



Carnegie Mellon University  
Language  
Technologies  
Institute

# 11-411 Natural Language Processing

## Introduction

---

David R. Mortensen

August 30, 2022

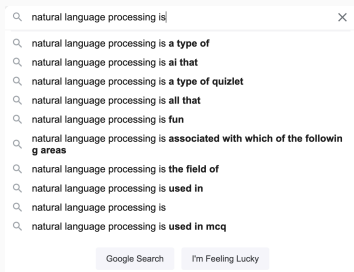
Language Technologies Institute

What is NLP?

---

# NLP is Everywhere

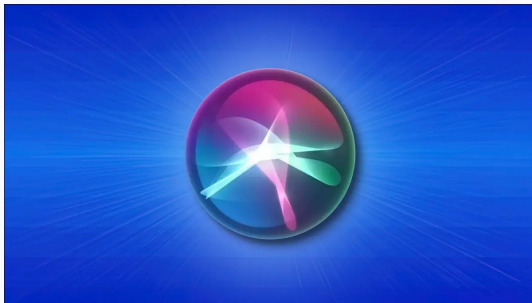
Did you ever wonder how web search engines work...



...or how Google can anticipate what you're searching for?

**That's NLP!**

Did you ever wonder how digital assistants work?



**Thats NLP!**

# NLP is Everywhere

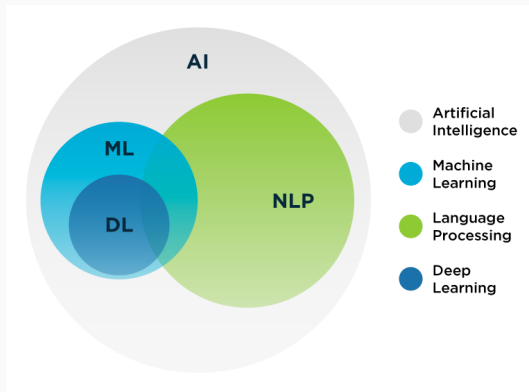
Did you ever wonder how the government is spying on your every word?



Thats also NLP!

# NLP is the Computational Analysis and Synthesis of Speech and Language

- NLP is one of the most important AI fields today
- It is about processing language with computers
- Engineering focus—solving practical problems



## NLP Is Different from Computational Linguistics

- Computational linguistics is the scientific study of language using computers
- Closely related to NLP, but with different aims



## Course Objectives and Overview

---



# Learning Objectives

At the end of this course, a student will be able to:

- Implement a range of fundamental NLP algorithms for question answering, classification, language modeling, sequence labeling, and sequence transduction.
- Leverage morphology, syntax, semantics, discourse, and speech to perform NLP tasks.
- Recognize the class of tasks to which a specific natural language task belongs and identify an appropriate type of model to solve it.
- Design and develop a medium-scale NLP project with multiple components.

At the end of this module, students will be able to:

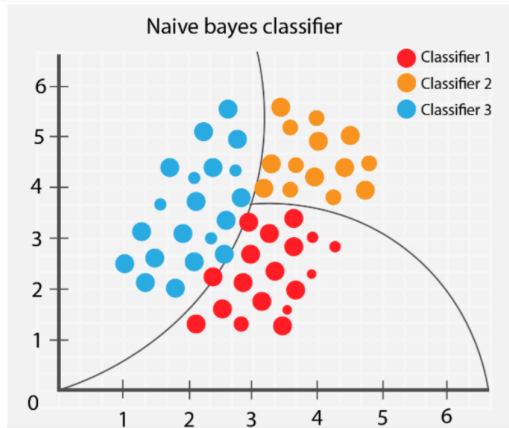
- Build a basic question answering system
- Annotate data for use in an NLP task
- Start on the semester-long project on QA



At the end of this module, students will be able to:

- Identify how the word-structure of a language will interact with NLP tasks by employing basic morphological concepts (root, prefix, suffix, lexeme, inflection, derivation)
- Build a lemmatizer for English verbs

## Module 3: Classification



At the end of this module, students will be able to:

- Characterize kinds of classification tasks in NLP
- Implement naive Bayes and logistic regression classifiers
- Build a basic language identification model

## Module 4: Language Models

At the end of this module, students will be able to:

