 520: Computational Analysis of English

Semester: FALL '16

Instructor: Sowmya Vajjala

Iowa State University, USA

23 August 2016

# Class outline

- ▶ Introductions
- ▶ What is Natural Language Processing (NLP)?
- ▶ Course objectives and Pre-requisites
- ▶ Course Logistics
- ▶ Syllabus
- ▶ Pre-course questionnaire

# Introductions

# About me

1. In ISU as Asst. Professor since January 2016.
2. Prior to that, PhD student in Computational linguistics, in Germany.
3. Taught 516 in Spring Semester
4. Flashback: 10 fall semesters ago, in Fall 2006, I attended a similar introductory NLP course as a masters student in India!

# About you?

1. Name
2. What do you do in ISU?
3. What are your interests related to computational analysis of language?

# Motivation for the course

# What is NLP?

1. NLP is a sub-field of Artificial intelligence that is concerned with analyzing, modeling and understanding human language using computational methods.
2. It explores how humans can interact with computers in human languages
3. The eventual goal is to make computers understand (and generate) human languages, and make them communicate with humans like humans
4. Because of its role in the process of human-computer interaction, NLP has a wide range of technological applications
5. It is also becoming popular as a research method in a broad range of disciplines in social sciences.

# Inter-disciplinary by nature

NLP is very inter-disciplinary. Draws from research in Computer Science, Linguistics, Mathematics, Statistics, Psychology etc.,



# History of NLP

1. Foundational ideas: 40s and 50s. WWII and Beyond.
2. Main NLP problem of that time (and even now): Machine Translation
3. First few decades: Work focused on the development of speech recognition systems, logic based language understanding systems, creating elaborate grammars to teach human language to computers, and automatic language generation.
4. Late 90s on: Advent of statistical methods and machine learning

# Computational Linguistics vs NLP

The terms are used synonymously. However, generally, NLP is typically used by people involved in engineering and technology development, and CL is typically used by traditionally linguistics groups who adapted computational methods.

# Where is NLP used in real-world?

1. Apple Siri and other such software that can understand and interpret human speech (okay, partially)
2. Google Translate and the likes
3. Search Engines
4. Question Answering (e.g., IBM Watson)
5. News recommendation - related articles features in News websites
6. Sentiment analysis of product reviews on Amazon, for example
7. Spam classification in Gmail, Yahooemail etc
8. Information extraction from text (e.g., identifying calendar entries automatically in gmail)
9. Dialog systems (having interactive conversations with users, to do flight bookings etc)

# NLP and Applied Linguistics

## Some Practical Examples

What are some applications of NLP in Applied Linguistics?

# NLP and Applied Linguistics

## Some Practical Examples

What are some applications of NLP in Applied Linguistics?

1. Spelling and Grammar checking
2. Plagiarism detection
3. Automated content analysis and scoring of student responses
4. Automatic generation of tests
5. Intelligent tutoring systems

... and so on.

# Some EdTech companies that use NLP

- ▶ Turnitin, LightingGrader etc
- ▶ ETS, Pearson etc
- ▶ Grammarly.com
- ▶ DuoLingo
- ▶ Cognii.com

... and so on

# Some research projects relevant for language learning

1. FLAIR project @University of Tübingen
  2. LARKA CALL Platform @Gothenburg University
  3. Lexical Complexity Analyser @Penn State
  4. Research Writing Tutor @ Iowa State
- .. and so on

# Course Objectives and Pre-requisites



# Goals for the course

1. Understand why language processing is hard
2. Understand some key problems and methods in NLP
3. Understand the limitations of NLP methods
4. You will gain some knowledge about common text processing methods, and their applications
5. You will also know how to apply existing language processing tools to develop your own applications

# Pre-requisites

1. Knowledge of some programming language (Perl, Python etc.)
2. Some knowledge about linguistics, and an interest in computational analysis of human language
3. Ability to understand problem descriptions, and write programs by yourself, understand and debug the errors you get while running the program.
4. An understanding that this course is not a typical Applied Linguistics course and will need different kind of effort.
5. Acknowledgement that your instructor is aware of her job and knows what it takes to learn the topics.

