

# Course Introduction

**CS 585, Fall 2018**

Introduction to Natural Language Processing

<http://people.cs.umass.edu/~miyyer/cs585/>

**Mohit Iyyer**

College of Information and Computer Sciences

University of Massachusetts Amherst

# natural language processing

# natural language processing

languages that evolved naturally through human use  
e.g., Spanish, English, Arabic, Hindi, etc.

NOT: controlled languages (e.g., Klingon)

NOT: programming languages

# natural language processing

supervised learning: *map text to **X***

unsupervised learning: *learn **X** from text*

generate text from **X**

# How

- Math! Algorithms!
- Data!
- Code!
  - Skill: translating from math to code
  - Skill: debugging math/linguistic/algorithm code
- A little bit of linguistics goes a long way

TAs:

Tu Vu

Subhojyoti Mukherjee

Varun Sharma

email all of us (including me!) at  
[cs585instructors@gmail.com](mailto:cs585instructors@gmail.com)

course website:

<https://people.cs.umass.edu/~miyyer/cs585>

my office hours: Fri 1-2PM in CS258

TA office hours: TBD

# anonymous questions / comments?

- submit questions/concerns/feedback to <https://goo.gl/forms/9Rk1vRfSzqqNI67i1>
- we will go over some/all submitted responses at the start of every class

# undergrad honors section

- COMPSCI H585 discussion section, 1 credit, TTh 11:30-12:30
- Open to any undergraduate (and fulfills honors requirement)
- General discussion of course topics
- Assignments: Paper reading, writeups, presentations
- Questions: [brenocon@cs.umass.edu](mailto:brenocon@cs.umass.edu)



# Prerequisites

- Comfort with programming, algorithmic thinking
  - Ever debugged a graph algorithm? Know its Big-O time and space requirements?
  - CS 220 or 230
- Comfort with probability and mathematical notation
  - Ever used Bayes Rule?
  - CS 240
- Excitement about language!
- Willingness to learn

# Requirements

- (10%) Participation and short exercises
  - Bring pencils/pens/paper to class
- (25%) Problem sets
  - Written: math and concepts
  - Programs: in Python
- (30%) Midterm (in-class, end of October)
- (35%) Final projects (groups of 3)
  - Choose a topic, or select a suggested topic
  - Project Proposal
  - Progress Report
  - In-class presentations
  - Final Report

# Logistics

- Main course website:  
<http://people.cs.umass.edu/~miyyer/cs585/>
- Gradescope/Moodle for homework submissions
- 585-01 and 585-02 sections are the same
- Due next week: HW0, math review.

# Readings

- No need to buy any textbooks!
- Readings will be provided as PDFs on website
- Often draft chapters from Jurafsky and Martin, *Speech and Language Processing*