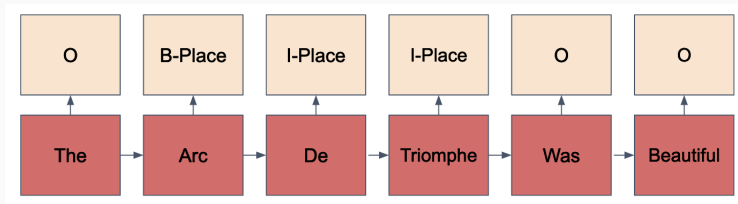
- Articulate what a language model does and the role it plays in NLP tasks
- Implement an n-gram language model
- Describe how neural language models work and what their relationship is to word (and sentence) embeddings



# Sequence Labeling Models

At the end of this module, students will be able to:

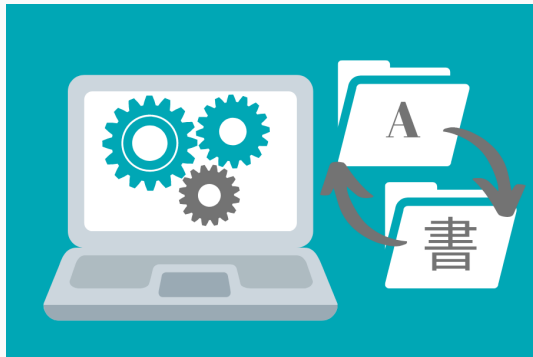
- Determine which NLP tasks can be reduced to sequence labeling
- Implement an HMM, the simplest practical sequence labeling model
- Use conditional random fields (CRFs) for sequence labeling
- Develop an HMM-based sequence-labeling model for part of speech (POS) tagging



# Deep Neural Networks for Sequence Processing

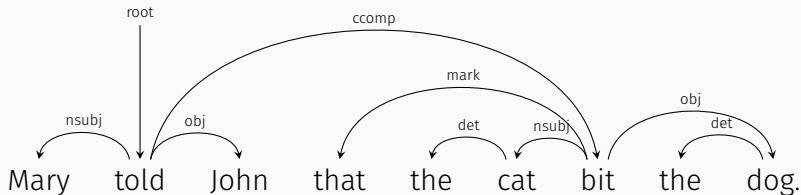
At the end of this module, students will be able to:

- Implement a sequence-to-sequence model using existing toolkits
- Train and test a simple machine translation model using OpenNMT (or possibly Fairseq)



At the end of this module, students will be able to:

- Identify two approaches to computational syntax: constituency and dependency grammar
- Identify tasks to which syntactic structure is relevant
- Implement constituency parsing using the CYK algorithm
- Implement dependency parsing using transition-based parsing





At the end of this module, students will be able to:

- Characterize the meanings of words in relation to one another using wordnets
- Articulate the desirable qualities of meaning representation languages
- Use syntactic parses to build a semantic role labeling system





At the end of this module, students will be able to:

- Explain how coreference resolution and entity linking work
- Articulate how discourse cohesion can be defined and quantified
- Characterize discourse coherence
- Describe the qualities that differentiate dialogue from other kinds of discourse (from the perspective of NLP)

At the end of this module, students will be able to:

- Identify ways in which processing speech is different from processing written language
- Describe appropriate approaches to speech recognition and speech synthesis



## Resources

---

Lectures will cover the fundamentals of the course content. To fully benefit from them, you will need to do the readings and the homework assignments (including the optional ones).

- **Slides** will be provided in advance of each lecture for note-taking purposes
- In-person students are expected to **attend each lecture**
- Remote students will be provided with a **Zoom link** for lectures; this will not be available to students in Pittsburgh except in cases of illness, etc.
- Recordings of the lectures will be made available within 24 hours of the lecture

## Textbook (Free)

- For years, the standard NLP textbook has been Jurafsky and Martin's "Speech and Language Processing," Second Edition (SLP2)
- It is very outdated, but Jurafsky and Martin have been working on a revision for several years
- SLP3 is almost done, and the draft is posted on Jurafsky's website at <https://web.stanford.edu/~jurafsky/slp3/>
- **SLP3 will be our sole textbook and all readings will be drawn from it**
- Some content relevant to assignments and exams will be found in the textbook but not the lectures