# Course Logistics

# Meeting and Location

- ▶ meets in Ross 312, on Tuesdays and Thursdays from 11 am-12:20 pm
- ▶ *Office hours:* Tuesdays and thursdays, 10-11 am (please email beforehand if there are specific issues to discuss. If this time does not work for you, send an email, and we can meet at a convenient time)
- ▶ course website: on BlackBoard.

# Credits

- ▶ Credit Points: 3  
(Expect to spend more time than you think! It is a difficult, graduate level course and needs hard work!)

## Format and Grading

# Course Format

- ▶ weekly lectures/practical sessions
- ▶ optional discussion forum
- ▶ weekly readings, sometimes additional videos.
- ▶ 5 assignments
- ▶ final exam - programming project.
- ▶ optional practice exercises on different topics.

# Assignments

- ▶ 5 assignments (80%) + 1 final project (20%)
- ▶ Problem sets will be given periodically for practice



# Final Exam

- ▶ Programming project - 20 M
- ▶ Choose a project by the end of october
- ▶ Either pick something from the list I give or come up with your idea.
- ▶ Evaluation: Show a working program + oral exam.

## Some general rules:

- ▶ attendance: not mandatory, but recommended.
- ▶ missing a deadline: .. is okay, but you will not get full credit.
- ▶ long absence due to illness etc: please inform and follow university procedures.
- ▶ cheating and plagiarism: see the course handbook, and university policy against plagiarism.
- ▶ classroom behavior: please be punctual and do not do personal work in the class.
- ▶ Disability accomodation: Please speak to Disability Resources Office (DRO) to officially request an accomodation.

# Other Issues

- ▶ validating enrollment: who is enrolled? who is just here?
- ▶ feedback about the course:
  1. Talk to me directly, or leave anonymous feedback at:  
<https://goo.gl/qCKax9> or leave a paper feedback in my mailbox.
  2. Be confident enough to confront me and talk to me if there is a concern.
  3. You can talk behind my back or wait until course feedback, but you will not benefit from that.. next batches perhaps will.

# Syllabus

# Topics

- ▶ Introduction to NLP
- ▶ Programming recap and Practice
- ▶ Morphological analysis
- ▶ Language modeling, Part of speech tagging
- ▶ Text classification
- ▶ Natural language parsing
- ▶ Other NLP problems: overview
- ▶ NLP for CALL

(Each topic will have an ungraded problem set for practice)

# Assignment Deadlines

- ▶ Assignment 1: 10 Sep 2016
- ▶ Assignment 2: 27 Sep 2016
- ▶ Assignment 3: 11 Oct 2016
- ▶ Assignment 4: 29 Oct 2016
- ▶ Assignment 5: 15 Nov 2016
- ▶ Final project decision + Prelim report: 5 Nov 2016
- ▶ Final project - presentation: December 6-13 2016

(All assignments are already uploaded. Final project options are up too)

# Text Book

1. Primary textbook: Speech and Language Processing by Jurafsky & Martin (2nd Edition)
  - ▶ Not mandatory, but very useful book in a long run, if you want to continue working on these things
  - ▶ 2nd Edition is what I will use to prepare lectures, but 3rd Edition draft chapters are already available for free.  
<https://web.stanford.edu/~jurafsky/slp3/>
2. This is for Python programmers: NLTK Book - follow the online version. Print version is outdated.  
<http://www.nltk.org/book/>

# Other Materials

1. Coursera course: Introduction to NLP by Dragomir Radev  
<https://www.coursera.org/learn/natural-language-processing> (please enroll)
  - ▶ One more NLP course (taught by Jurafsky and Manning) will restart in September 2016.
  - ▶ There used to be another, more technical course by Michael Collins. I have the lectures downloaded - anyone interested can get them from me.
2. Language and Computers (Dickinson, Brew and Meurers) - optional. Good to get a general overview, without getting into programming.
3. NACLO website - good resource for some brainstorming about language processing problems.  
<http://www.nacloweb.org/>



Any questions so far?

# Next Class ..

- ▶ To do before next class:
  1. Read the syllabus handbook carefully
  2. Start listening Week 1 lectures in Radev's Coursera course.
- ▶ Next class:
  1. Overview of common problems and tasks in NLP
  2. Assignment 1 description

# One request

I need 2,3 volunteers for next tuesday's class. They should be willing to do a 5 minute demo of any programming project they did so far (e.g., in 516).

Please fill up the questionnaire