# Levels of linguistic structure

Discourse

Semantics

Syntax: Constituents

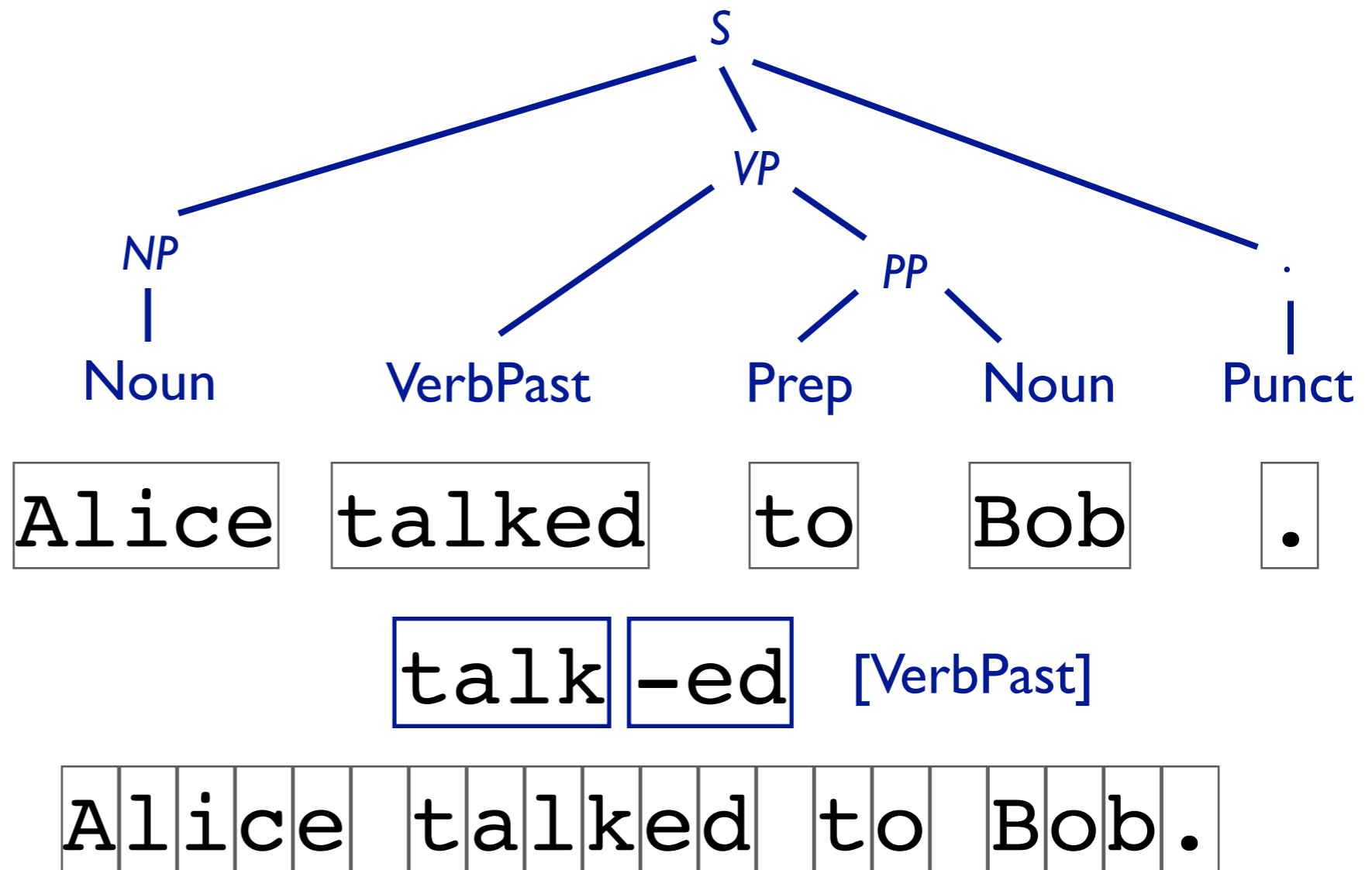
Syntax: Part of Speech

Words

Morphology

Characters

CommunicationEvent(e)    SpeakerContext(s)  
Agent(e, Alice)            TemporalBefore(e, s)  
Recipient(e, Bob)



demos!  
([allennlp.org](http://allennlp.org))

# machine translation

The screenshot shows the Google Translate interface. At the top left is the Google logo, and at the top right is a 'Sign in' button. Below the logo is the word 'Translate' in red. The interface is divided into two main sections: the source text on the left and the translated text on the right. The source text is in Chinese, and the translated text is in English. The source text is enclosed in a blue border, and the translated text is enclosed in a grey border. The source text is: 新华网北京9月2日电（记者刘华）中共中央政治局常委、中央书记处书记刘云山2日在北京会见了由主席斯塔尼舍夫率领的欧洲社会党代表团。刘云山说，中方高度重视中欧关系，愿与欧方一道落实习近平主席与欧方领导人达成的重要共识，围绕打造和平、增长、改革、文明四大伙伴关系，执行好《中欧合作2020战略规划》，深化各领域务实合作。中国共产党愿与欧洲社会党发展良好党际关系，通过多种形式深化治国理政经验交流，促进中欧全面战略伙伴关系持续健康发展。刘云山应询介绍了中共党建有关情况。 The translated text is: Xinhua Beijing September 2 (Reporter Liu) Politburo Standing Committee of the CPC Central Committee Liu Yunshan the 2nd met with the delegation led by the main 席斯塔尼舍夫 European Socialists in Beijing. Liu Yunshan said China attaches great importance to China-EU relations and is willing to work together with the EU President Xi Jinping and implement the important consensus reached by leaders of the EU, around to create peace, growth, reform and civilization of the four partnerships, implementation of good "2020 China-EU cooperation in strategic planning.", deepen pragmatic cooperation in various fields. Communist Party of China is willing to develop good inter-party relations with the European Socialist Party, through various forms of

# NLP today: Story generation

INVESTING 8/28/2014 @ 10:00AM | 121 views



## Earnings for OmniVision Technologies Expected to Fall

Narrative Science

By [Narrative Science](#)

FOLLOW

+ [Comment Now](#) + [Follow Comments](#)

[full bio](#) →

Opinions expressed by Forbes Contributors are their own.

Wall Street is expecting lower profit for **OmniVision Technologies** when the company reports its first quarter results on Thursday, August 28, 2014. Analysts are expecting earnings per share of 39 cents after the company booked a profit of 42 cents a share a year earlier.

The consensus estimate has risen from 16 cents over the past three months. Analysts are expecting earnings of 99 cents per share for the fiscal year. Revenue is projected to eclipse the year-earlier total of \$373.7 million by 2%, finishing at \$381.5 million for the quarter. For the year, revenue is projected to come in at \$1.39 billion.

<http://www.forbes.com/sites/narrativescience/>



# NLP today: Search/summarization

**yelp** Find tacos, cheap dinner, Max's Near amherst, ma Sign Up Log In

Home About Me Write a Review Find Friends Messages Talk

## Oriental Flavor

★ ★ ★ ★ ★ 54 reviews Details

\$\$ · Szechuan, Dim Sum, Cantonese Edit

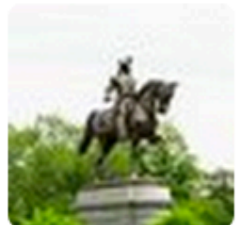
Write a Review Add Photo Share Bookmark



“We had the **crystal shrimp dumplings** that is seen wowing all yelpers, and they were superb.” in 5 reviews



“In addition to having the best pork buns in the area, they also have the best **scallion pancakes** and xiaolongbao.” in 8 reviews



“Usually I eat at the large dim sum restaurants in Boston with the rolling **carts**, but this experience was as good as any.” in 5 reviews

# Ambiguity

- Juvenile Court to Try Shooting Defendant
- Hospitals Are Sued by 7 Foot Doctors
- Alice saw Bob with a telescope.
- Our company is training workers.
- They found that in order to attract settlers -- and make a profit from their holdings -- they had to offer people farms, not just tenancy on manorial estates.

- Check out HW0 on the website
- See you on Thursday