**THE TEXTBOOK IS FREE!**

- Each TA will hold two office hours (one hour each) each week
- The instructors will schedule one office hour per week
- Hours are to be announced this week
- You are strongly encouraged to attend office hours frequently

- **Canvas:** <https://canvas.cmu.edu>
  - Submit project milestones
  - Download slides and access videos
  - Check your grade
- **Gradescope:** <https://www.gradescope.com>
  - Submit homework assignments
  - Complete exams
- **Piazza:** <https://piazza.com>
  - Receive course announcements (very important)
  - Ask questions and read answers



# Assessments

---

# Overview of Assessments

ASSESSMENT	POINTS	
Project	101	40.40%
Homework	49	19.60%
Midterm Exam	50	20.00%
Final Exam	50	20.00%
Course Total	250	100.00%

# Project

The largest component of the course will be a semester-long group project: question generation and question answering

- Groups of four, self-selected
- Open-ended (teams may chose from a wide variety of techniques)
- Preliminary evaluation by teaching staff
- Competitive evaluation by students
- Progress report (video)
- Final Report (technical paper)

# Project Components

COMPONENT	POINTS	PERCENTAGE	DUE DATE
Literature Search	3	2.88%	Sep 15
Initial Plan	8	7.69%	Sep 22
Progress Report and Meeting	15	14.42%	Oct 10
Working Question Answering System	15	14.42%	Oct 25
Working Question Generation System	15	14.42%	Nov 15
Final Evaluation	22	21.15%	Nov 29
Final Report	26	25.00%	Dec 9
<b>Project Total</b>	<b>101</b>	<b>100.00%</b>	

## Homework Assignments (× 9)

- There will be nine short homework assignments
- Most will be coding assignments (autograded)
- They are designed to be completed in approximately four hours
- **The lowest two scores will be dropped** (including zero scores due to a student not completing the assignment)
- This means that you may choose which seven assignments to complete if you feel confident in your ability to do well on the others

There will be two exams, consisting of long, multipart problems (typically 3–5 questions).

- **Midterm Exam:** Exam covering Module 1–5 (Oct 13). **3 questions.**
- **Final Exam:** Comprehensive exam, focusing on Modules 6–10 (scheduled by registrar). **6 questions.**

# Time Management

---

## Weekly Task Breakdown

You are expected to spend 12 hours per week on this course. The breakdown of a typical week might be as follows:

TASK	HOURS/WEEK
Study readings and notes	2
Attend lectures	3
Complete HW assignment	3
Work on project	4
<b>Total</b>	<b>12</b>

We encourage you to schedule time now. If you have difficulty completing these tasks in the allotted times, please let us know.



# Policies

---

Life is full of unexpected events. **If you encounter such an event, and are having trouble completing work on time, contact us!** We will try to make a plan with you that allows you to finish the course work in a reasonable time frame. This may include extended deadlines.

We cannot grant you an extension retroactively except in exceptional circumstances (an adverse event occurred very close to the deadline). However, we are happy to extend deadlines for **deaths and funerals (whether yours or those of family members), illnesses, mental health crises or episodes, weddings, important religious and national holidays, job interviews, AND SO FORTH.**

Many people have disabilities, including members of our own families. **We see disabilities as deficits not in disabled people but in the institutions and societies that are structured such that disabled people are disadvantaged.** We wish to do our part to overcome this disparate treatment. If you have a disability (visible or invisible), please let us know as soon as possible (you don't need to tell us the nature of the disability) and work with Disability Services to develop a set of accommodations which we can then approve. These may include extra time on exams, a quiet place in which to take an exam, alt text on all images, documents that work for people with differences in vision, sign language interpretation, captioning, etc.

Throughout human history, some people have been denied the rights and opportunities available to others on the basis of their race, gender, economic class, caste, ancestry, language community, age, religion, beliefs, political affiliation, and abilities (visible and invisible). A single course cannot undo the injustices of history, but we—as a teaching staff—are committed to fighting inequity and promoting inclusion. **We encourage you to join us.** If you feel that you, or those around you, have been treated unfairly based upon their identity (or perceived identity) by us, by other members of the teaching staff, or by other students in the course, we ask that you bring it to our attention so that we can address the wrongs. **We will not retaliate against you, even if your feedback is critical.**

Questions?